ADVANCES IN ECONOMETRICS
VOLUME 20

ECONOMETRIC ANALYSIS OF FINANCIAL AND ECONOMIC TIME SERIES - PART A

DEK TERRELL
THOMAS B. FOMBY
Editors

ECONOMETRIC ANALYSIS OF FINANCIAL AND ECONOMIC TIME SERIES

ADVANCES IN ECONOMETRICS

Series Editors: Thomas B. Fomby and R. Carter Hill

Recent Volumes:

Volume 15:	Nonstationary Panels, Panel Cointegration, and Dynamic Panels, Edited by Badi Baltagi
Volume 16:	Econometric Models in Marketing, Edited by P. H. Franses and A. L. Montgomery
Volume 17:	Maximum Likelihood of Misspecified Models: Twenty Years Later, Edited by Thomas B. Fomby and R. Carter Hill
Volume 18:	Spatial and Spatiotemporal Econometrics, Edited by J. P. LeSage and R. Kelley Pace
Volume 19:	Applications of Artificial Intelligence in Finance and Economics, Edited by J. M. Binner, G. Kendall and S. H. Chen

ADVANCES IN ECONOMETRICS VOLUME 20, PART A

ECONOMETRIC ANALYSIS OF FINANCIAL AND ECONOMIC TIME SERIES

EDITED BY

DEK TERRELL

Department of Economics, Louisiana State University, Baton Rouge, LA 70803

THOMAS B. FOMBY

Department of Economics, Southern Methodist University, Dallas, TX 75275

ELSEVIER
JAI

Amsterdam – Boston – Heidelberg – London – New York – Oxford
Paris – San Diego – San Francisco – Singapore – Sydney – Tokyo

ELSEVIER B.V.
Radarweg 29
P.O. Box 211
1000 AE Amsterdam,
The Netherlands

ELSEVIER Inc.
525 B Street, Suite 1900
San Diego
CA 92101-4495
USA

ELSEVIER Ltd
The Boulevard, Langford Lane, Kidlington
Oxford OX5 1GB
UK

ELSEVIER Ltd
84 Theobalds Road
London
WC1X 8RR
UK

First edition 2006

British Library Cataloguing in Publication Data
A catalogue record is available from the British Library.

ISBN-10: 0-7623-1274-2
ISBN-13: 978-0-7623-1274-0
ISSN: 0731-9053 (Series)

⑧ The paper used in this publication meets the requirements of ANSI/NISO Z39.48-1992 (Permanence of Paper).
Printed in The Netherlands.

CONTENTS

This page intentionally left blank

DEDICATION

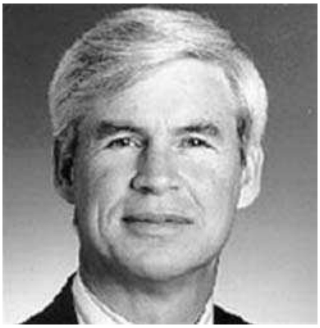
Volume 20 of *Advances in Econometrics* is dedicated to Rob Engle and Sir Clive Granger, winners of the 2003 Nobel Prize in Economics, for their many valuable contributions to the econometrics profession. The Royal Swedish Academy of Sciences cited Rob "for methods of analyzing economic time series with time-varying volatility (ARCH)," while Clive was cited "for methods of analyzing economic time series with common trends (cointegration)." Of course, these citations are meant for public consumption but we specialists in time-series analysis know their contributions go far beyond these brief citations. Consider *some* of Rob's other contributions to our literature: Aggregation of Time Series, Band Spectrum Regression, Dynamic Factor Models, Exogeneity, Forecasting in the Presence of Cointegration, Seasonal Cointegration, Common Features, ARCH-M, Multivariate GARCH, Analysis of High Frequency Data, and CAViaR. *Some* of Sir Clive's additional contributions include Spectral Analysis of Economic Time Series, Bilinear Time Series Models, Combination Forecasting, Spurious Regression, Forecasting Transformed Time Series, Causality, Aggregation of Time Series, Long Memory, Extreme Bounds, Multi-Cointegration, and Non-linear Cointegration. No doubt, their Nobel Prizes are richly deserved. And the 48 authors of the two parts of this volume think likewise. They have authored some very fine papers that contribute nicely to the same literature that Rob's and Clive's research helped build.

For more information on Rob's and Clive's Nobel prizes you can go to the Nobel Prize website http://nobelprize.org/economics/laureates/2003/index.html. In addition to the papers that are contributed here we are publishing remarks by Rob and Clive on the nature of innovation in econometric research that were given during the Third Annual *Advances in Econometrics* Conference at Louisiana State University in Baton Rouge, November 5–7, 2004. We think you will enjoy reading their remarks. You come away with the distinct impression that, although they may claim they were "lucky" or "things just happened to fall in place," having the orientation of building models that solve practical problems has been an orientation that has served them and our profession very well.

We hope the readers of this two-part volume enjoy its contents. We feel fortunate to have had the opportunity of working with these fine authors and putting this volume together.

Thomas B. Fomby
Department of Economics
Southern Methodist University
Dallas, Texas 75275

Dek Terrell
Department of Economics
Louisiana State University
Baton Rouge, Louisiana 70803



Robert F. Engle III
2003 Nobel Prize Winner
in Economics

Sir Clive W. J. Granger, Knight's designation (KB)
2003 Nobel Prize Winner
in Economics

LIST OF CONTRIBUTORS

Elena Andreou	University of Cyprus, Nicosia, Cyprus
Dirk Baur	Joint Research Centre, EU Commission, Ispra, Italy
Ray Y. Chou	Institute of Economics, Academia Sinica and National Chiao Tung University, Taiwan
Giovanni De Luca	Università di Napoli (Parthenope), Napoli, Italy
Dick van Dijk	Erasmus University, Rotterdam, The Netherlands
Jean-Marie Dufour	Université de Montréal, Montreal, Quebec, Canada
Philip Hans Franses	Erasmus University, Rotterdam, The Netherlands
Marc G. Genton	Texas A&M University, College Station, TX, USA
Eric Ghysels	University of North Carolina, Chapel Hill, NC, USA
Christian M. Hafner	Erasmus University, Rotterdam, The Netherlands
Maria S. Heracleous	American University, Washington, DC, USA
Borus Jungbacker	Vrije Universiteit, Amsterdam, The Netherlands
Siem Jan Koopman	Vrije Universiteit, Amsterdam, The Netherlands
Kajal Lahiri	University at Albany – SUNY, Albany, NY
Fushang Liu	University at Albany – SUNY, Albany, NY

Nicola Loperfido	Università di Urbino, (Carlo Bo), Urbino, Italy
John P. Owens	Victoria University of Wellington, Wellington, New Zealand
Dimitris N. Politis	University of California, San Diego, CA, USA
Chor-yiu Sin	Hong Kong Baptist University, Hong Kong, China
Aris Spanos	Virginia Polytechnic Institute, Blacksburg, VA, USA
Douglas G. Steigerwald	University of California, Santa Barbara, CA, USA
Pascale Valéry	HEC, Montreal, Quebec, Canada
Peter A. Zadrozny	Bureau of Labor Statistics, Washington, DC, USA

INTRODUCTION

Dek Terrell and Thomas B. Fomby

The editors are pleased to offer the following papers to the reader in recognition and appreciation of the contributions to our literature made by Robert Engle and Sir Clive Granger, winners of the 2003 Nobel Prize in Economics. Please see the previous dedication page of this volume. This part of Volume 20 of Advances in Econometric focuses on volatility models. The contributions cover a variety of topics and are organized into three broad categories to aid the reader. The first five papers focus broadly on multivariate Generalised auto-regressive conditional heteroskedasticity (GARCH) models. The first four papers propose new models that enhance existing models, while the final paper proposes a test for multivariate GARCH in the models with non-stationary variables. The next three papers examine topics related to high frequency-data. The first of these papers compares asymptotically mean square error (MSE)-equivalent sampling frequencies and window lengths, while the other two papers in this group consider the problem of estimating volatility in the presence of microstructure noise. The last five papers are contributions relevant primarily to univariate volatility models. Of course, we are also pleased to include Rob's and Clive's remarks on their careers and their views on innovation in econometric theory and practice that were given at the third annual Advances in Econometrics Conference held at Louisiana State University, Baton Rouge, on November 5–7, 2004.

Let us briefly review the specifics of the papers presented here. Dirk Baur's "A Flexible Dynamic Correlation Model" proposes a new model that parameterizes the conditional correlation matrix directly to ensure a positive-definite covariance matrix. In the Flexible Dynamic Correlation (FDC) model, the number of exogenous variables can be increased without any risk of an indefinite covariance matrix and an extension, which allows for asymmetric effects of shocks on the time-varying correlation. Simulations are used to compare the performance of FDC to other models and an empirical application uses FDC to model daily and monthly stock market returns for Germany, Japan, the United Kingdom, and the United States.

Results reveal that correlations exhibit less persistence than variances in returns to shocks. Correlations rise in response to jointly positive or negative shocks. However, jointly negative shocks do not appear to increase correlations any more than jointly positive shocks.

In their paper "A Multivariate Skew-GARCH Model," Giovanni De Luca, Marc G. Genton, and Nicola Loperfido introduce a new GARCH model based on the skew-normal distribution. The paper begins by describing the properties of the skew-normal distribution and integrating it into the GARCH model. The paper then turns to an application using returns computed from close to close prices for Dutch, Swiss, Italian, and US markets. The authors note that the US return serves as a proxy for the world news for the period between the closing and opening of the European markets. Results suggest that the US return influences the mean, variance, and higher moments of the European market returns.

Christian M. Hafner, Dick van Dijk, and Philip Hans Franses propose a new semi-parametric model for conditional correlations in the paper "Semi-parametric Modelling of Correlation Dynamics." The number of parameters expands very rapidly in many multivariate GARCH models with the number of assets, which can make estimation intractable for large portfolios. One of the greatest challenges for multivariate GARCH models is to create a flexible formulation for correlations across assets that are tractable even for a large number of assets. This paper attacks this problem by combining univariate GARCH specifications for individual asset volatilities with nonparametric kernel specifications for conditional correlations. The conditional correlations are allowed to depend on exogenous factors, and nonparametric kernel regressions are applied to each conditional correlation individually. An empirical application to the 30 Dow Jones stocks provides an interesting test for the model. Results reveal that the model performs well compared to parametric Dynamic Conditional Correlation (DCC) models in constructing minimum variance portfolios. Interestingly, this semi-parametric model required less time to estimate than simple DCC models with two parameters.

Dimitris Politis' "A Multivariate Heavy-Tailed Distribution for ARCH/GARCH Residuals" extends his earlier univariate work to propose a new distribution for residuals in the multivariate GARCH setting. After reviewing the properties of his proposed heavy-tailed distribution in the multivariate GARCH model, the paper suggests a procedure for maximum likelihood estimation of the parameters. An empirical example using the daily stock returns of IBM and Hewlett–Packard demonstrates the model and finds that the bivariate results are substantially different from those obtained from univariate models.

Chor-yiu Sin's "A Portmanteau Test for Multivariate GARCH when the Conditional Mean is ECM: Theory and Empirical Applications" addresses the issue of testing for multivariate GARCH when the model includes non-stationary variables. The paper first establishes that the Pormanteau test of squared residuals is distributed asymptotically normal, even when the conditional mean is an error correction model. Simulations are then used to assess the performance of the test in samples of size 200, 400, and 800. The test is then used to test for GARCH effects in both the yearly and quarterly Nelson–Plosser datasets and for the intra-daily Hang Seng Index and its derivatives. The results find substantial evidence of GARCH effects in most of the data and suggest that other findings from these error correction models may be sensitive to the omission of GARCH terms.

Empirical volatility estimates depend on the weighting scheme, sampling frequency, and the window length. In their paper, "Sampling Frequency and Window Length Trade-Offs in Data-Driven Volatility Estimation: Appraising the Accuracy of Asymptotic Approximations," Elena Andreou and Eric Ghysels focus on comparisons across different sampling frequencies and window lengths. They focus on volatility filters using monthly, daily, and intra-day observations in the context of a continuous time stochastic volatility model. The paper first presents asymptotically MSE-equivalent one-sided volatility filters for different sampling frequencies for both daily and monthly benchmarks. Monte Carlo simulations are then used to assess the accuracy of these asymptotic results for sample sizes and filters typically used in empirical studies.

Borus Jungacker and Siem Jan Koopman's "Model-Based Measurement of Actual Volatility in High Frequency Data," focuses on the problem of measuring actual volatility in the presence of microstructure noise. The paper first augments a basic stochastic volatility model to account for microstructure noise and intra-daily volatility patterns. The paper then turns to the problem of estimation, which is complicated due to the fact that no closed form for the likelihood function exists. An importance sampling approach is proposed to address these challenges and an application of the model using three months of tick-by-tick data for IBM obtained from the Trades and Quotes database is included to demonstrate the model. Results indicate that models with microstructure noise lead to slightly lower volatility estimates than those without microstructure noise.

John Owens and Douglas Steigerwald also focus on the problem of microstructure noise in "Noise Reduced Realized Volatility: A Kalman Filter Approach." They propose using a Kalman filter to extract the latent squared returns. The optimal smoother is infeasible because it requires knowledge of a latent stochastic volatility state variable. However, the paper uses a rolling regression proxy for high-frequency volatility to create a feasible Kalman smoothing algorithm. Simulation results reveal that the feasible algorithm reduces the mean squared error of volatility, even in the presence of time-varying volatility.

In his paper, ''Modeling the Asymmetry of Stock Movements Using Price Ranges,'' Ray Chou introduces the Asymmetric Conditional Autoregressive Range (ACARR) model. This model extends his earlier work modeling price ranges to model the upward and downward range separately. Interestingly, traditional GARCH software can easily be used to estimate the ACARR model. The paper suggests several potential modifications to reduce sensitivity to outliers and then turns to an empirical example using S&P 500 data for 1962–2000. Focusing on the daily high/low range, daily return squared, and absolute value of daily returns, the results show that the forecasted volatility of ACARR dominates that of Conditional Autoregressive Range (CARR). The paper closes with several potentially promising suggestions for future research.

Although theoretically appealing, the number of empirical applications using stochastic volatility models has been limited by the complexity of estimation procedures required. In the paper ''On a Simple Two-Stage Closed-Form Estimator for a Stochastic Volatility in a General Linear Regression,'' Jean-Marie Dufour and Pascale Valéry propose a computationally inexpensive estimator for the stochastic volatility model. After estimating the conditional mean using ordinary least squares, the parameters of the stochastic volatility model are estimated using a method-of-moments estimator based on three moments. This two-stage three moment (2S–3 M) estimator takes less than a second to compute, compared to the several hours it often takes for generalized method of moments estimators to converge. Under general regularity conditions, the 2S–3 M estimator is asymptotically normal and simulation results from the paper suggest that the estimator performs well in small samples. The paper concludes with an application of the estimator using daily data for the Standard and Poors Composite Price Index.

In their paper, ''The Student's *t* Dynamic Linear Regression: Re-examining Volatility Modeling,'' Maria S. Heracleous and Aris Spanos propose an alternative volatility model. The paper begins by using the Probabilistic Reduction approach to derive the Student's *t* Dynamic Linear Regression (St-DLR). The paper then compares the properties of the St-DLR to those of typical GARCH models and discusses how the St-DLR addresses key issues in the literature. After discussing estimation of the St-DLR model, the paper turns to an example of modeling daily Dow Jones returns in the St-DLR model with the three-month Treasury Bill rate included as an exogenous variable. The empirical results and specification test favor the St-DLR over GARCH models estimated for comparison.

In the paper, "ARCH Models for Multi-period Forecast Uncertainty – A Reality Check Using a Panel of Density Forecasts," Kajal Lahiri and Fushang Liu compare forecast uncertainty of the Survey of Professional Forecasters (SPF) to that predicted by time series models. The paper first focuses on the problem of accurately measuring the uncertainty of survey forecasts by building a model that addresses both the average forecast error variance and the disagreement across forecasters. The paper then compares the survey measures of forecast uncertainty to those from variants of the GARCH and VAR-GARCH models. Results indicate that the survey uncertainty is similar to that of GARCH models during periods of stable inflation, but tend to diverge during periods of structural change.

Peter A. Zadrozny's "Necessary and Sufficient Restrictions for Existence of a Unique Fourth Moment of a Univariate GARCH(p,q) Process" addresses an interesting question that has received substantial attention over the years. His paper proves that an eigenvalue restriction is a necessary and sufficient condition for the existence of a unique 4th moment of the underlying variable of a univariate GARCH process. The approach leads to an inequality restriction that is easy to compute for low-dimensional GARCH processes. The final section of the paper illustrates the approach by computing the restrictions for several models in the literature.

This page intentionally left blank

GOOD IDEAS

Robert F. Engle III

The Nobel Prize is given for good ideas–very good ideas. These ideas often shape the direction of research for an academic discipline. These ideas are often accompanied by a great deal of work by many researchers.

Most good ideas don't get prizes but they are the centerpieces of our research and our conferences. At this interesting *Advances in Econometrics* conference hosted by LSU, we've seen lots of new ideas, and in our careers we have all had many good ideas. I would like to explore where they come from and what they look like.

When I was growing up in suburban Philadelphia, my mother would sometimes take me over to Swarthmore College to the Physics library. It was a small dusty room with windows out over a big lawn with trees. The books cracked when I opened them; they smelled old and had faded gold letters on the spine. This little room was exhilarating. I opened books by the famous names in physics and read about quantum mechanics, elementary particles and the history of the universe. I didn't understand too much but kept piecing together my limited ideas. I kept wondering whether I would understand these things when I was older and had studied in college or graduate school. I developed a love of science and the scientific method. I think this is why I studied econometrics; it is the place where theory meets reality. It is the place where data on the economy tests the validity of economic theory.

Fundamentally I think good ideas are simple. In Economics, most ideas can be simplified until they can be explained to non-specialists in plain language. The process of simplifying ideas and explaining them is extremely important. Often the power of the idea comes from simplification of a collection of complicated and conflicting research. The process of distilling out the simple novel ingredient is not easy at all and often takes lots of fresh starts and numerical examples. Discouragingly, good ideas boiled down to their essence may seem trivial. I think this is true of ARCH and Cointegration and many other Nobel citations. But, I think we should not be offended by this simplicity, but rather we should embrace it. Of course it is easy to do this after 20 years have gone by; but the trick is to recognize good ideas early. Look for them at seminars or when reading or refereeing or editing.

Good ideas generalize. A good idea, when applied to a new situation, often gives interesting insights. In fact, the implications of a good idea may be initially surprising. Upon reflection, the implications may be of growing importance. If ideas translated into other fields give novel interpretations to existing problems, this is a measure of their power.

Often good ideas come from examining one problem from the point of view of another. In fact, the ARCH model came from such an analysis. It was a marriage of theory, time series and empirical evidence. The role of uncertainty in rational expectations macroeconomics was not well developed, yet there were theoretical reasons why changing uncertainty could have real effects. From a time series point of view a natural solution to modeling uncertainty was to build conditional models of variance rather than the more familiar unconditional models. I knew that Clive's test for bilinearity based on the autocorrelation of squared residuals was often significant in macroeconomic data, although I suspected that the test was also sensitive to other effects such as changing variances. The idea for the ARCH model came from combining these three observations to get an autoregressive model of conditional heteroskedasticity.

Sometimes a good idea can come from attempts to disprove proposals of others. Clive traces the origin of cointegration to his attempt to disprove a David Hendry conjecture that a linear combination of the two integrated series could be stationary. From trying to show that this was impossible, Clive proved the Granger Representation theorem that provides the fundamental rationale for error correction models in cointegrated systems.

My first meeting in Economics was the 1970 World Congress of the Econometric Society in Cambridge England. I heard many of the famous economists of that generation explain their ideas. I certainly did not understand everything but I wanted to learn it all. I gave a paper at this meeting at a session organized by Clive that included Chris Sims and Phoebus Dhrymes. What a thrill. I have enjoyed European meetings of the Econometric Society ever since.

My first job was at MIT. I had a lot of chances to see good ideas; particularly good ideas in finance. Myron Scholes and Fischer Black were working on options theory and Bob Merton was developing continuous time finance. I joined Franco Modigliani and Myron on Michael Brennan's dissertation committee where he was testing the CAPM. Somehow I missed the opportunity to capitalize on these powerful ideas and it was only many years later that I moved my research in this direction.

I moved to UCSD in 1976 to join Clive Granger. We studied many fascinating time series problems. Mark Watson was my first PhD student at UCSD. The ARCH model was developed on sabbatical at LSE, and when I returned, a group of graduate students contributed greatly to the development of this research. Tim Bollerslev and Dennis Kraft were among the first, Russ Robins and Jeff Wooldridge and my colleague David Lilien were instrumental in helping me think about the finance applications. The next 20 years at UCSD were fantastic in retrospect. I don't think we knew at the time how we were moving the frontiers in econometrics. We had great visitors and faculty and students and every day there were new ideas.

These ideas came from casual conversations and a relaxed mind. They came from brainstorming on the blackboard with a student who was looking for a dissertation topic. They came from "Econometrics Lunch" when we weren't talking about gossip in the profession. Universities are incubators of good ideas. Our students come with good ideas but they have to be shaped and interpreted. Our faculties have good ideas, which they publish and lecture on around the world. Our departments and universities thrive on good ideas that make them famous places for study and innovation. They also contribute to spin-offs in the private sector and consulting projects. Good ideas make the whole system work and it is so important to recognize them in all their various forms and reward them.

As a profession we are very protective of our ideas. Often the origin of the idea is disputable. New ideas may have only a part of the story that eventually develops; who gets the credit? While such disputes are natural, it is often better in my opinion to recognize previous contributions and stand on their shoulders thereby making your own ideas even more important. I give similar advice for academics who are changing specialties; stand with one foot in the old discipline and one in the new. Look for research that takes your successful ideas from one field into an important place in a new discipline.

Here are three quotations that I think succinctly reflect these thoughts.

- "*The universe is full of magical things, patiently waiting for our wits to grow sharper*"
  Eden Philpotts
- "*To see what is in front of one's nose requires a constant struggle.*"
  George Orwell
- "*To select well among old things is almost equal to inventing new ones*"
  Nicolas Charles Trublet

There is nothing in our chosen career that is as exhilarating as having a good idea. But a very close second is seeing someone develop a wonderful new application from your idea. The award of the Nobel Prize to Clive and me for our work in time series is an honor to all of the authors who contributed to the conference and to this volume. I think the prize is really given to a field and we all received it. This gives me so much joy. And I hope that someone in this volume will move forward to open more doors with powerful new ideas, and receive her own Nobel Prize.

Robert F. Engle III
Remarks Given at Third Annual Advances in Econometrics Conference
Louisiana State University
Baton Rouge, Louisiana
November 5–7, 2004

THE CREATIVITY PROCESS

Sir Clive W. J. Granger, KB

In 1956, I was searching for a Ph.D. topic and I selected time series analysis as being an area that was not very developed and was potentially interesting. I have never regretted that choice. Occasionally, I have tried to develop other interests but after a couple of years away I would always return to time series topics where I am more comfortable.

I have never had a long-term research topic. What I try to do is to develop new ideas, topics, and models, do some initial development, and leave the really hard, rigorous stuff to other people. Some new topics catch on quickly and develop a lot of citations (such as cointegration), others are initially ignored but eventually become much discussed and applied (causality, as I call it), some develop interest slowly but eventually deeply (fractionally integrated processes), some have long term, steady life (combination of forecasts), whereas others generate interest but eventually vanish (bilinear models, spectral analysis).

The ideas come from many sources, by reading literature in other fields, from discussions with other workers, from attending conferences (time distance measure for forecasts), and from general reading. I will often attempt to take a known model and generalize and expand it in various ways. Quite frequently these generalizations turn out not to be interesting; I have several examples of general I(d) processes where d is not real or not finite. The models that do survive may be technically interesting but they may not prove useful with economic data, providing an example of a so-called "empty box," bilinear models, and I(d), d non-integer could be examples.

In developing these models one is playing a game. One can never claim that a new model will be relevant, only that it might be. Of course, when using the model to generate forecasts, one has to assume that the model is correct, but one must not forget this assumption. If the model is correct, the data will have certain properties that can be proved, but it should always be remembered that other models may generate the same properties, for example I(d), d a fraction, and break processes can give similar "long memory" autocorrelations. Finding properties of data and then suggesting that a particular model will have generated the data is a dangerous game.

Of course, once the research has been done one faces the problem of publication. The refereeing process is always a hassle. I am not convinced that delaying an interesting paper (I am not thinking of any of my own here) by a year or more to fix a few minor difficulties is actually helping the development of our field. Rob and I had initial rejections of some of our best joint papers, including the one on cointegration. My paper on the typical spectral shape took over three and a half years between submission and publication, and it is a very short paper.

My favorite editor's comment was that "my paper was not very good (correct) but is was very short," and as they just had that space to fill they would accept. My least favorite comment was a rejection of a paper with Paul Newbold because "it has all been done before." As we were surprised at this we politely asked for citations. The referee had no citations, he just thought that must have been done before. The paper was published elsewhere.

For most of its history time series theory considered conditional means, but later conditional variances. The next natural development would be conditional quantiles, but this area is receiving less attention than I expected. The last stages are initially conditional marginal distributions, and finally conditional multivariate distributions. Some interesting theory is starting in these areas but there is an enormous amount to be done.

The practical aspects of time series analysis are rapidly changing with improvements in computer performance. Now many, fairly long series can be analyzed jointly. For example, Stock and Watson (1999) consider over 200 macro series. However, the dependent series are usually considered individually, whereas what we are really dealing with is a sample from a 200-dimensional multivariate distribution, assuming the processes are jointly stationary. How to even describe the essential features of such a distribution, which is almost certainly non-Gaussian, in a way that is useful to economists and decision makers is a substantial problem in itself.

My younger colleagues sometimes complain that we old guys solved all the interesting easy questions. I do not think that was ever true and is not true now. The higher we stand the wider our perspective; I hope that Rob and I have provided, with many others, a suitable starting point for the future study in this area.

REFERENCE

Stock, J.H., & Watson, M.W. (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: R. F. Engle, & H. White (Eds.), *Cointegration, causality, and forecasting: A festschrift in honour of Clive W.J. Granger.* Oxford: Oxford University Press.

Sir Clive W. J. Granger, KB
Remarks Read at Third Annual Advances in Econometrics Conference
Louisiana State University
Baton Rouge, Louisiana
November 5–7, 2004

This page intentionally left blank

PART I:
MULTIVARIATE VOLATILITY MODELS

This page intentionally left blank

A FLEXIBLE DYNAMIC CORRELATION MODEL

Dirk Baur

ABSTRACT

Existing multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models either impose strong restrictions on the parameters or do not guarantee a well-defined (positive-definite) covariance matrix. I discuss the main multivariate GARCH models and focus on the BEKK model for which it is shown that the covariance and correlation is not adequately specified under certain conditions. This implies that any analysis of the persistence and the asymmetry of the correlation is potentially inaccurate. I therefore propose a new Flexible Dynamic Correlation (FDC) model that parameterizes the conditional correlation directly and eliminates various shortcomings. Most importantly, the number of exogenous variables in the correlation equation can be flexibly augmented without risking an indefinite covariance matrix. Empirical results of daily and monthly returns of four international stock market indices reveal that correlations exhibit different degrees of persistence and different asymmetric reactions to shocks than variances. In addition, I find that correlations do not always increase with jointly negative shocks implying a justification for international portfolio diversification.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 3–31

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20001-4

1. INTRODUCTION

The knowledge of the time-varying behavior of correlations and covariances between asset returns is an essential part in asset pricing, portfolio selection and risk management. Whereas unconditional correlations can easily be estimated, this is not true for time-varying correlations. One approach to estimate conditional covariances and correlations is within a multivariate GARCH model. Other approaches as a moving average specification for the covariances and the variances provide estimates for time-varying correlations but do not parameterize the conditional correlations directly. Interestingly, studies comparing the existing multivariate GARCH models are rare in relation to the existing studies that compare univariate time-varying volatility models (see Engle & Ng, 1993, Pagan & Schwert, 1990 among others). For multivariate GARCH models, there are studies of Bauwens, Laurent, and Rombouts (2005), Engle (2002); Engle and Sheppard (2001) and Kroner and Ng (1998). While Kroner and Ng (1998) compare the main existing models within an empirical analysis, Engle (2002) and Engle and Sheppard (2001) use Monte-Carlo simulations to analyze different models with a focus on the Dynamic Conditional Correlation (DCC) estimator. The study of Bauwens et al. (2005) is the only real survey and includes almost all existing multivariate GARCH models.

The first multivariate GARCH model is proposed by Bollerslev, Engle, and Wooldridge (1988). This model uses the VECH operator and is thus referred to as the VECH model. It does not guarantee a positive-definite covariance matrix and the number of parameters is relatively large. Baba, Engle, Kroner, and Kraft (1991) proposed a multivariate GARCH model, called BEKK (named after the authors), that guarantees the positive definiteness of the covariance matrix.[1] Interestingly, it seems that even restricted versions of the BEKK model have too many parameters since commonly only bivariate models are estimated (see Bekaert & Wu, 2000; Engle, 2002; Karolyi & Stulz, 1996; Kroner & Ng, 1998; Longin & Solnik, 1995; Ng, 2000). In addition, I am not aware of any multivariate GARCH model that is estimated with a higher lag order than GARCH(1,1).

The Constant Correlation Model (CCM) of Bollerslev (1990) does also circumvent the problem of possible non-positive definiteness of the covariance matrix but is restricted since it does not allow correlations to be time varying.

Asymmetric extensions of the existing models are introduced by Kroner and Ng (1998) who proposed the general asymmetric dynamic covariance (ADC) model that nests the VECH, the Factor GARCH, the BEKK model and the CCM.[2] An asymmetric version of the DCC model is proposed by Cappiello, Engle, and Sheppard (2002).

Tse and Tsui (2002) proposed a new multivariate GARCH model that parameterizes the conditional correlation directly by using the empirical correlation matrix and Engle (2002) suggested a time-varying correlation model, called DCC that also parameterizes the conditional correlation directly and enables a two-stage estimation strategy. The Flexible Dynamic Correlations (FDC) estimator suggested in this paper also specifies the conditional correlation directly, but is only a bivariate model. However, in its bivariate form, it is shown to be the most flexible.

The remainder of this paper is as follows: Section 2 discusses the main multivariate GARCH models and focusses on the full and restricted BEKK model and its asymmetric extensions. I also discuss the CCM of Bollerslev (1990) and use this model as a benchmark to analyze the effect of the estimation of time-varying correlations on volatility estimates. The discussion of existing multivariate GARCH models is less complete as in Bauwens et al. (2005) and focusses on additional aspects such as the interpretation of the parameters and potential bias of these parameters due to opposed restrictions in order to guarantee a positive-definite covariance matrix. Section 2.6 introduces the FDC Model. Section 3 presents results of Monte-Carlo simulations for a selection of the discussed models. Section 4 shows the estimation results of the FDC model and the diagonal BEKK model for daily and monthly data and focusses on the persistence and the asymmetry of time-varying correlations. Finally, Section 5 concludes.

2. EXISTING MULTIVARIATE GARCH MODELS

Extending the univariate GARCH model to a n-dimensional multivariate model requires the estimation of n different mean and corresponding variance equations and $(n^2 - n)/2$ covariance equations. I use a simple specification for the mean equation since our interest is the time-varying covariance matrix. Thus, returns are modelled as follows:

$$r_t = \mu + \varepsilon_t, \quad \varepsilon_t \mid \Omega_{t-1} \sim \mathrm{N}(0, H_t) \tag{1}$$

where r_t is a vector of appropriately defined returns and μ a $(N \times 1)$ vector of parameters that estimates the mean of the return series. The residual vector is ε_t with the corresponding conditional covariance matrix H_t given the available information set Ω_{t-1}.

2.1. The VECH Model

The equivalent to a univariate GARCH(1,1) model is given as follows:[3]

$$\mathrm{vech}(H_t) = \Omega + A\mathrm{vech}\left(\varepsilon_{t-1}\varepsilon_{t-1}'\right) + B\mathrm{vech}(H_{t-1}) \tag{2}$$

where H_t is the time-varying $(N \times N)$ covariance matrix, Ω denotes an $(N \times N)(N(N+1)/2 \times 1)$ vector and A and B are $(N(N+1)/2 \times N(N+1)/2)$ matrices. The VECH operator stacks the lower portion of an $(N \times N)$ symmetric matrix as an $(N(N+1)/2 \times 1)$ vector which can be done since the covariance matrix is symmetric by definition. In the bivariate VECH model, the matrices are all (3×3) matrices thus leading to 27 parameters to be estimated.

$$\begin{pmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{pmatrix} = \Omega + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{pmatrix} \tag{3}$$

The diagonal VECH model reduces the number of parameters by using diagonal matrices A and B. However, even for this special case a positive-definite covariance matrix is not guaranteed.[4] Hence, I do not present this model in its asymmetric extension and dispense with a discussion.

2.2. The BEKK Model

The BEKK model was introduced by Baba et al. (1991) and can be seen as an improvement to the VECH model (introduced by Bollerslev et al., 1988). First, the number of parameters is reduced and second, the positive definiteness of the covariance matrix is guaranteed.

I initially present the full (unrestricted) BEKK model and its asymmetric extension and then restrict this model to the diagonal BEKK.[5]

The covariance matrix of the unrestricted BEKK model is

$$H_t = A'A + B'\varepsilon_{t-1}\varepsilon_{t-1}'B + C'H_{t-1}C \tag{4}$$

A, B and C are matrices of parameters with appropriate dimensions. It is obvious from the equation above that the covariance matrix is guaranteed to be positive definite as long as $A'A$ is positive definite. Furthermore, the parameters are squared or cross-products of themselves leading to variance and covariance equations without a univariate GARCH counterpart (see also Eq. (7)). Note that this is not true for the diagonal VECH model which is a simple extension of univariate GARCH models to a multivariate form.

The asymmetric extension of this model introduced by Kroner and Ng (1998) bases on the univariate asymmetric GARCH model proposed by Glosten, Jagannathan, and Runkle (1993). Here, the covariance matrix is given as follows:

$$H_t = A'A + B'\varepsilon_{t-1}\varepsilon'_{t-1}B + C'H_{t-1}C + D'\eta_{t-1}\eta'_{t-1}D \tag{5}$$

where $\eta_{i,t} = \min\{\varepsilon_{i,t}, 0\}$ and $\eta_t = (\eta_{1,t}, \eta_{2,t}, \ldots)'$. Thus, this extension can capture asymmetric effects of shocks by additionally including negative shocks and still guarantees the positive definiteness of the covariance matrix.

To clarify the difficulties in interpreting the parameters of the covariance matrix, I consider the general BEKK model in bivariate form. $h_{11,t}$ and $h_{22,t}$ denote the conditional variances of the underlying return series and $h_{12,t}$ is their covariance:

$$\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix} = A'A + \begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix} \begin{pmatrix} \varepsilon^2_{1,t-1} & \varepsilon_{t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon^2_{2,t-1} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$
$$+ \begin{pmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{pmatrix} \begin{pmatrix} h_{11,t-1} & h_{12_t-1} \\ h_{12,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \tag{6}$$

Without using matrices (see Eq. (6) above), I get the following form:

$$\begin{aligned} h_{11,t} &= a^2_{11} + b^2_{11}\varepsilon^2_{1,t-1} + 2b_{11}b_{21}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b^2_{21}\varepsilon^2_{2,t-1} + c^2_{11}h_{11,t-1} \\ &\quad + 2c_{11}c_{21}h_{12,t-1} + c^2_{21}h_{22,t-1} \\ h_{12,t} &= a_{12}a_{11} + b_{11}b_{12}\varepsilon^2_{1,t-1} + (b_{12}b_{21} + b_{11}b_{22})\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b_{21}b_{22}\varepsilon^2_{2,t-1} \\ &\quad + c_{11}c_{12}h_{11,t-1} + (c_{12}c_{21} + c_{11}c_{22})h_{12,t-1} + c_{21}c_{22}h_{22,t-1} = h_{21,t} \\ h_{22,t} &= a^2_{12} + a^2_{22} + b^2_{12}\varepsilon^2_{2,t-1} + 2b_{12}b_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + b^2_{22}\varepsilon^2_{2,t-1} + c^2_{12}h_{11,t-1} \\ &\quad + 2c_{12}c_{22}h_{12,t-1} + c^2_{22}h_{22,t-1} \end{aligned} \tag{7}$$

The latter formulation clarifies that even for the bivariate model the interpretation of the parameters may be misleading since there is no equation that does exclusively possess its own parameters, i.e. parameters that exclusively govern an equation. Hence, it is possible that a parameter estimate is biased by the fact that it influences two equations simultaneously or by the sole number of regressors (see also Tse, 2000). For example, the regressors $\varepsilon^2_{2,t-1}$ and the regressor $h_{22,t-1}$ in the first variance equation ($h_{11,t}$) could both be viewed as a volatility spillover from the second return. In addition, the statistical significance of the parameters is also unclear due to the combinations of different parameters serving as new coefficients for particular regressors.

These critics do not all apply to the diagonal BEKK model where both parameter matrices are diagonal. Thus, the off-diagonal elements are all equal to zero (apart from the constant term $A'A$). The number of parameters to be estimated is significantly lower while maintaining the main advantage of this specification, the positive definiteness of the conditional covariance matrix. The diagonal BEKK model is given by the following equations:

$$\begin{aligned}
h_{11,t} &= a^2_{11} + b^2_{11}\varepsilon^2_{1,t-1} + c^2_{11}h_{11,t-1} \\
h_{22,t} &= a^2_{11} + a^2_{22} + b^2_{22}\varepsilon^2_{2,t-1} + c^2_{22}h_{22,t-1} \\
h_{12,t} &= h_{21,t} = a_{11}a_{22} + b_{11}b_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{11}c_{22}h_{12,t-1} \\
h_{21,t} &= h_{12,t}
\end{aligned} \tag{8}$$

This model exhibits essentially the same problems as the Full BEKK model since there is no parameter in any equation that exclusively governs a particular covariance equation. Hence, it is not clear whether the parameters for h_{12} are just the result of the parameter estimates for h_{11} and h_{22} or if the covariance equation alters the parameter estimates of the variance equations. In addition, the model is not very flexible and can therefore be misspecified. For example, assuming that the persistence of the volatility is relatively high for both return series, say $b^2_{ii} + c^2_{ii} = 0.05 + 0.90 = 0.95$ for $i = 1, 2$, then the persistence of the covariance must be equally high, $b_{ii}b_{jj} + c_{ii}c_{jj} = 0.05 + 0.9 = 0.95$ for $i = 1$ and $j = 2$. However, if the covariance exhibits a different degree of persistence than the volatilities, it is clear that either the volatility or the covariance process is misspecified.

2.3. Dynamic Correlation Models

While the previously discussed models parameterize the covariance directly, we now briefly discuss the existing models that parameterize the correlation process directly. Tse and Tsui (2002)[6] assume that the time-varying conditional correlation matrix Γ_t is given as follows:

$$\Gamma_t = (1 - \theta_1 - \theta_2)\Gamma + \theta_1 \Gamma_{t-1} + \theta_2 \Psi_{t-1} \tag{9}$$

where Γ is an unconditional $K \times K$ correlation matrix and Ψ_{t-1} is a $K \times K$ matrix that contains lagged cross-sectional observations of the underlying returns r_t.

The parameters θ_1 and θ_2 are assumed to be non-negative and must also satisfy $\theta_1 + \theta_2 \leq 1$. Since Γ_t is a weighted average of Γ, Γ_{t-1} and Ψ_{t-1}, it is positive definite if its elements are positive definite. Of course, the model performance also depends on the specification of Ψ_{t-1}. In addition, it has to be noted that θ_1 and θ_2 have to be restricted in the optimization process and that the evolution of small and large correlation matrices only depends on two parameters.

Engle (2002) makes essentially use of the same model but specifies Ψ_{t-1} as $z_{t-1}z'_{t-1}$ where z_t are standardized residuals of the underlying returns. Engle (2002) additionally shows that a two-step estimation strategy is feasible. In the first step, the variances are estimated univariate and in the second step, the correlation processes are estimated. Again, even large correlation matrices do only depend on two parameters.

2.4. Constant Correlation Model and Zero Correlation Model

In order to test whether the estimates of the parameters in the conditional-variance matrix are robust with respect to the correlation assumption (see Tse, 2000) I also analyze the CCM of Bollerslev (1990) and a restricted version with a correlation coefficient of zero, i.e. Zero Correlation Model (ZCM).

The bivariate CCM is given by

$$\begin{aligned} h_{11,t} &= a_{11} + b_{11}\varepsilon^2_{1,t-1} + c_{11}h_{11,t-1} \\ h_{22,t} &= a_{22} + b_{22}\varepsilon^2_{2,t-1} + c_{22}h_{22,t-1} \\ h_{12,t} &= \rho\sqrt{h_{11,t}h_{22,t}} \\ h_{21,t} &= h_{12,t} \end{aligned} \tag{10}$$

where ρ is a parameter that can be estimated almost freely (ρ must be in the range $[-1, 1]$) and is equal to the empirical correlation coefficient (see Bollerslev, 1990). In contrast to the BEKK model there is one parameter (ρ) in the CCM that exclusively governs the covariance equation. Note that the CCM exhibits time-varying covariances but only constant correlations.[7]

Setting ρ to zero implies a model that is called the ZCM. I will use both the CCM and the ZCM to analyze how the modelling of the covariances affects the variance estimates.

2.5. Asymmetric Extensions

While it is straightforward in the diagonal BEKK Model to analyze whether the covariance exhibits the same degree of persistence as the variances, the relevant parameter estimates measuring the persistence of shocks are potentially influenced by each other leading to biased parameter estimates. This is also true for the full BEKK Model and possibly more severe due to the larger number of parameters. The same problem arises for the asymmetric extensions of the models. To illustrate this, I analyze the asymmetric extensions proposed by Kroner and Ng (1998) and focus on the diagonal BEKK model.

The following asymmetric covariance equations for the bivariate case within the full BEKK model are:

$$\begin{aligned} h_{11,t} &= \ldots + d_{11}^2\eta_{1,t-1}^2 + 2d_{11}d_{21}\eta_{1,t-1}\eta_{2,t-1} + d_{21}^2\eta_{2,t-1}^2 \\ h_{22,t} &= \ldots + d_{12}^2\eta_{1,t-1}^2 + 2d_{12}d_{22}\eta_{1,t-1}\eta_{2,t-1} + d_{22}^2\eta_{2,t-1}^2 \\ h_{12,t} &= \ldots + d_{11}d_{12}\eta_{1,t-1}^2 + (d_{12}d_{21} + d_{11}d_{22})\eta_{1,t-1}\eta_{2,t-1} + d_{21}d_{22}\eta_{2,t-1}^2 \end{aligned} \tag{11}$$

where $\eta_{i,t} = \min\{\varepsilon_{i,t}, 0\}$ and $\eta_t = (\eta_{1,t}, \eta_{2,t}, \ldots)'$. Eq. (11) shows that the number of parameters and its combinations render an interpretation of any asymmetry of the impact of shocks on the conditional (co-)variance difficult.

For the diagonal BEKK model (see Eq. (8)) the asymmetric extension is

$$\begin{aligned} h_{11,t} &= \ldots + d_{11}^2\eta_{1,t-1}^2 \\ h_{22,t} &= \ldots + d_{22}^2\eta_{2,t-1}^2 \\ h_{12,t} &= \ldots + d_{11}d_{22}\eta_{1,t-1}\eta_{2,t-1} \end{aligned} \tag{12}$$

where $\eta_{i,t} = \min\{\varepsilon_{i,t}, 0\}$ and $\eta_t = (\eta_{1,t}, \eta_{2,t}, \ldots)'$.

Here, the covariance reacts to negative shocks $\eta_{i,t}$ as determined by the asymmetry implied by the variance equations or vice versa. For example, assuming that variance h_{11} does not react asymmetrically to positive and negative shocks ($d_{11} = 0$) and variance h_{22} does ($d_{22} = 0.2$), the asymmetric effect for the covariance would be zero ($d_{11}d_{22} = 0$). Consequently, if there is an asymmetric effect of the covariance, either the variance equation or the covariance equation will be misspecified. Another example is the case where the asymmetry of the covariance is equal to 0.2. Then, the parameters d_{11} or d_{22} would have to be very large to capture this covariance asymmetry (e.g. $d_{11} = d_{22} = \sqrt{0.2}$).[8] This problem is potentially more severe in the full BEKK model.

The asymmetric extension of the CCM (see Eq. (10)) introduced by Kroner and Ng (1998) has the variance equations of the diagonal BEKK model and the covariance equation as given in the original model. Again, this model could be used as a benchmark to analyze how variance estimates change when correlations are modelled time varying. This question is further examined in the simulation study in Section 3. Cappiello *et al*. (2002) develop an asymmetric version of the DCC model of Engle (2002).

The next section introduces a new bivariate model that reduces the number of parameters compared to the full BEKK model and extends the flexibility compared to the other multivariate GARCH models.

2.6. Flexible Dynamic Correlations (FDC)

I propose a new bivariate model that is more flexible than the models discussed and also parameterizes the conditional correlations directly.[9] I write the covariance matrix H_t in the following form:

$$H_t = D_t R_t D_t \tag{13}$$

where D_t is a diagonal matrix with the roots of the variances on the main diagonal and R_t is a correlation matrix. In the bivariate case, the correlation matrix R_t is

$$R_t = \begin{pmatrix} 1 & \rho_t \\ \rho_t & 1 \end{pmatrix} \tag{14}$$

with ρ_t denoting the correlation between two series. H_t is positive definite if R_t is positive definite. This is guaranteed as long as $|\rho_t| < 1$. Thus I restrict $|\rho_t|$ to be smaller than 1 by using the following transformation:

$$\rho_t^\star = \frac{\rho_t}{\sqrt{1+\rho_t^2}} \tag{15}$$

where $\rho_t^\star$ is the correlation restricted to lie in the interval $(-1; 1)$. This restriction allows the use of own parameters for the correlation (covariance) equation and to include additional regressors without risking semi-definite or indefinite covariance matrices.

Tsay (2002) uses the same idea but restricts ρ_t by a Fisher transformation as follows:

$$\rho_t^{\star} = \frac{\exp(\rho_t) - 1}{\exp(\rho_t) + 1} \tag{16}$$

However, Tsay does not evaluate the performance of this estimator as is done in Section 3. Furthermore, I also present evidence that the Fisher transformation is more restrictive and thus less adequate than the transformation proposed in Eq. (15).

The FDC model is specified by the following equations:[10]

$$\begin{aligned} h_{11,t} &= a_{11} + b_{11}\varepsilon_{1,t-1}^2 + c_{11}h_{11,t-1} \\ h_{22,t} &= a_{22} + b_{22}\varepsilon_{2,t-1}^2 + c_{22}h_{22,t-1} \\ \rho_t &= a_{12} + b_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{12}\rho_{t-1} \\ \rho_t^{\star} &= \frac{\rho_t}{\sqrt{1 + (\rho_t)^2}} \\ h_{12,t} &= \rho_t^{\star} \cdot \sqrt{h_{11,t}h_{22,t}} \end{aligned} \tag{17}$$

The FDC model is a dynamic correlation model since $\rho_t^{\star}$ is a time-varying process. The covariance does possess its own parameters and the covariance matrix is always guaranteed to be positive definite. The model allows to assess the degree of persistence and to compare this persistence with the volatility persistence.

Apart from using the cross product $\varepsilon_{1,t-1}\varepsilon_{2,t-1}$ to model the correlation equation (see Eq. (17)), I additionally use the cross product of the standardized residuals $z_{1,t-1}$ and $z_{2,t-1}$ in order to analyze the different behavior of the correlation process.[11] Tse (2000) points out that there is no a priori reason to expect the standardized residuals to be a better specification. However, it can be argued that the use of z_t is a more natural specification for the conditional correlations (e.g. see Engle, 2002; Tse, 2000). I refer to the model using the raw residuals ε_t as FDC_ε (see Eq. (17)) and to the model using the standardized residuals z_t as FDC_z. The correlation equation for the FDC_z model is given by

$$\rho_t = a_{12} + b_{12}z_{1,t-1}z_{2,t-1} + c_{12}\rho_{t-1} \tag{18}$$

The next subsection demonstrates the potential extensions of the FDC model including asymmetries.

2.7. Extensions of the FDC Model

An extension of the presented FDC models can also capture asymmetric effects of the time-varying correlation. Thus, h_{11} and h_{22} and ρ are specified as follows:

$$
\begin{aligned}
h_{11,t} &= a_{11} + b_{11}\varepsilon_{1,t-1}^2 + c_{11}h_{11,t-1} + d_{11}\eta_{1,t-1}^2 \\
h_{22,t} &= a_{22} + b_{22}\varepsilon_{2,t-1}^2 + c_{22}h_{22,t-1} + d_{22}\eta_{2,t-1}^2 \\
\rho_t &= a_{12} + b_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{12}\rho_{t-1} + d_{12}\eta_{1,t-1}\eta_{2,t-1}
\end{aligned}
\tag{19}
$$

Again, $\eta_{i,t} = \min\{\varepsilon_{i,t}, 0\}$ with η_t containing only the negative shocks of the returns at t. One important feature of the FDC model is that it does not require similar variance and correlation equations as this is necessary for many other multivariate GARCH models. For example, the FDC model is well defined (does not risk an indefinite covariance matrix) even if the variance equations are specified without any asymmetric regressors and the asymmetry is only modelled in the correlation equation. This feature can also be used to include additional regressors (e.g. thresholds or spillover effects) in the correlation equation without risking an indefinite covariance matrix. For example, the conditional correlation equation could also be specified as follows (independently of the variance equations):

$$
\rho_t = a_{12} + b_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{12}\rho_{t-1} + d_{12}\eta_{1,t-1}\eta_{2,t-1} + d_{12}^{\star}X \tag{20}
$$

where $d_{12}^{\star}$ is a vector capturing the effect of the matrix X of exogenous variables.

It should also be noted that the parameter b_{12} in Eq. (20) captures all possible combinations of shocks: two positive shocks, two negative shocks and shocks of opposite signs. Hence, this parameter contains more information than the ARCH parameter in an original univariate GARCH model. However, the FDC model would offer the possibility to estimate the different impact of shocks on their joint conditional correlation.

Essentially, the FDC model is different from the other models since the correlation matrix need not be a weighted average of positive-definite matrices. Thus, the model is more flexible.

2.8. Estimation

The estimation of the models based on a sample of T observations of the return vector r_t is done through numerical maximization of a likelihood function by using the BHHH algorithm of Berndt, Hall, Hall, and Hausman (1974) and by assuming normally distributed returns:

$$\log L(\theta; r_1, \ldots, r_T) = -T/2 \log(2\pi) - 1/2 \log(|H_t|) - 1/2\varepsilon_t' H_t^{-1}\varepsilon_t \quad (21)$$

By using $H_t = D_t R_t D_t$, I can write the above likelihood function also as

$$\log L(.) = -1/2 \sum_t \left(n \log(2\pi) + 2 \log |D_t| + \log |R_t| + z_t' R_t^{-1} z_t\right) \quad (22)$$

This separation shows that a two-step estimation procedure is feasible and that variances and correlations can be estimated separately (see Engle, 2002). The two-stage approach has mainly the advantage that the dimensionality of the maximization problem is reduced which accelerates the maximization process. Note that the standard errors are corrected for the two stages in the estimation as in Engle and Sheppard (2001). Furthermore, the standard errors of the estimates in each stage are calculated using the quasi-maximum likelihood methods of Bollerslev and Wooldridge (1992), i.e. the standard errors are robust to the density function underlying the residuals.

3. SIMULATIONS

In this section, I compare the covariance estimates of the diagonal BEKK, the FDC_ε model, the FDC_z model, the CCM and the ZCM. I use the CCM and the ZCM to compare the variance estimates and to analyze the impact of the covariance specification on the variance estimates (Tse, 2000 suggested such an analysis[12]). The simulations and tests are partially similar to the ones undertaken by Engle (2002).[13] I simulate different bivariate GARCH models 200 times with 1,000 observations. The data-generating process consists of $T = 1,000$ Gaussian random numbers ε_i for $i = 1, 2$ with mean zero and variance one transformed to a bivariate GARCH model with a time-varying covariance matrix H_t with a given (time-varying) correlation (see below) and the following variance equations:

$$\begin{aligned} h_{11,t} &= 0.01 + 0.04\varepsilon_{1,t-1}^2 + 0.95h_{11,t-1} \\ h_{22,t} &= 0.01 + 0.20\varepsilon_{2,t-1}^2 + 0.50h_{22,t-1} \end{aligned} \quad (23)$$

The variance given by h_{11} is highly persistent and the variance h_{22} is less persistent. Given these variances I use different correlation processes in the simulations:

(i) constant correlations: $\rho_t^\star = 0.5$ and (ii) highly persistent time-varying correlations $\rho_t^\star = \alpha + \beta \sin(t/(50f))$ with a fast sine function given by $\alpha = 0$, $\beta = 0.5$ and $f = 1$ and a slow sine given by $\alpha = 0$, $\beta = 0.9$ and $f = 5$.

The different correlation processes are plotted in Fig. 1 and additionally show how the transformations of the FDC model and the Tsay (2002) model change the true correlation processes. It can be seen that the Fisher transformation changes the underlying correlation process considerably more than the transformation used by the FDC model. For example, in the middle panel of Fig. 1 it is evident that the Fisher transformation (dotted line) lowers the amplitude of the sinus function significantly more than the FDC transformation (dashed line) which reduces the true amplitude only slightly.

For the asymmetric extensions of the models I use the following variance equations:

$$\begin{aligned} h_{11,t} &= 0.01 + 0.04\varepsilon_{1,t-1}^2 + 0.85h_{11,t-1} + 0.1\eta_{1,t-1}^2 \\ h_{22,t} &= 0.01 + 0.10\varepsilon_{2,t-1}^2 + 0.50h_{22,t-1} + 0.2\eta_{2,t-1}^2 \end{aligned} \tag{24}$$

The correlation processes are the same as for the non-asymmetric models.

I compare the estimates for $h_{11,t}$, $h_{22,t}$ and $\rho_t^\star$ with the true variance and covariance series by (i) the mean absolute deviation (MAD) and (ii) the means of the correlations of the true covariance series ($h_{ij,t}$ for $i, j = 1, 2$) with the estimated covariance series. The means of the correlations are also computed since they provide a measure of the fit of the estimated model compared to the simulated one.

3.1. Simulation Results

Table 1 presents the results for the simulated models (diagonal BEKK, FDC_ε, FDC_z, CCM and ZCM) in a non-asymmetric specification. The table contains the results for the mean absolute deviation (MAD) and the mean of the correlation of the estimated process (variances and correlations) with the true simulated series. The values denoted with a star indicate the minimum MAD or maximum (mean of correlation) value among the estimated models and among the different correlation processes (constant correlations, fast sine function and sine function).

Constant correlations are best estimated by the CCM and time-varying correlations are best estimated by the FDC_z model and the diagonal BEKK model. The diagonal BEKK model performs best for the fast sine function.

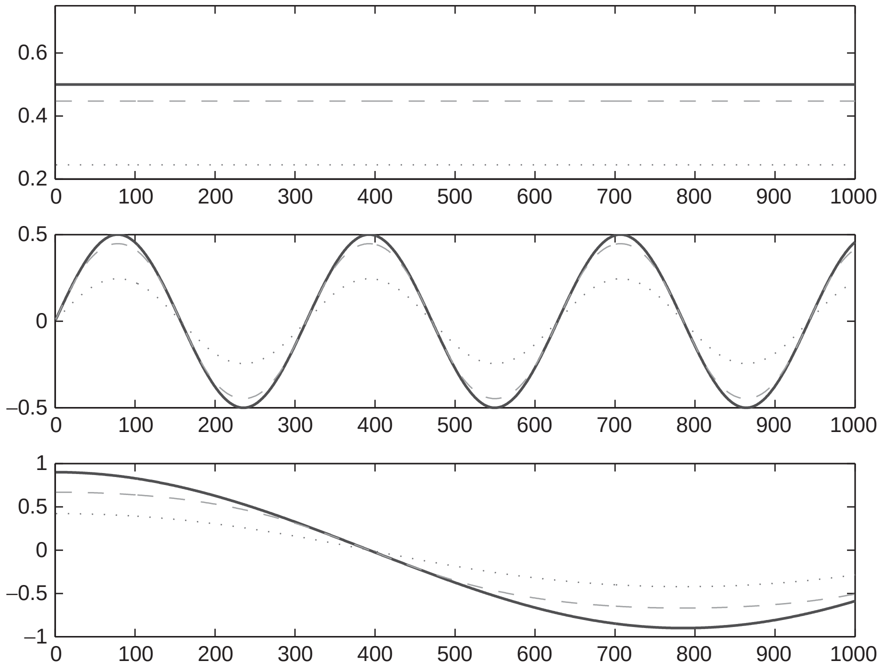


Fig. 1. Simulated Correlation Processes. Solid Line, True Correlation; Dashed Line, FDC model; Dotted Line, Tsay model.

Table 1. Simulation Results – Multivariate GARCH Models.

	D.BEKK	FDC_{ε}	FDC_z	CCM	ZCM
MAD					
Correlations					
Constant	0.2011	0.0303	0.0295	0.0169[a]	0.4995
Fast sine	0.1619[a]	0.2172	0.1859	0.5662	0.5670
Sine	0.1734	0.1244	0.1202[a]	0.2199	0.2935
Variance 1					
Constant	0.1045	0.0617	0.0562[a]	0.0585	0.0636
Fast sine	0.1668	0.0744	0.0997	0.0660[a]	0.0667
Sine	0.0631	0.0601	0.0586[a]	0.0600	0.0587
Variance 2					
Constant	0.0041	0.0024	0.0023	0.0022[a]	0.0024
Fast sine	0.0134	0.0041	0.0045	0.0026	0.0023[a]
Sine	0.0038	0.0026	0.0029	0.0024	0.0023[a]
Mean of correlations					
Correlations					
Constant	0.0761	0.4667	0.4751	1.0000[b]	
Fast sine	0.9456[b]	0.9056	0.9267	0.0008	
Sine	0.6188	0.7790	0.7795[b]	0.0252	
Variance 1					
Constant	0.9832	0.9881	0.9887[b]	0.9868	0.9795
Fast sine	0.9620	0.9840[b]	0.9661	0.9796	0.9801
Sine	0.9937[b]	0.9851	0.9876	0.9856	0.9801
Variance 2					
Constant	0.9543	0.9892	0.9917	0.9925	0.9926[b]
Fast sine	0.8156	0.9809	0.9878	0.9887	0.9920[b]
Sine	0.9467	0.9891	0.9904	0.9924[b]	0.9890

Note: $T = 1{,}000$, 200 iterations.
[a]Denotes the best model (minimum value in row).
[b]Denotes the best model (maximum value in row).

However, the difference in the FDC_z model is small. The good performance of the diagonal BEKK model can be explained with the similar characteristics of the correlation and the variance processes. In this case, the diagonal BEKK model can be assumed to be least biased.[14] Moreover, it is important to emphasize that the FDC model performs clearly better for constant correlations compared to the diagonal BEKK model. This can be attributed to the greater flexibility of the FDC model.

The results for the variances show that the CCM, the ZCM and the FDC_z model perform best. The higher correlation of the estimates of the variances with the true variances for the ZCM are an indication that the correlation process is not relevant for the variance estimates.[15] Estimating time-varying correlations (instead of assuming a zero or constant correlation) does even seem to influence variance estimates negatively.

For the asymmetric models, results do not change considerably. Consequently, I do not report these results. However, I further analyze the behavior of the asymmetric FDC_z model when a two-step procedure is used (see Engle, 2002). This model is estimated for the previous correlation processes with an asymmetric effect. Additionally, the simulations are performed for $T = 1{,}000$ and 2,000. Results in Table 2 show that the two-step estimation leads to similar results as the one-step estimation strategy according to the mean of the correlation coefficient (of the true correlation process and its estimate).

The constant correlation process is an exception since the values are considerably higher compared to Table 1. This can be explained with the fact that the addition of an asymmetric term introduces a time-varying component into the correlation process. Furthermore, the performance increased for all correlation processes with a larger sample size of $T = 2{,}000$.

I conclude from the simulation results that correlation estimates are the closest to the true values in the FDC model for time-varying correlations and constant correlations among the time-varying correlation models. However, constant correlations are best estimated by a CCM. In addition, the FDC_z model performs better than the FDC_ε model which I attribute to the variance correction of the shocks that potentially leads to less noise in the correlation process. Evidently, the true advantages of the FDC model are not fully demonstrated by this simulation study since no exogenous regressors or thresholds in the correlation process are analyzed.

Table 2. Simulations Results – FDC_z Model (Two-Step Procedure).

Asymmetric FDC_z Model Mean of Correlation	$T = 1{,}000$	$T = 2{,}000$
Constant + asymmetry	0.8397	0.9237
Fast sine + asymmetry	0.9044	0.9054
Slow sine + asymmetry	0.7470	0.9289

Note: 200 iterations, variances are assumed to be perfectly estimated.
Data-generating process for asymmetric correlation is: $\rho_t^\star = \cdots + 0.1_{\eta,t-1\ \eta 2,t-1}$($\rho_t^\star$ is restricted as in the FDC_z model to guarantee positive-definite covariance matrices).

4. EMPIRICAL RESULTS

I estimate the asymmetric versions of the FDC_z and the diagonal BEKK model and use daily (close-to-close) continuously compounded returns of the following MSCI stock indices: Germany, Japan, the United Kingdom and the United States of America. The indices span a time period of approximately 5 years from April 30, 1997 until December 30, 2001 with $T = 1{,}176$ observations for each stock index. Non-trading days in a market are included to synchronize the data.[16] This implies that conditional correlations decrease on non-trading days. I am aware of the fact that estimates for close-to-close daily data can be biased if trading hours differ (e.g. Japan and the US).[17] Owing to these potential problems, I also use monthly data spanning from December 1969 until April 2002 with $T = 389$ observations for each index. I additionally use this type of data to analyze any differences in the characteristics of time-varying correlations between daily and monthly data.

Table 3 contains the descriptive statistics for the daily and monthly returns and Table 4 reports the unconditional empirical correlation coefficient for these return series, respectively.

Interestingly, whereas the unconditional correlations are higher for monthly data than for daily data, the correlation of Germany and the

Table 3. Descriptive Statistics.

	Mean	Variance	Skewness	Kurtosis
Daily Data				
Germany	0.000	2.441	−0.860	27.959
Japan	−0.026	2.729	2.137	48.669
UK	0.002	1.351	−0.184	7.327
US	0.023	1.603	−0.517	14.943
Monthly Data				
Germany	0.606	35.513	−85.396	5167.311
Japan	0.772	42.308	0.439	6131.782
UK	0.573	42.634	129.330	15664.007
US	0.591	20.136	−50.565	2235.682

Note: number of observations for daily data: 1,175, number of observations for monthly data: 389.

$$r_t = 100 \log(p_t) - \log(p_{t-1})$$

Table 4. Unconditional Correlation.

Daily correlation (monthly correlation)	Germany	Japan	UK	US
Germany	1.000 (1.000)	0.182 (0.373)	0.655 (0.454)	0.377 (0.427)
Japan		1.000 (1.000)	0.189 (0.376)	0.069 (0.307)
UK			1.000 (1.000)	0.366 (0.525)
US				1.000 (1.000)

Note: number of observations: 1,175 (389), monthly estimates in parentheses.

UK is an exception. In this case, monthly returns exhibit a lower correlation than daily data. This counterintuitive finding could be a result of the use of close-to-close returns because the trading hours of both markets are not synchronous. Figs. 2 and 3 show the evolution of the returns of these series.

Tables 5 and 6 present the results for the correlation estimates for the FDC_z model in its asymmetric specification for daily and monthly data, respectively. Comparing the values $b_{12} + c_{12}$ among the index pairs for daily data show that the persistence of shocks varies considerably among the estimated time-varying correlations. The correlations between Germany and the UK and Japan and the UK have the highest persistence with respect to shocks while the persistence of Germany and the US is very low. For monthly data, the correlations are mainly constant or also have a very low persistence.

In general, there is a very weak evidence of a considerable correlation persistence. In most cases, the persistence of shocks of the correlation processes is much lower than the persistence of shocks to variances. An exception is the daily correlation of Germany and the UK that exhibits a degree of persistence which is comparable with a typical volatility persistence.

Figs. 4–6 show the correlation estimates for the FDC_z model for daily and monthly data, respectively. Note that the finding of constant and non-persistent conditional correlations observed in two cases for daily returns and in five cases for monthly returns estimated with the FDC Model is a new result and can be attributed to the fact that the often used BEKK models cannot reveal such a result due to the structure of the models as explained above.

Table 7 presents results for the asymmetric diagonal BEKK model.

Comparing the parameter estimates for the variance equations (a_{ii}, b_{ii}, c_{ii}, d_{ii}) among the different index pairs shows that the parameter estimates vary substantially for the same return series. For example, the parameter

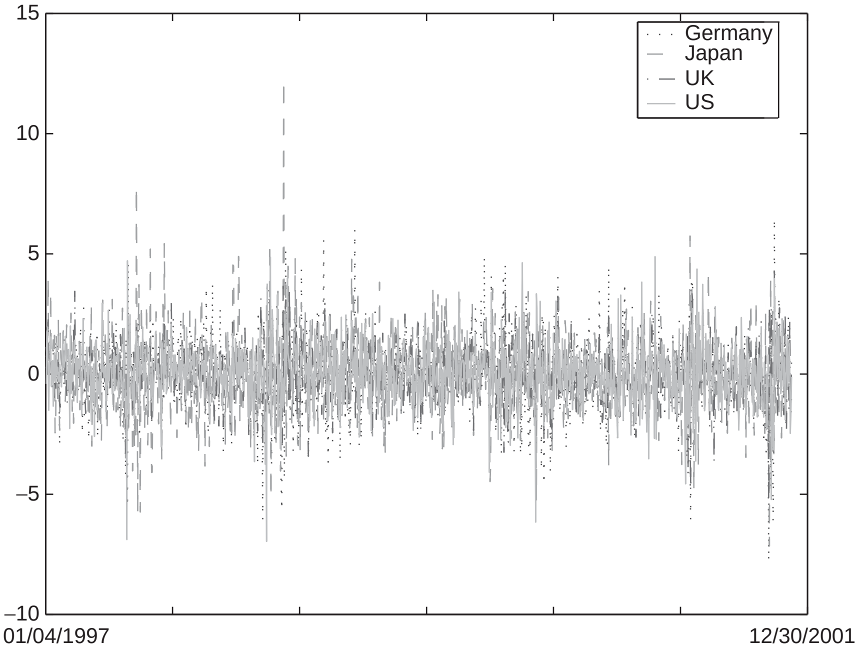


Fig. 2. Daily Returns (Germany Japan, UK, US).

estimates of the UK when estimated with Japan are given by the vector $(0.092, -0.046, 0.883, 0.418)$ and by $(0.401, -0.287, 0.888, -0.098)$ when estimated with the US. These differences are clear evidence that the variance estimates are influenced by the second return series and by the estimated covariance. Thus, parameter estimates are biased. Since the BEKK model does not estimate the correlation process directly but by the ratio of the covariance and the squared root of the product of the variances, I can only analyze the persistence and asymmetry of the variances and the *covariance* which makes a direct comparison of the estimates between the FDC_z model and the diagonal BEKK model infeasible. For this reason, estimation results of the diagonal BEKK model for the monthly data are not reported.

4.1. The Asymmetry of Correlations

Asymmetric effects of volatilities to positive and negative shocks are well documented in the literature and explained with the leverage effect (Black,

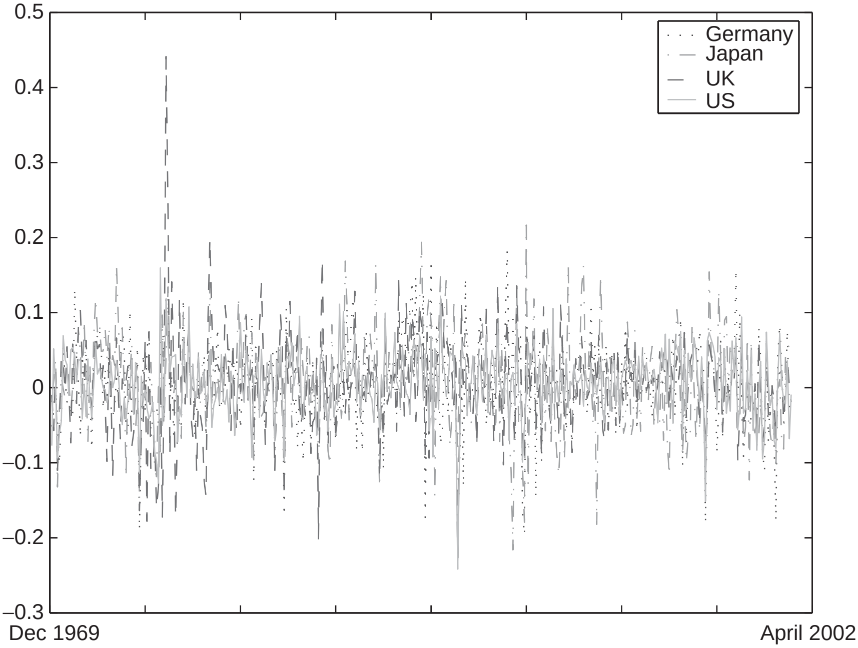


Fig. 3. Monthly Returns (Germany Japan, UK, US).

1976; Christie, 1982) and the volatility feedback effect (Campbell & Hentschel, 1992). However, little is known about the temporal behavior of stock return correlations (see Andersen, Bollerslev, Diebold, & Ebens, 2001; Andersen, Bollerslev, Diebold, & Labys, 2003) and even less of the potential asymmetric effects of positive and negative shocks. Theories explaining asymmetric effects of correlations are in development (e.g. see Connolly & Wang, 2001; Karolyi & Stulz, 2001) but are still rare compared to the studies investigating such effects for volatilities.

The estimation results of the asymmetric FDC_z model show that there is an asymmetric effect of correlations and that this asymmetry is not similar to the one observed for volatilities. This result differs from the findings in the literature where similar asymmetric effects of the conditional covariance are reported (e.g. see Kroner & Ng, 1998). The difference can be explained by the fact that the FDC model analyzes the correlation directly whereas commonly the covariance is examined. Differences in other empirical results such as for the asymmetric DCC model as reported in Cappiello et al. (2002) can also be attributed to the different structure of the FDC model. For

Table 5. Estimation Results of Asymmetric FDC Model (Daily Data).

	GER/JAP	GER/UK	GER/US	JAP/UK	JAP/US	UK/US
Variance estimates						
a_{11}	0.0776	0.0776	0.0776	0.0913*	0.0913*	0.0700*
b_{11}	0.0859***	0.0859***	0.0859***	0.0770***	0.0770***	0.0784***
c_{11}	0.8725***	0.8725***	0.8725***	0.8812***	0.8812***	0.8297***
d_{11}	0.0416	0.0416	0.0416	0.0418	0.0418**	0.0920*
a_{22}	0.0913*	0.0700*	0.0946*	0.0700*	0.0946*	0.0946*
b_{22}	0.0770***	0.0784***	0.0782***	0.0784***	0.0782***	0.0782***
c_{22}	0.8812***	0.8297***	0.8327***	0.8297***	0.8327***	0.8327***
d_{22}	0.0418	0.0920*	0.0891	0.0920*	0.0891	0.0891
Correlation estimates						
a_{12}	0.1877***	0.0394*	0.3609***	0.0684	0.0662***	0.3768***
b_{12}	0.0000	0.0403	0.0000	0.0310	0.0000	0.0000
c_{12}	0.0000	0.9187***	0.0522	0.6297	0.0316	0.0544
d_{12}	0.0306	0.0000	0.0000	0.0000	0.0000	0.0000

Model:

$$h_{11,t} = a_{11} + b_{11}\varepsilon^2_{1,t-1} + c_{11}h_{11,t-1} + d_{11}\eta^2_{1,t-1}$$
$$h_{22,t} = a_{22} + b_{22}\varepsilon^2_{2,t-1} + c_{22}h_{22,t-1} + d_{22}\eta^2_{2,t-1}$$
$$\rho_t = a_{12} + b_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{12}\rho_{t-1} + d_{12}\eta_{1,t-1}\,\eta_{2,t-1}$$

*significant at 10% level.
**significant at 5% level.
***significant at 1% level.

example, an individual asymmetric effect in the DCC model is estimated as the product of two parameter estimates.[18]

To interpret the asymmetric effect, I focus on the parameter estimates for b_{12} and d_{12} in Tables 5 and 6. Results reveal that correlations increase with joint positive shocks (and decrease with shocks of unequal signs) for all country pairs. Whether correlations react differently to joint negative shocks can be assessed by the estimate for d_{12}. Correlations increase more with jointly negative shocks than with jointly positive shocks for Germany and Japan for daily and monthly data and for Germany and the UK, Germany and the US and for Japan and the UK for monthly data. In many cases (five out of six for daily correlations and two out of six for monthly correlations), the correlation does not increase more by joint negative shocks than by joint positive shocks.

The asymmetry of correlations is closely related to the empirical finding that correlations increase in bear markets calling into question the

Table 6. Estimation Results of Asymmetric FDC Model (Monthly Data).

	GER/JAP	GER/UK	GER/US	JAP/UK	JAP/US	UK/US
Variance estimates						
a_{11}	0.0018***	0.0018***	0.0018***	0.0014***	0.0014***	0.0001
b_{11}	0.2064*	0.2064*	0.2064*	0.1040*	0.1040*	0.0849
c_{11}	0.3016*	0.3016*	0.3016*	0.4837***	0.4837***	0.9049***
d_{11}	0.0000	0.0000	0.0000	0.0009	0.0009	0.0000
a_{22}	0.0014***	0.0001	0.0014*	0.0001	0.0014*	0.0014*
b_{22}	0.1040*	0.0849	0.0318	0.0849	0.0318	0.0318
c_{22}	0.4837***	0.9049***	0.0000	0.9049***	0.0000	0.0000
d_{22}	0.0009	0.0000	0.0013**	0.0000	0.0013**	0.0013**
Correlation estimates						
a_{12}	0.3729***	0.5145***	0.4457***	0.2774***	0.2531	0.4036*
b_{12}	0.0010	0.1525*	0.0744	0.0185	0.00100	0.0188
c_{12}	0.0112	0.0000	0.0000	0.3315*	0.2607	0.3832
d_{12}	0.1084*	0.0656	0.0153	0.0794	0.0000	0.0000

Model:

$$h_{11,t} = a_{11} + b_{11}\varepsilon^2_{1,t-1} + c_{11}h_{11,t-1} + d_{11}\eta^2_{1,t-1}$$
$$h_{22,t} = a_{22} + b_{22}\varepsilon^2_{2,t-1} + c_{22}h_{22,t-1} + d_{22}\eta^2_{2,t-1}$$
$$\rho_t = a_{12} + b_{12}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{12}\rho_{t-1} + d_{12}\eta_{1,t-1}\,\eta_{2,t-1}$$

*significant at 10% level.
**significant at 5% level.
***significant at 1% level.

desirability of international portfolio diversification (see De Santis & Gerard, 1997; Longin & Solnik, 1995; Longin & Solnik, 2001; Ng, 2000; Ramchand & Susmel, 1998; Engle & Susmel, 1993). Focussing on negative shocks, we can answer this question by analyzing the parameter estimates for b_{12} and d_{12}. The daily correlations between Germany and the UK and Japan and the UK and all monthly correlations except for Japan and the US replicate the findings in the literature, i.e. international portfolio diversification is not effective when it is needed most since the correlation increases with joint negative shocks. However, the constancy of the daily correlations estimates for Germany and the US, Japan and the US (also monthly) and the UK and the US do not confirm this finding. On the contrary, these results do encourage international portfolio diversification since correlations do not change.

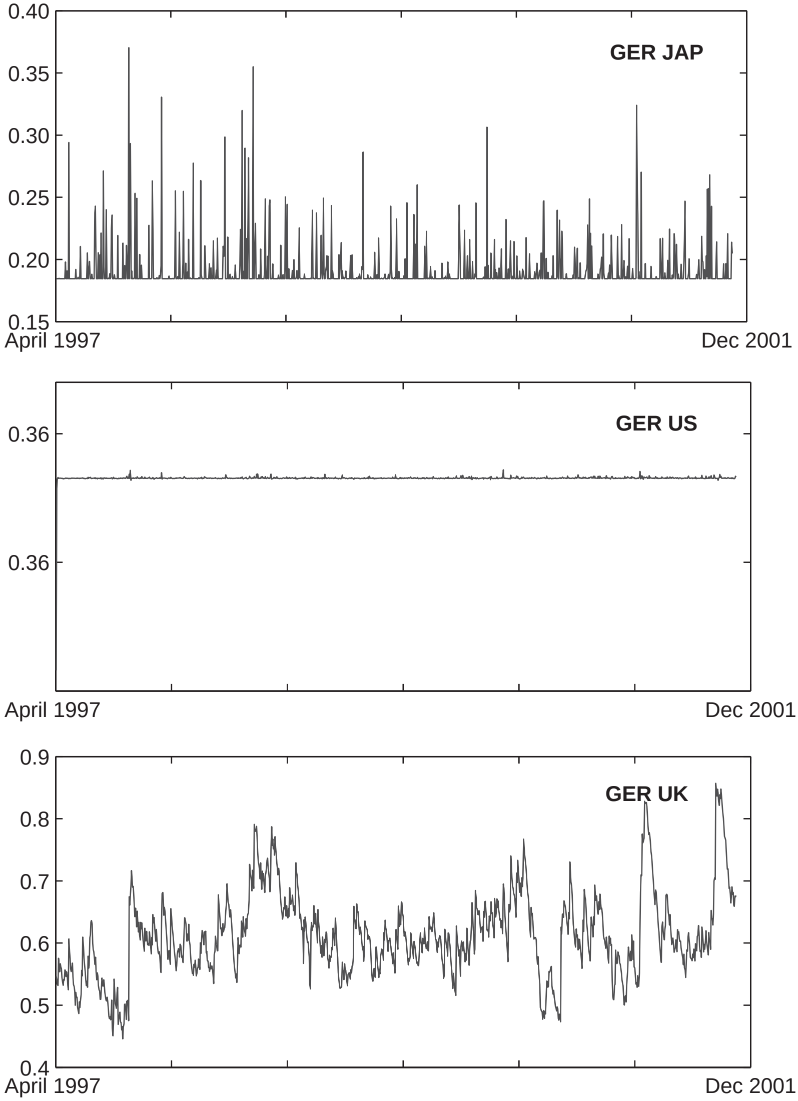


Fig. 4. Correlation Estimates (Daily Data).

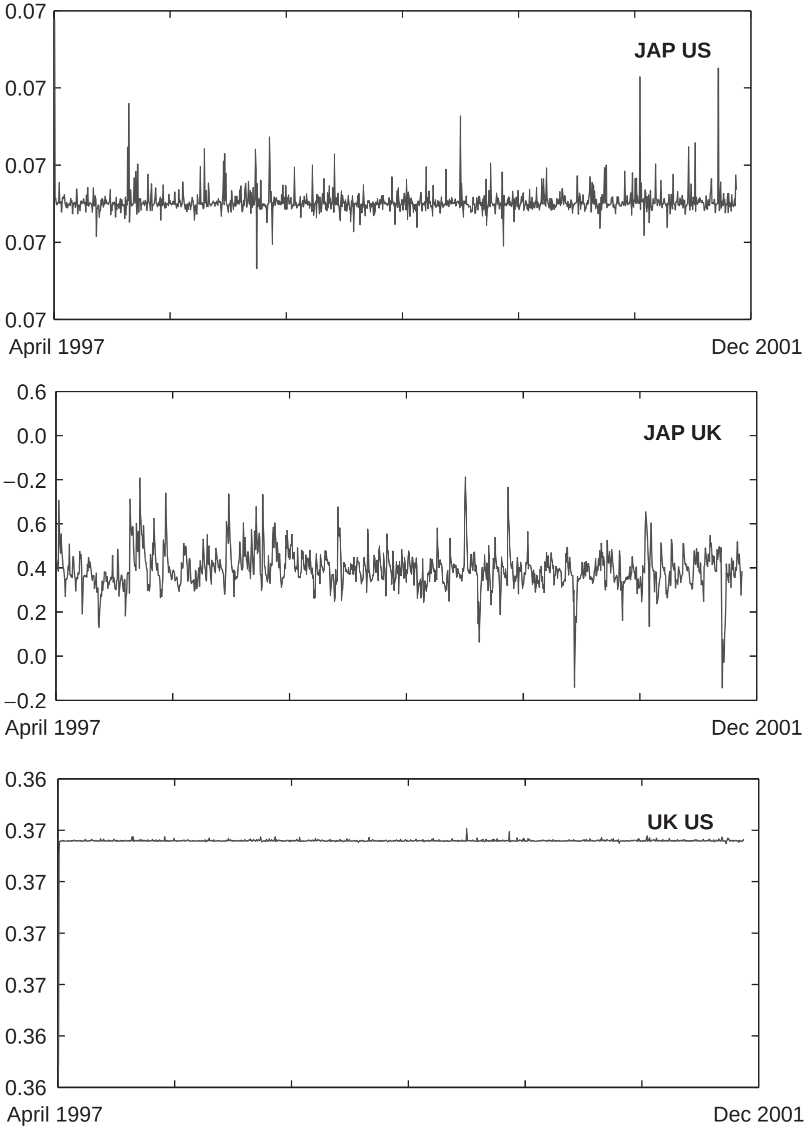


Fig. 5. Correlation Estimates (Daily Data).

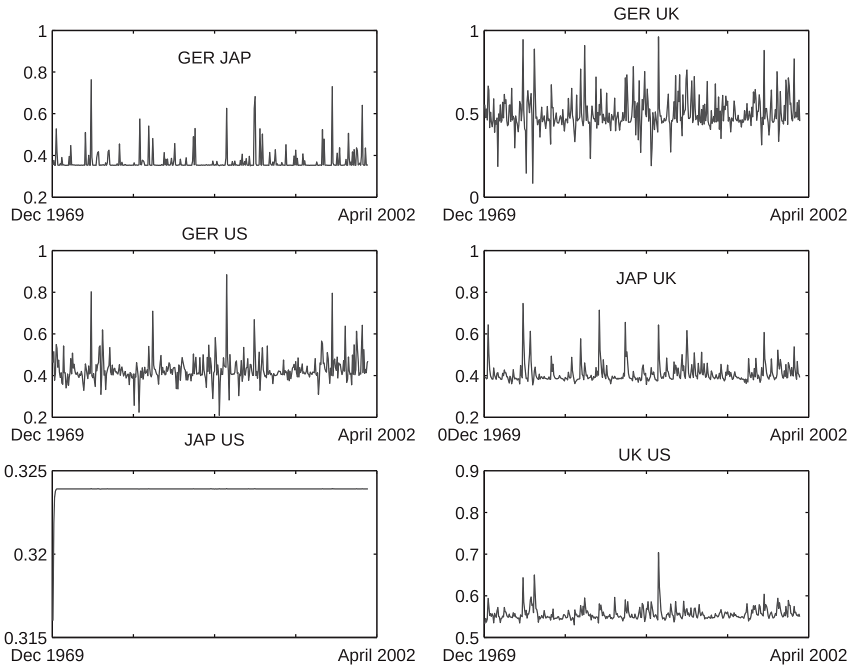


Fig. 6. Correlation Estimates (Monthly Data).

5. CONCLUSIONS

I have demonstrated how the restrictions in multivariate GARCH models to guarantee positive-definite covariance matrices can affect the interpretability of the parameters and the precision of the estimates. The FDC Model introduced here performs clearly better in this regard. Consequently, the empirical results are considerably different from the findings in the literature. I estimated the FDC model for four international stock market indices and found that correlations exhibit a different temporal behavior compared to volatilities, i.e. correlations are more often constant and less persistent than volatilities and the asymmetry of shocks on volatility is more pronounced and more similar among volatilities themselves than the asymmetric effects of jointly positive or negative shocks on correlations.

Future research could be done by studying the impact of the distributional assumptions on the persistence and asymmetry of correlations. In addition,

Table 7. Daily Data – Asymmetric Diagonal BEKK Model.

Estimates	GER/JAP	GER/UK	GER/US	JAP/UK	JAP/US	UK/US
a_{11}	0.457***	0.321***	0.381***	0.408***	0.427***	0.401***
a_{22}	0.106***	0.228***	0.160***	0.092***	0.041	0.304***
b_{11}	0.266***	0.000	0.172***	0.245***	0.297***	−0.287***
b_{22}	−0.050	−0.097	−0.063*	−0.046	−0.067 ***	−0.133*
c_{11}	0.912***	0.930***	0.924***	0.923***	0.918***	0.888***
c_{22}	0.902***	0.929***	0.877***	0.883***	0.869***	0.843***
d_{11}	0.217*	0.332***	0.330***	0.248***	0.105**	−0.098
d_{22}	0.425***	0.334***	0.501***	0.418***	0.511***	0.468***

Covariance equations:

$$h_{11,t} = a_{11}^2 + b_{11}^2 \varepsilon_{1,t-1}^2 + c_{11}^2 h_{11,t-1} + d_{11}^2 \eta_{1,t-1}^2$$
$$h_{22,t} = a_{22}^2 + b_{22}^2 \varepsilon_{2,t-1}^2 + c_{22}^2 h_{22,t-1} + d_{22}^2 \eta_{2,t-1}^2$$
$$h_{12,t} = h_{21,t} = a_{11}a_{22} + b_{11}b_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + c_{11}c_{22}h_{12,t-1} + d_{11}d_{22}\eta_{1,t-1}\,\eta_{2,t-1}$$

*significant at 10% level.
**significant at 5% level.
***significant at 1% level.

although mentioned, further research is necessary to answer the question as to how correlation estimates change with the specification of the volatility equations and vice versa.

NOTES

1. Restricting the BEKK model to be diagonal reduces the number of parameters that must be estimated. The factor GARCH model (Engle, Ng, & Rothschild, 1990) reduces the number of parameters and can be transformed to a BEKK model.
2. Note that the nested ADC model requires further restrictions to guarantee a positive-definite covariance matrix.
3. I subsequently assume a simple mean equation as given by Eq. (1) and do exclusively focus on the covariance matrix.
4. A positive-definite covariance matrix would imply that the determinant of

$$\mathbf{H}_t = \begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{12,t} & h_{22,t} \end{pmatrix}$$

is positive. That means that $h_{11,t}$ and $h_{22,t} > h^2_{12,t}$ which is not guaranteed since the parameters a_{ij} and b_{ij} are freely estimated for all $i, j = 1, 2$.

5. The multivariate factor GARCH model will not be presented here since it can be derived from a full BEKK model (see Kroner & Ng, 1998).

6. This model was first presented in 1998 and is therefore described before the DCC model of Engle (2002).

7. To guarantee positive variances I use the variance equations of the diagonal BEKK model for the variance equations of the CCM as suggested by Kroner and Ng (1998).

8. Ang and Chen (2002) report misspecifications of the asymmetric effect in an asymmetric GARCH-M model without being specific in pointing to this problem.

9. In most multivariate GARCH models conditional correlations are derived from the ratio of the covariance and the product of the roots of the conditional variances (see Eqs. (6) and (8)).

10. The conditional variances are specified as a simple GARCH(1,1) model. However, the FDC model can be estimated with any other specification of the conditional variances, e.g. a EGARCH model.

11. The standardized residuals are given by $z_t = \varepsilon_t / \sigma_t$.

12. In other words, are the estimates of the parameters in the conditional-variance estimates robust with respect to the constant correlation assumption? (p. 109).

13. Engle compares the DCC model with the scalar BEKK, the diagonal BEKK, a moving average process, an exponential smoother and a principle components GARCH.

14. See discussion of the diagonal BEKK model above. In this case, the high persistence of the correlation process mainly contributes to this result.

15. Tse (2000) proposed an analysis whether or to which extent correlation estimates improve or change variance estimates.

16. A non-trading day means that no information is processed in the market. Consequently, returns are zero.

17. Existing methods to synchronize the data are only recently developed (see Engle, Burns, & Mezrich, 1998; Forbes & Rigobon, 2002; Martens & Poon, 2001).

18. See Capiello et al. (2002, Eq. (5), p. 8). Of course, differences can also accrue due to different data samples.

ACKNOWLEDGMENTS

The author thanks three anonymous referees for helpful comments and conference participants of the 10th Conference on the Globalization of Securities Markets in Kaohsiung in December 2001 and of the Econometric Society's North American Summer Meeting in Los Angeles in June 2002.

REFERENCES

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.

Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Review of Financial Studies*, *63*(3), 443–494.

Baba, Y., Engle, R. F., Kraft, D., & Kroner, K. F. (1991). *Multivariate simultaneous generalized ARCH*. MS. Thesis University of California, Department of Economics, San Diego.

Bauwens, L., Laurent, S., & Rombouts, J. V. K. (2005). Multivariate Garch models: A survey. *Journal of Applied Econometrics*, forthcoming.

Bekaert, G., & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *The Review of Financial Studies*, *13*, 1–42.

Berndt, E., Hall, B., Hall, R., & Hausman, J. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, *3/4*, 653–665.

Black, F. (1976). Studies of stock market volatility changes. *Proceeding of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, *72*, 1155–1180.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, *96*, 116–131.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*, 143–172.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, *31*, 281–318.

Cappiello, L., Engle, R. F., Sheppard, K. (2002). *Asymmetric dynamics in the correlations of global equity and bond returns*. Unpublished working paper, European Central Bank.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, *10*, 407–432.

Connolly, R., Wang, F.-K. A. (2001). *On stock market return co-movements: Macroeconomic news, dispersion of beliefs, and contagion*? mimeo.

De Santis, G., & Gerard, B. (1997). International asset pricing and portfolio diversification with time-varying risk. *Journal of Finance*, *LII*(5), 1881–1912.

Engle, R. F. (2002). Dynamic conditional correlations – A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, *20*(3), 339–350.

Engle, R. F., Burns, P., & Mezrich, J. (1998). Correlations and volatilities of asynchronous data. *Journal of Derivatives*, Summer, 1–12.

Engle, R. F., & Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*, 1749–1778.

Engle, R. F., Ng, V., & Rothschild, M. (1990). Asset pricing with a factor ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics*, *45*, 213–238.

Engle, R. F., & Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH*. NBER Working Paper W8554.

Engle, R. F., & Susmel, R. (1993). Common volatility in international equity markets. *Journal of Business and Economic Statistics*, *11*(2), 167–176.

Forbes, K., & Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market co-movements. *Journal of Finance*, *57*(5), 2223–2261.

Glosten, L. R., Jagannathan, R., & Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*, 1779–1801.

Karolyi, G. A., & Stulz, R. M. (1996). Why do markets move together? An investigation of US–Japan stock return comovements. *Journal of Finance*, *51*(3), 951–986.

Karolyi, G. A., & Stulz, R. M. (2001). Are financial asset prices locally or globally? In: Constantinides, Harris & Stulz (Eds), *The handbook of economics of finance*. Amsterdam: North-Holland.

Kroner, K. F., & Ng, V. K. (1998). Modeling asymmetric comovements of asset returns. *The Review of Financial Studies*, *11*(4), 817–844.

Longin, F., & Solnik, B. (1995). Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance*, *14*(1), 3–26.

Longin, F., & Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, *56*, 651–678.

Martens, M., & Poon, S.-H. (2001). Returns synchronization and daily correlation dynamics between international stock markets. *Journal of Banking and Finance*, *25*, 1805–1827.

Ng, A. (2000). Volatility spillover effects from Japan and the US to the Pacific-Basin. *Journal of International Money and Finance*, *19*, 207–233.

Pagan, A., & Schwert, G. W. (1990). Alternative models of conditional stock volatility. *Journal of Econometrics*, *45*, 267–290.

Ramchand, L., & Susmel, R. (1998). Volatility and cross correlation across major stock markets. *Journal of Empirical Finance*, *5*, 397–416.

Tsay, R. S. (2002). *Analysis of financial time series*, (2nd ed.). Wiley Series in Probability and Statistics. New York: Wiley.

Tse, Y. K. (2000). A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics*, *98*(1), 107–127.

Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, *20*(3), 351–362.

This page intentionally left blank

A MULTIVARIATE SKEW-GARCH MODEL

Giovanni De Luca, Marc G. Genton and Nicola Loperfido

ABSTRACT

Empirical research on European stock markets has shown that they behave differently according to the performance of the leading financial market identified as the US market. A positive sign is viewed as good news in the international financial markets, a negative sign means, conversely, bad news. As a result, we assume that European stock market returns are affected by endogenous and exogenous shocks. The former raise in the market itself, the latter come from the US market, because of its most influential role in the world. Under standard assumptions, the distribution of the European market index returns conditionally on the sign of the one-day lagged US return is skew-normal. The resulting model is denoted Skew-GARCH. We study the properties of this new model and illustrate its application to time-series data from three European financial markets.

1. INTRODUCTION

The pioneering work of Engle (1982) has represented the starting point of a tremendous scientific production with the aim of modeling and forecasting

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 33–57

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20002-6

the volatility of financial time series. The AutoRegressive Conditional Het-eroskedasticity (ARCH) model has been dealt with in depth. Many variants have been proposed. Among them, we emphasize its most popular generalizations, the Generalized ARCH model (Bollerslev, 1986) and the Exponential GARCH model (Nelson, 1991) allowing for the inclusion of the asymmetric effect of volatility. Moving from a univariate to a multivariate perspective, the multivariate GARCH model is quite interesting because it can shed light on the common movements of the volatilities across markets (Bollerslev, Engle, & Wooldridge, 1988; Bollerslev, 1990; Engle, 2002).

When the analysis focuses on one or more markets, the possible relevance of an external leading market is usually ignored. Nonetheless, it is an important point which can help explaining some empirically detected features. Actually, a wide literature has dealt with the issue of the international transmission of stock markets movements. Eun and Shim (1989) stressed the most influential role of the US stock market. Innovations in the US market are transmitted to the other markets. Conversely, none of the other markets can affect the US market movements. The time horizon of the transmission is very short: the other stock exchanges' responses rapidly decrease after one day. The conclusion of Eun and Shim (1989) suggests to consider the US market as *the most important producer of information affecting the world stock market.* The contemporaneous and lead/lag relationships among stock markets are also studied in Koch and Koch (1991).

The analysis of univariate and multivariate GARCH models has traditionally neglected this aspect, with few exceptions. For instance, Lin, Engle, and Ito (1994) carried out an empirical investigation of the relationship between returns and volatilities of Tokyo and New York markets. The peculiarity of the study is the use of a decomposition of daily return into two components: the daytime return and the overnight return. The conclusion is the existence of cross-market interdependence in returns and volatilities. Karolyi (1995) detected the interdependence between the US and Canadian markets through a bivariate GARCH model.

Our analysis starts from the empirical detection of the different behavior of three European markets, according to the performance of the leading market identified as the US market. From a statistical perspective, we assume that European stock market returns are affected by endogenous and exogenous shocks. The former raise in the market itself, the latter come from the US market, defined as the leading market because of its most influential role in the world. Moreover, the flow of information from the US market to the European markets is asymmetric in its direction as well as in its effects. We recognize a negative (positive) performance of the US market as a proxy for bad (good) news for the world stock market. European financial markets returns behave in a different way according to bad or good news. Moreover, they are more reactive to bad news than to good ones.

The European and the US markets are not synchronous. When European markets open on day t, the US market is still closed; when European markets close on the same day, the US market is about to open or has just opened (depending on the European country). This implies a possible causal relationship from the US market return at time $t-1$ to the European returns at time t.

The distribution of European returns changes according to the sign of the one-day lagged performance of the US market. Average returns are negative (positive) in presence of bad (good) news and they are very similar in absolute value. Volatility is higher in presence of bad news. Skewness is negative (positive) and more remarkable in presence of bad (good) news. In both cases, a high degree of leptokurtosis is observed. Finally, bad news involves a stronger correlation between present European returns and the one-day lagged US return.

Allowing for a GARCH structure for taking into account the heteroskedastic nature of financial time series, under standard assumptions, the distribution of the European returns conditionally on news (that is, on the sign of the one-day lagged US return) and past information turns out to be skew-normal (Azzalini, 1985). This is a generalization of the normal distribution with an additional parameter to control skewness. The two conditional distributions are characterized by different features according to the type of news (bad or good). In particular, the skewness can be either negative (bad news) or positive (good news). The resulting model is denoted Skew-GARCH (henceforth SGARCH). The theoretical features of the model perfectly match the empirical evidence.

The basic idea can be extended to a multivariate setting. The international integration of financial markets is more remarkable in presence of a geographical proximity. The European markets tend to show common movements. Under standard assumptions, the joint distribution of European stock market returns conditionally on the sign of the one-day lagged US market return and past information is a multivariate skew-normal distribution (Azzalini & Dalla Valle, 1996), whose density is indexed by a location vector, a scale matrix and a shape vector. Finally, unconditional (with respect to the performance of the US market) returns have some features in concordance with empirical evidence.

The paper is organized as follows. Section 2 describes the theory of the skew-normal distribution. In Section 3, the multivariate SGARCH model is presented. In Section 4, the conditional distribution and related moments are obtained. Some special cases are described, including the univariate model when the dimensionality reduces to one. Section 5 refers to the unconditional distribution. Section 6 exhibits the estimates of the univariate and multivariate models applied to three small financial markets in Europe: Dutch, Swiss and Italian. The results show the relevance of the performance of the leading market supporting the proposal of the SGARCH model. Section 7 concludes. Some proofs are presented in the appendix.

2. THE SKEW-NORMAL DISTRIBUTION

The distribution of a random vector z is multivariate skew-normal (*SN*, henceforth) with location parameter ζ, scale parameter Ω and shape parameter α, that is $z \sim SN_p(\zeta, \Omega, \alpha)$, if its probability density function (pdf) is

$$f(z; \zeta, \Omega, \alpha) = 2\phi_p(z - \zeta; \Omega)\Phi\left[\alpha^{\mathrm{T}}(z - \zeta)\right], \quad z, \zeta, \alpha \in \mathbb{R}^p,\ \Omega \in \mathbb{R}^{p \times p}$$

where $\Phi(\cdot)$ is the cdf of a standardized normal variable and $\phi_p(z - \zeta; \Omega)$ is the density function of a p-dimensional normal distribution with mean ζ and variance Ω. For example, $Z \sim SN_1(\zeta, \omega, \alpha)$ denotes a random variable whose distribution is univariate *SN* with pdf

$$f(z; \zeta, \omega, \alpha) = \frac{2}{\sqrt{\omega}}\phi\left(\frac{z - \zeta}{\sqrt{\omega}}\right)\Phi[\alpha(z - \zeta)]$$

where $\phi(\cdot)$ is the pdf of a standard normal variable.

Despite the presence of an additional parameter, the *SN* distribution resembles the normal one in several ways, formalized through the following properties:

Inclusion: The normal distribution is an *SN* distribution with shape parameter equal to zero:

$$z \sim SN_p(\zeta, \Omega, 0) \Leftrightarrow z \sim N_p(\zeta, \Omega)$$

Greater norms of α imply greater differences between the density of the multivariate $SN_p(\zeta, \Omega, \alpha)$ and the density of the multivariate $N_p(\zeta, \Omega)$.

Linearity: The class of *SN* distributions is closed with respect to linear transformations. If A is a $k \times p$ matrix and $\mathrm{b} \in \mathbb{R}^k$, then

$$z \sim SN_p(\zeta, \Omega, \alpha) \Rightarrow Az + \mathrm{b} \sim SN_k\left(A\zeta + b, A\Omega A^{\mathrm{T}}, \bar{\alpha}\right)$$

$$\bar{\alpha} = \frac{\left(A\Omega A^{\mathrm{T}}\right)^{-1} A\Omega\alpha}{\sqrt{1 + \alpha^{\mathrm{T}}\Omega\left(\Omega^{-1} - A^{\mathrm{T}}\left(A\Omega A^{\mathrm{T}}\right)^{-1} A\right)\Omega\alpha}}$$

It follows that the *SN* class is closed under marginalization: subvectors of *SN* vectors are *SN*, too. In particular, each component of an *SN* random vector is univariate *SN*.

Invariance: The matrix of squares and products $(\mathrm{z} - \zeta)(\mathrm{z} - \zeta)^{\mathrm{T}}$ has a Wishart distribution:

$$(\mathrm{z} - \zeta)(\mathrm{z} - \zeta)^{\mathrm{T}} \sim W(\Omega, 1)$$

Notice that the distribution of $(\mathrm{z} - \zeta)(\mathrm{z} - \zeta)^{\mathrm{T}}$ does not depend on the shape parameter. In the univariate case, it means that $Z \sim SN_1(\zeta, \omega, \alpha)$ implies $(Z - \zeta)^2/\omega \sim \chi_1^2$.

All moments of the *SN* distribution exist and are finite. They have a simple analytical form. However, moments of the *SN* distribution differ from the normal ones in several ways:

- Location and scale parameters equal mean and variance only if the shape parameter vector α equals zero.
- Moments are more conveniently represented through the parameter $\delta = \Omega\alpha/\sqrt{1 + \alpha^{\mathrm{T}}\Omega\alpha}$.
- Tails of the *SN* distribution are always heavier than the normal ones, when the shape parameter vector α differs from zero.

Table 1 reports the expectation, variance, skewness and kurtosis of the *SN* distribution, in the multivariate and univariate cases.

Multivariate skewness and kurtosis are evaluated through Mardia's indices (Mardia, 1970). Notice that in the univariate case Mardia's index of kurtosis equals the fourth moment of the standardized random variable. On the other hand, in the univariate case, Mardia's index of skewness equals the square of the third moment of the standardized random variable. In the following sections, when dealing with skewness, we shall refer to Mardia's index in the multivariate case and to the third moment of the standardized random variable in the univariate case.

Table 1. Expectation, Variance, Skewness and Kurtosis of the *SN* Distribution.

	$z \sim SN_p(\zeta, \Omega, \alpha)$	$Z \sim SN_1(\zeta, \omega, \alpha)$
Expectation	$\zeta + \sqrt{\frac{2}{\pi}}\delta$	$\zeta + \sqrt{\frac{2}{\pi}}\delta$
Variance	$\Omega - \frac{2}{\pi}\delta\delta^{\mathrm{T}}$	$\omega - \frac{2}{\pi}\delta^2$
Skewness	$2(4-\pi)^2\left(\frac{\delta^{\mathrm{T}}\Omega^{-1}\delta}{\pi - 2\delta^{\mathrm{T}}\Omega^{-1}\delta}\right)^3$	$\sqrt{2}(4-\pi)\left(\frac{\delta}{\sqrt{\pi\omega - 2\delta^2}}\right)^3$
Kurtosis	$8(\pi-3)\left(\frac{\delta^{\mathrm{T}}\Omega^{-1}\delta}{\pi - 2\delta^{\mathrm{T}}\Omega^{-1}\delta}\right)^2$	$8(\pi-3)\left(\frac{\delta^2}{\pi\omega - 2\delta^2}\right)^2$

3. THE SGARCH MODEL

Let Y_t be the *leading* (US) market return at time t. A simple GARCH(p,q) model is assumed. Then we can write

$$\begin{aligned} Y_t &= \eta_t \varepsilon_t \\ \eta_t^2 &= \delta_0 + \sum_{i=1}^{q} \delta_i (\eta_{t-i}\varepsilon_{t-i})^2 + \sum_{j=q+1}^{q+p} \delta_j \eta_{t+q-j}^2 \end{aligned} \tag{1}$$

where $\{\varepsilon_t\}$~i.i.d. N(0,1). $\{\varepsilon_t\}$ is the innovation (or shock) of the US market and is hypothesized to be Gaussian. In order to ensure the positivity of η_t^2,δ_0 has to be positive and the remaining parameters in (1) non-negative. After denoting $Z^{t-1} = [z_{t-1}, z_{t-2}, \ldots]$, it turns out that

$$Y_t | Y^{t-1} \sim N(0, \eta_t^2)$$

Let x_t be the $p \times 1$ return vector of the European markets at time t. We assume that returns at time t depend both on an endogenous (local) shock and an exogenous (global) shock. The endogenous shocks do have relationships with each other (common movements are usually observed in neighboring markets). The $p \times 1$ local shock vector is denoted ζ_t. The exogenous or global shock is an event that has an influence across more markets. For the European markets we identify the global shock as the innovation of the US market one-day before, that is ε_{t-1}. The lag is due to the mentioned nonsynchronicity of the markets.

The function $\mathrm{f}(\varepsilon_{t-1})$, specified below, describes the relationship between the return vector x_t and ε_{t-1}.

The local and the global shocks are assumed to be independent and to have a joint $(p+1)-$ dimensional normal distribution,

$$\begin{pmatrix} \varepsilon_{t-1} \\ \zeta_t \end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0^T \\ 0 & \Psi \end{pmatrix} \right) \tag{2}$$

where ψ is a correlation matrix with generic off-diagonal entry ρ_{ij}. If the hypothesis of Gaussianity is far from true for returns, it appears to be consistent for shocks, even if some authors propose more general distributions (e.g. Bollerslev, 1987). Moreover, we assume that the variances have a multivariate GARCH structure. We can then write

$$\mathrm{x}_t = D_t[\mathrm{f}(\varepsilon_{t-1}) + \zeta_t] \tag{3}$$

$$\mathrm{f}(\varepsilon_{t-1}) = \sqrt{\frac{2}{\pi}}\gamma + \beta\varepsilon_{t-1} - \gamma|\varepsilon_{t-1}| \tag{4}$$

where

$$D_t = \begin{pmatrix} \sigma_{1t} & 0 & & 0 \\ 0 & \sigma_{2t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & & 0 & \sigma_{pt} \end{pmatrix} \tag{5}$$

$$\sigma_{kt}^2 = \omega_{0k} + \sum_{i=1}^{q} \omega_{ik}\left(\sigma_{k,t-i}\zeta_{k,t-i}\right)^2 + \sum_{j=q+1}^{q+p} \omega_{jk}\sigma_{k,t+q-j}^2 + \omega_{q+p+1,k}\eta_{t-1}^2 \tag{6}$$

The last term takes into account the possible volatility spillover from the US to the European markets. If at time $t-1$ the US stock exchange is closed, then $\eta_{t-1}^2 = \eta_{t-2}^2$ The positivity of σ_{kt}^2 is ensured by the usual constraints on the parameters.

Assumptions (2–6) compound the SGARCH model. The function $\mathrm{f}(\varepsilon_{t-1})$ models the effect of the exogenous shock ε_{t-1} on the vector x_t and $\{\zeta_t\}$ is a sequence of serially independent random vectors. The parameter vectors β and γ are constrained to be non-negative. Moreover, $\beta \geq \gamma \geq 0$. The former describes the direct effect of the past US innovations on x_t, the latter the feedback effect. Volatility feedback theory (Campbell & Hentschel, 1992) implies that news increases volatility, which in turn lowers returns. Hence the direct effect of good (bad) news is mitigated (strengthened) by the feedback effect. A point in favor of the SGARCH model is the formalization of the two effects.

The conditional distribution of the return vector is

$$\mathrm{x}_t|I_{t-1} \sim N_p(D_t\mathrm{f}(\varepsilon_{t-1}), D_t\Psi D_t)$$

where I_{t-1} denotes the information at time $t-1$.

The SGARCH model does not involve a conditional null mean vector. Instead, the mean does depend on the volatility. Moreover, returns are more reactive to bad news than good news. In fact,

$$\left.\frac{\partial \mathrm{x}_t}{\partial \varepsilon_{t-1}}\right|_{\varepsilon_{t-1}>0} = D_t(\beta - \gamma)$$

$$\left.\frac{\partial \mathrm{x}_t}{\partial \varepsilon_{t-1}}\right|_{\varepsilon_{t-1}<0} = D_t(\beta + \gamma)$$

4. CONDITIONAL DISTRIBUTIONS AND RELATED MOMENTS

We are interested in the p-variate distribution of x_t conditional on D_t and on news from the US market, that is on the sign of Y_{t-1}.

Theorem 1. Under the SGARCH model's assumptions, the following distributions hold:

Good news

$$\left(D_t^{-1}\mathrm{x}_t|\,Y_{t-1}>0\right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_+, \alpha_+\right)$$

where

$$\Omega_+ = \Psi + \delta_+\delta_+^{\mathrm{T}}, \quad \alpha_+ = \frac{\Psi^{-1}\delta_+}{\sqrt{1+\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}} \quad \text{and} \quad \delta_+ = \beta - \gamma$$

Bad news

$$\left(D_t^{-1}\mathrm{x}_t|\,Y_{t-1}<0\right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_-, \alpha_-\right)$$

where

$$\Omega_{-} = \Psi + \delta_{-}\delta_{-}^{\mathrm{T}}, \quad \alpha_{-} = \frac{-\Psi^{-1}\delta_{-}}{\sqrt{1+\delta_{-}^{\mathrm{T}}\Psi^{-1}\delta_{-}}} \text{ and } \quad \delta_{-} = \beta + \gamma$$

The proof is given in the appendix.

Applying the results of linear transformations of a multivariate *SN* distribution, we can obtain the conditional distributions of returns $(\mathrm{x}_t|D_t, Y_{t-1}>0)$ and $(\mathrm{x}_t|D_t, Y_{t-1}<0)$. Their formal properties are found to be consistent with empirical findings. In more detail:

Expectation: On average, good (bad) news determines positive (negative) returns. The effect of both kinds of news is equal in absolute value. Moreover, it is proportional to the direct effect (modeled by the parameter β) mitigated or amplified by the volatility structure (modeled by the diagonal matrix D_t),

$$E(\mathrm{x}_t|D_t, Y_{t-1}>0) = \sqrt{\frac{2}{\pi}}D_t\beta$$

$$E(\mathrm{x}_t|D_t, Y_{t-1}<0) = -\sqrt{\frac{2}{\pi}}D_t\beta$$

This result could be interpreted as the presence of an arbitrage opportunity in the market with implications regarding market efficiency. Actually, our analysis is carried out using returns computed from two close prices (close-to-close). During the time from the closing to the opening of a European exchange, there is the closing of the US exchange and its effect is reflected mainly in the open prices of the European exchanges. If we used open-to-close returns, this apparent arbitrage opportunity would disappear.

Variance: Variance is higher (lower) in the presence of bad (good) news. More precisely, the elements on the main diagonal of the covariance matrix are greater (smaller) when previous day's US market returns were negative (positive)

$$V(x_t|D_t, Y_{t-1}>0) = D_t\left(\Psi + \frac{\pi-2}{\pi}\delta_{+}\delta_{+}^{\mathrm{T}}\right)D_t$$

$$V(\mathrm{x}_t|D_t, Y_{t-1}<0) = D_t\left(\Psi + \frac{\pi-2}{\pi}\delta_{-}\delta_{-}^{\mathrm{T}}\right)D_t$$

Skewness: Symmetry of conditional returns would imply that news from the US are irrelevant. In this framework, univariate skewness is negative (positive) in presence of bad (good) news and higher (smaller) in absolute value. On the other hand, multivariate skewness is always positive but its level cannot be related to the kind of news. Skewness of x_t when $Y_{t-1}>0$ can be either lower or higher than skewness of x_t when $Y_{t-1}<0$, depending on the parameters. The two indices are

$$S(\mathrm{x}_t|D_t, Y_{t-1}>0) = 2(4-\pi)^2\left(\frac{\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}{\pi+(\pi-2)\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}\right)^3$$

$$S(\mathrm{x}_t|D_t, Y_{t-1}<0) = 2(4-\pi)^2\left(\frac{\delta_-^{\mathrm{T}}\Psi^{-1}\delta_-}{\pi+(\pi-2)\delta_-^{\mathrm{T}}\Psi^{-1}\delta_-}\right)^3$$

Kurtosis: In the SGARCH model, relevant news (good or bad) always lead to leptokurtotic returns. Again, there is no relationship between the kind of news (good or bad) and multivariate kurtosis (high or low). However, SGARCH models imply that a higher kurtosis is related to a higher skewness. The two indices are

$$K(\mathrm{x}_t|D_t, Y_{t-1}>0) = 8(\pi-3)\left(\frac{\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}{\pi+(\pi-2)\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}\right)^2$$

$$K(\mathrm{x}_t|D_t, Y_{t-1}<0) = 8(\pi-3)\left(\frac{\delta_-^{\mathrm{T}}\Psi^{-1}\delta_-}{\pi+(\pi-2)\delta_-^{\mathrm{T}}\Psi^{-1}\delta_-}\right)^2$$

Correlation: Let ρ_+ (ρ_-) be the $p\times 1$ vector whose ith component ρ_{i+} (ρ_{i-}) is the correlation coefficient between the ith European return X_{it} and the previous day's US return Y_{t-1} conditionally on good (bad) news, that is $Y_{t-1}>0$, ($Y_{t-1}<0$), and volatility D_t. The correlation of $X_{it}, Y_{t-1}|D_t, Y_{t-1}>0$ ($X_{it}, Y_t|D_t, Y_{t-1}<0$) is the same under the multivariate and univariate SGARCH model (the former being a multivariate generalization of the latter). Hence, we can recall a property of the univariate SGARCH model (De Luca & Loperfido, 2004) and write

$$\rho_{i+} = \frac{(\beta_i-\gamma_i)\sqrt{\pi-2}}{\sqrt{\pi+(\pi-2)(\beta_i-\gamma_i)^2}}, \quad \rho_{i-} = \frac{(\beta_i+\gamma_i)\sqrt{\pi-2}}{\sqrt{\pi+(\pi-2)(\beta_i+\gamma_i)^2}}$$

for $i=1,\ldots,p$. Implications of the above results are twofold. In the first place, ρ_+ and $\rho-$ are functions of $\delta_+=\beta-\gamma$ and $\delta_-=\beta+\gamma$, respectively. In the second place, a little algebra leads to the inequalities $0\le\rho_+\le\rho_-\le 1_p$, where 1_p is a p-dimensional vector of ones. It follows that bad news strengthen the association between European returns and previous day's US returns.

The two p-variate distributions in Theorem 1 have different characteristics according to the news coming from the US market. As a result, the conditioning appears to be relevant.

It is interesting to consider some special cases. Firstly, if the parameter vector γ is zero, then $\delta_+ = \delta_- = \beta$ and $\Omega_+ = \Omega_- = \Omega$. The two conditional distributions in Theorem 1 differ for the shape parameter which has the same absolute value but a different sign, that is

$$\left(D_t^{-1}\mathrm{x}_t | Y_{t-1} > 0\right) \sim SN_p(0, \Omega, \alpha)$$

where

$$\Omega = \Psi + \beta\beta^{\mathrm{T}}, \quad \alpha = \frac{\Psi^{-1}\beta}{\sqrt{1 + \beta^{\mathrm{T}}\Psi^{-1}\beta}}$$

and

$$\left(D_t^{-1}\mathrm{x}_t | Y_{t-1} < 0\right) \sim SN_p(0, \Omega, -\alpha).$$

In this case, no feedback effect exists. However, it still makes sense to condition on the type of news and to introduce *SN* distributions.

Secondly, if the parameter vector β is zero (implying that also γ is zero), there is no evidence of any (direct or feedback) effect of the US news on European returns. The two markets are independent. The *SN* distributions in Theorem 1 turn out to have a zero shape parameter which shrinks them to the same normal distribution. As a result, the multivariate skewness and kurtosis indices also shrink to zero.

Finally, if the dimensionality parameter p equals one, that is if we move from a multivariate framework to a univariate perspective, the multivariate SGARCH model equals the univariate SGARCH model (De Luca & Loperfido, 2004). In this case, denoting by X_t a European return at time t, we have

$$\frac{X_t}{\sigma_t} = \sqrt{\frac{2}{\pi}}\gamma + \beta\varepsilon_{t-1} - \gamma|\varepsilon_{t-1}| + \zeta_t$$

where β and γ are now scalars and ζ_t is a unidimensional random variable. The main features of the model in a univariate context (parameters of the distribution of X_t/σ_t given the sign of Y_{t-1} and moments of X_t given σ_t and the sign of Y_{t-1}) are summarized in Table 2.

Table 2. Features of the Univariate SGARCH Model.

	Good News ($Y_{t-1} > 0$)	Bad News ($Y_{t-1} < 0$)	Overall
Location	$\sqrt{\frac{2}{\pi}}\gamma$	$\sqrt{\frac{2}{\pi}}\gamma$	—
Scale	$\sqrt{1+(\beta-\gamma)^2}$	$\sqrt{1+(\beta+\gamma)^2}$	—
Shape	$\frac{\beta-\gamma}{\sqrt{1+(\beta-\gamma)^2}}$	$\frac{-(\beta+\gamma)}{\sqrt{1+(\beta+\gamma)^2}}$	—
Expectation	$\sqrt{\frac{2}{\pi}}\beta\sigma_t$	$-\sqrt{\frac{2}{\pi}}\beta\sigma_t$	0
Variance	$\sigma_t^2\left[1+\frac{\pi-2}{\pi}(\beta-\gamma)^2\right]$	$\sigma_t^2\left[1+\frac{\pi-2}{\pi}(\beta+\gamma)^2\right]$	$\sigma_t^2\left[1+\frac{\pi-2}{\pi}\gamma^2+\beta^2\right]$
Skewness	$\frac{4-\pi}{2}\left[\frac{2(\beta-\gamma)^2}{\pi+(\pi-2)(\beta-\gamma)^2}\right]^{1.5}$	$\frac{\pi-4}{2}\left[\frac{2(\beta+\gamma)^2}{\pi+(\pi-2)(\beta+\gamma)^2}\right]^{1.5}$	$\frac{\sqrt{2}\gamma\left(2\pi\gamma^2-3\pi\beta^2-4\gamma^2\right)}{\left[\pi+(\pi-2)\gamma^2+\pi\beta^2\right]^{1.5}}$
Kurtosis	$2(\pi-3)\left[\frac{2(\beta-\gamma)^2}{\pi+(\pi-2)(\beta-\gamma)^2}\right]^2$	$2(\pi-3)\left[\frac{2(\beta+\gamma)^2}{\pi+(\pi-2)(\beta+\gamma)^2}\right]^2$	$3+\frac{8\gamma^2(\pi-3)}{\left[\pi+(\pi-2)\gamma^2+\pi\beta^2\right]^2}$
Correlation with Y_{t-1}	$\frac{(\beta-\gamma)\sqrt{\pi-2}}{\sqrt{\pi+(\pi-2)(\beta-\gamma)^2}}$	$\frac{(\beta+\gamma)\sqrt{\pi-2}}{\sqrt{\pi+(\pi-2)(\beta+\gamma)^2}}$	$\frac{\beta\sqrt{\pi}}{\sqrt{\pi+(\pi-2)\gamma^2+\pi\beta^2}}$

5. UNCONDITIONAL DISTRIBUTIONS

We refer to the expression "unconditional distribution" to indicate the distribution of the vector x_t unconditionally on Y_{t-1}. However, it is still conditional on its own past history.

Expectation: The expected value of the returns is the null vector, consistently with empirical findings and economic theory,

$$E(\mathrm{x}_t|D_t) = 0$$

Variance: The variance of x_t can be seen as decomposed into the sum of two components, an endogenous component, determined by the market internal structure, and an exogenous component, determined by news from the US market. The latter can be further represented as the sum of a component depending on the direct effect and another component depending on the feedback effect:

$$V(\mathrm{x}_t|D_t) = D_t\left(\Psi + \beta\beta^{\mathrm{T}} + \frac{\pi-2}{\pi}\gamma\gamma^{\mathrm{T}}\right)D_t$$

Skewness: In the multivariate SGARCH model, the feedback effect determines the asymmetric behavior of returns. More formally, multivariate skewness, as measured by Mardia's index, is zero if and only if the feedback parameter is the zero vector:

$$\gamma = 0 \Rightarrow S(\mathrm{x}_t|D_t) = 0, \quad \gamma \neq 0 \Rightarrow S(\mathrm{x}_t|D_t) > 0$$

The proof of this result is in the appendix.

Kurtosis: In the univariate SGARCH model ($p = 1$), we can show the presence of kurtosis. The same result holds for a linear combination of returns following the model. In the multivariate case we conjecture the existence of multivariate kurtosis as measured by Mardia's index.

Correlation: Let ρ be the $p \times 1$ vector whose ith component ρ_i is the correlation coefficient between the ith European return X_{it} and the previous day's US return Y_{t-1}, conditionally on volatility D_t. It easily follows from the definition of x_t as a function of the US shocks and ordinary properties of covariance that

$$\rho_i = \frac{\beta_i\sqrt{\pi}}{\sqrt{\pi + (\pi-2)\gamma_i^2 + \pi\beta_i^2}}, \quad i = 1, \ldots, p$$

as reported in the last row of Table 2. This result implies that $0 \leq \rho \leq 1_p$, where 1_p is a p-dimensional vector of ones. It also implies that ρ_i is an increasing function of β_i and a decreasing function of γ_i. It follows that association between European returns and previous day's US returns is directly related to the vector of direct effects β and inversely related to the vector of feedback effects γ.

The same statistics are reported for the univariate case in the last column of Table 2. The complete unconditional distribution could be obtained by simulation.

6. ANALYSIS OF SOME FINANCIAL MARKETS

We focus on a univariate and multivariate analysis of three European financial markets: the Dutch, the Swiss and the Italian market. They are small capitalized markets compared to the US market. The differences in sizes between the US market and the three European exchanges are evident. The weight of the capitalization of the US market on the world capitalization is over 40%, while the weights of each of the latter does not exceed 3%.

The close-to-close log-returns of representative market indexes have been observed in the period 18/01/1995–02/05/2003. The three markets (Dutch, Swiss and Italian) are represented by the AEX, SMI and MIB indexes, respectively. The returns of the US market are represented by the Standard & Poor 500 (S&P), the most popular market index for the New York Stock Exchange.

6.1. Univariate Analysis

The analysis proceeds by looking at the most important features of the returns in absence and in presence of the conditioning on the performance of the one-day lagged US market.

Table 3 reports the most salient statistics for the indexes when no conditioning is taken into account. They describe the typical features of financial returns. The average returns are very close to zero and a certain degree of negative skewness is apparent. Fat tails in the distribution are revealed by the kurtosis indices. Finally, the correlations with the one-day lagged S&P returns are reported. A dynamic linkage from the US market to the European markets does exist.

Then, we divide the entire samples into two subsamples, according to the sign of the one-day lagged S&P return. In order to take into account the

Table 3. Descriptive Statistics for the Three Stock Indexes.

	Average	Standard deviation	Skewness	Kurtosis	Correlation
AEX	0.019	1.575	−0.093	6.549	0.322
SMI	0.027	1.328	−0.134	6.689	0.276
MIB	0.022	1.564	−0.064	4.804	0.210

Table 4. Descriptive Statistics for the Three Stock Indexes According to the Sign of One-Day Lagged S&P Returns.

	Average	Standard deviation	Skewness	Kurtosis	Correlation
AEX					
S&P > 0	0.384	1.498	0.346	8.020	0.178
S&P < 0	−0.378	1.550	−0.412	5.071	0.275
SMI					
S&P > 0	0.252	1.241	0.192	7.517	0.157
S&P < 0	−0.219	1.366	−0.297	6.214	0.286
MIB					
S&P > 0	0.259	1.502	0.028	5.098	0.084
S&P < 0	−0.226	1.584	−0.038	4.494	0.210

differences in closing days of the stock exchanges, some hypotheses have to be made. We assume that if the European exchange is open at time t and the US exchange is closed at time $t-1$, then ε_{t-1} is set to zero (there is no information from the US market). If the European exchange is closed at time t, the US exchange information at time $t-1$ is useless; the next European exchange return (at time $t+1$) is related to ε_t. We compute the same statistics as above in this setting, summarized in Table 4.

The resulting statistics are very interesting. They show a different behavior of the European market indexes according to the sign of the last trading day in the American stock exchange.

In the three markets the average return is positive (negative) when the one-day lagged return of the S&P is positive (negative). The standard deviation of the returns of the European market indexes is always greater in presence of a negative sign coming from the US market. The skewness coefficient is negative and stronger when the American stock exchange return is negative; it is positive in the opposite case. Finally, the relationship between present European returns and past US returns is clearly stronger when the S&P return is negative. All these empirical findings match the

Table 5. Maximum Likelihood Estimates (Standard Error) of the Univariate SGARCH Models for the Three Indexes' Returns.

	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\omega}_2$	$\hat{\omega}_3$	$\hat{\beta}$	$\hat{\gamma}$
AEX	0.012	0.115	0.866	0.020	0.343	0.072
	(0.006)	(0.016)	(0.019)	(0.011)	(0.022)	(0.033)
SMI	0.035	0.133	0.840	0.006	0.267	0.121
	(0.011)	(0.019)	(0.021)	(0.009)	(0.024)	(0.034)
MIB	0.083	0.115	0.836	0.020	0.186	0.113
	(0.022)	(0.018)	(0.025)	(0.016)	(0.021)	(0.033)

theoretical features of the univariate SGARCH model (De Luca & Loperfido, 2004). Only the observed kurtosis is always smaller in presence of a negative US return, which contradicts the model.

The autocorrelation functions of the squared European returns steadily decrease, hinting that $p = q = 1$ in the variance equation. Table 5 contains the maximum likelihood estimates of the parameters of the three univariate models. According to the ratios between estimate and standard error, all the parameters are significant but the volatility spillover coefficients. This involves the advantage of a model including the parameters β and γ. The highest value of the quantity $\beta - \gamma$ refers to the AEX returns involving the major distance between the behaviors of the index conditionally on the signals coming from the US market.

On the whole, the economic interpretation is straightforward: it is relevant to distinguish between bad and good news from the US market. The inclusion of the effect of exogenous news can significantly improve the predictive performance.

In order to check the stability of the coefficients (particularly β and γ) we carried out recursive estimates. The first sample is composed of the observations from 1 to 1000. For each subsequent sample we added an observation. Fig. 1 shows the dynamics of the two parameters. The differences between β and γ are approximately constant for AEX and MIB. For SMI, the distance tends to be slightly more variable.

In order to evaluate the performance of the model in out-of-sample forecasting, we computed one-step-ahead forecasts of the volatility, $\sigma^2_{t|t-1}$, for $t = 1601, \ldots$, using the SGARCH model. We compared them with benchmark forecasts obtained from a standard GARCH(1,1) model. Following Pagan and Schwert (1990), we ran a regression of log squared returns versus log forecasted volatility, and then computed the F-test for the

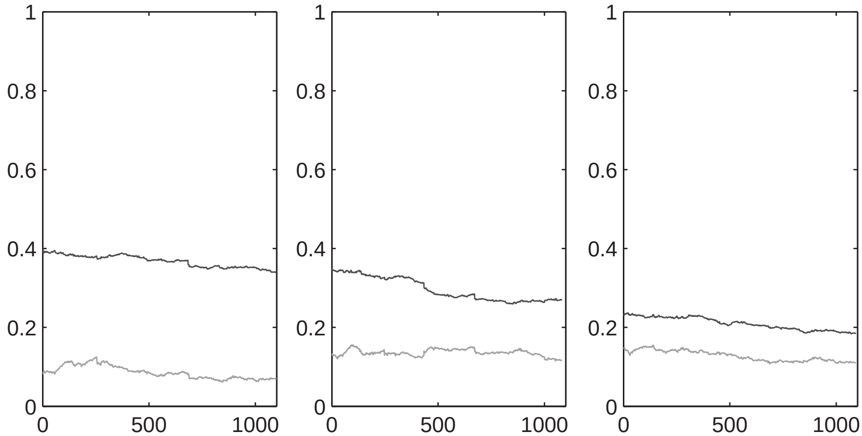


Fig. 1. Recursive Estimates of β (Top Curve) and γ (Bottom Curve) for AEX (Left), SMI (Middle) and MIB (Right).

hypothesis of a null intercept and a unit slope. The *F*-statistics for SGARCH and GARCH, respectively, are 53.14 and 69.24 (AEX), 59.25 and 69.24 (SMI), 70.00 and 72.65 (MIB) and thus favor the SGARCH model. Moreover, we computed how many times the sign of the return had been correctly predicted with the SGARCH model. The percentages of correct negative signs are 64.37 (AEX), 58.98 (SMI) and 56.15 (MIB). For positive sign, they are 56.78 (AEX), 51.33 (SMI) and 49.14 (MIB).

6.2. Multivariate Analysis

When a multivariate model is considered, the focus on daily data poses some problems. In fact there are unavoidably some missing values, due to the different holidays of each country. As an example, on the 1st of May, there is the Labor Holiday in Italy and Switzerland and the stock exchanges do not operate. But on the same day the stock exchange in the Netherlands is open. Deleting the day implies missing a datum. In order to overcome this drawback, we assumed that the Italian and Swiss variance in that day, σ^2_{1stMay}, has the same value as on the last opening day of the stock exchange ($\sigma^2_{30thApril}$ if not Saturday or Sunday). The next day variance, σ^2_{2ndMay} if not Saturday or Sunday, was computed according to a GARCH

(1,1) model without considering the past holiday. In general, for the k-th market index,

$$\sigma_{kt}^2 = \begin{cases} \omega_{0k} + \omega_{1k}\left(\sigma_{k,t-1}\zeta_{k,t-1}\right)^2 + \omega_{2k}\sigma_{k,t-1}^2 + \omega_{3k}\eta_{t-1}^2, & t-1 = \text{open} \\ \omega_{0k} + \omega_{1k}\left(\sigma_{k,t-2}\zeta_{k,t-2}\right)^2 + \omega_{2k}\sigma_{k,t-1}^2 + \omega_{3k}\eta_{t-1}^2, & t-1 = \text{close} \end{cases}$$

Maximum likelihood estimation has been performed using a Gauss code written by the authors. The algorithm converged after a few iterations. The parameters are again all significant, but the spillover parameters. Their values, reported in Table 6, are not very far from the corresponding estimates in the univariate context. In addition we obtain the estimates of the correlation coefficients, indicated by ρ_{ij}.

Table 6. Maximum Likelihood Estimates and Standard Errors of the Multivariate SGARCH Model. The Subscript Letters have the following Meanings: A = AEX, S = SMI, M = MIB.

Parameter	Estimate	Standard Error
ω_{0A}	0.042	0.008
ω_{1A}	0.085	0.010
ω_{2A}	0.865	0.019
ω_{3A}	0.021	0.010
ω_{0S}	0.075	0.012
ω_{1S}	0.100	0.013
ω_{2S}	0.826	0.021
ω_{3S}	0.008	0.006
ω_{0M}	0.148	0.027
ω_{1M}	0.087	0.013
ω_{2M}	0.852	0.023
ω_{3M}	−0.017	0.011
ρ_{AM}	0.665	0.012
ρ_{AS}	0.705	0.011
ρ_{MS}	0.605	0.014
β_A	0.335	0.022
γ_A	0.106	0.035
β_S	0.235	0.021
γ_S	0.148	0.032
β_M	0.192	0.029
γ_M	0.168	0.041

The diagnostic of the SGARCH model can be based on the squared norms $\{S_t\}$ of the residuals $\{\mathrm{r}_t\}$, defined as follows:

$$S_t = \mathrm{r}_t^{\mathrm{T}}\mathrm{r}_t,\ \mathrm{r}_t = \begin{cases} \hat{\Omega}_+^{-1/2}\left(\hat{D}_t^{-1}\mathrm{x}_t - \hat{\gamma}\sqrt{\frac{2}{\pi}}\right), & Y_{t-1} > 0 \\ \hat{\Omega}_-^{-1/2}\left(\hat{D}_t^{-1}\mathrm{x}_t - \hat{\gamma}\sqrt{\frac{2}{\pi}}\right), & Y_{t-1} < 0 \end{cases}$$

If the model is correctly specified, the following results hold:

1. The squared norms $\{S_t\}$ are i.i.d. and $S_t \sim \chi_p^2$.
2. If n is the number of observed returns, then

$$\sqrt{\frac{n}{2p}}\left(\frac{1}{n}\sum_{t=1}^{n} S_t - p\right) \overset{a}{\sim} N(0,1) \tag{7}$$

The proof of this result is given in the appendix. We computed the statistic in (7). Its value is -0.031, such that the hypothesis of skew-normality of the conditional distributions cannot be rejected.

7. CONCLUSIONS

The multivariate Skew-GARCH model is a generalization of the GARCH model aimed at describing the behavior of a vector of dependent financial returns when an exogenous shock coming from a leading financial market is taken into account. We analyzed returns from three European markets, while the leading market was identified as the US market. It turned out to be significant to consider the effects of the exogenous shock. The distributions of the European returns show different features according to the type of news arriving from the leading market. When the above assumptions are not consistent, the estimation step reveals the drawback. In this case, some parameters of the model are zero and the multivariate Skew-GARCH model shrinks to the simple multivariate GARCH model with constant correlation coefficients. A future extension of our proposed model would be to replace the multivariate skew-normal distribution by a multivariate skew-elliptical distribution, see the book edited by Genton (2004). For example, a multivariate skew-t distribution would add further flexibility by introducing an explicit parameter controlling tail behavior.

REFERENCES

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, *12*, 171–178.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics*, *69*, 542–547.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *The Review of Economics and Statistics*, *72*, 498–505.

Bollerslev, T., Engle, R., & Wooldridge, J. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, *96*, 116–131.

Campbell, J. Y., & Hentschel, L. (1992). No news is a good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, *31*, 281–331.

Dalla Valle, A. (2004). The skew-normal distribution. In: M. G. Genton (Ed.), *Skew-elliptical distributions and their applications: A journey beyond normality* (pp. 3–24). Boca Raton, FL: Chapman & Hall/CRC.

De Luca, G., & Loperfido, N. (2004). A Skew-in-Mean GARCH model. In: M. G. Genton (Ed.), *Skew-elliptical distributions and their applications: A journey beyond normality* (pp. 205–222). Boca Raton, FL: Chapman & Hall/CRC.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*, *50*, 987–1008.

Engle, R. F. (2002). Dynamic conditional correlation. A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, *20*, 339–350.

Eun, C. S., & Shim, S. (1989). International transmission of stock markets movements. *Journal of Financial and Quantitative Analysis*, *24*, 241–256.

Genton, M. G. (Ed.) (2004). *Skew-elliptical distributions and their applications: A journey beyond normality*. Boca Raton: Chapman & Hall/CRC.

Genton, M. G., & Loperfido, N. (2005). Generalized skew-elliptical distributions and their quadratic forms. *Annals of the Institute of Statistical Mathematics*, *57*, 389–401.

Karolyi, G. A. (1995). A multivariate GARCH model of international transmissions of stock returns and volatility: The case of United States and Canada. *Journal of Business and Economic Statistics*, *13*, 11–25.

Koch, P. D., & Koch, T. W. (1991). Evolution in dynamic linkages across daily national stock indexes. *Journal of International Money and Finance*, *10*, 231–251.

Lin, W., Engle, R., & Ito, T. (1994). Do bulls and bears move across borders? International transmission of stock return and volatility. *The Review of Financial Studies*, *7*, 507–538.

Loperfido, N. (2004). Generalized skew-normal distributions. In: M. G. Genton (Ed.), *Skew-elliptical distributions and their applications: A journey beyond normality* (pp. 65–80). Boca Raton: Chapman & Hall/CRC, FL.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns. *Econometrica*, *59*, 347–370.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*, 267–290.

APPENDIX

Proof of Theorem 1. We shall prove the theorem for $Y_{t-1} > 0$ only. The proof for $Y_{t-1} < 0$ is similar. By assumption, the joint distribution of random shocks is

$$\begin{pmatrix} \varepsilon_{t-1} \\ \zeta_t \end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0^{\mathrm{T}} \\ 0 & \Psi \end{pmatrix} \right)$$

By definition, $\Omega_+ = \Psi + \delta_+ \delta_+^{\mathrm{T}}$, so that

$$\begin{pmatrix} \varepsilon_{t-1} \\ \gamma\sqrt{2/\pi} + \delta_+ \varepsilon_{t-1} + \zeta_t \end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix} 0 \\ \gamma\sqrt{2/\pi} \end{pmatrix}, \begin{pmatrix} 1 & \delta_+^{\mathrm{T}} \\ \delta_+ & \Omega_+ \end{pmatrix} \right)$$

The conditional distribution of $\gamma\sqrt{2/\pi} + \delta_+ \varepsilon_{t-1} + \zeta_t$ given $\varepsilon_{t-1} > 0$ is multivariate *SN* (Dalla Valle, 2004):

$$\left(\gamma\sqrt{\frac{2}{\pi}} + \delta_+ \varepsilon_{t-1} + \zeta_t | \varepsilon_{t-1} > 0 \right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_+, \frac{\Omega_+^{-1}\delta_+}{\sqrt{1 - \delta_+^{\mathrm{T}} \Omega_t^{-1} \delta_+}} \right)$$

By definition, $\delta_+ = \beta - \gamma$, so that standard properties of absolute values lead to

$$\left(\gamma\sqrt{\frac{2}{\pi}} + \beta\varepsilon_{t-1} - \gamma|\varepsilon_{t-1}| + \zeta_t \middle| \varepsilon_{t-1} > 0 \right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_+, \frac{\Omega_+^{-1}\delta_+}{\sqrt{1 - \delta_+^{\mathrm{T}} \Omega_+^{-1} \delta_+}} \right)$$

By definition, US returns and European returns are $Y_{t-1} = \eta_{t-1}\varepsilon_{t-1}$ and $\mathrm{x}_t = D_t(\gamma\sqrt{2/\pi} + \beta\varepsilon_{t-1} - \gamma|\varepsilon_{t-1}| + \zeta_t)$. Hence,

$$\left(D_t^{-1}\mathrm{x}_t | Y_{t-1} > 0 \right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_+, \frac{\Omega_+^{-1}\delta_+}{\sqrt{1 - \delta_+^{\mathrm{T}} \Omega_+^{-1} \delta_+}} \right)$$

Apply now the Sherman-Morrison formula to the matrix $\Omega_+ = \Psi + \delta_+\delta_+^{\mathrm{T}}$:

$$\Omega_+^{-1} = \Psi^{-1} - \frac{\Psi^{-1}\delta_+\delta_+^{\mathrm{T}}\Psi^{-1}}{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}$$

A little algebra leads to the following equations:

$$\begin{aligned}
\Omega_+^{-1}\delta_+ &= \left(\Psi^{-1} - \frac{\Psi^{-1}\delta_+\delta_+^{\mathrm{T}}\Psi^{-1}}{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}\right)\delta_+ \\
&= \Psi^{-1}\delta_+\left(1 - \frac{\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}\right) \\
&= \frac{\Psi^{-1}\delta_+}{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+} \\
1 - \delta_+^{\mathrm{T}}\Omega_+^{-1}\delta_+ &= 1 - \frac{\delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+} = \frac{1}{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}
\end{aligned}$$

Recall now the definition of the vector α_+:

$$\begin{aligned}
\frac{\Omega_+^{-1}\delta_+}{\sqrt{1 - \delta_+^{\mathrm{T}}\Omega_+^{-1}\delta_+}} &= \frac{\Psi^{-1}\delta_+/\left(1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+\right)}{1/\sqrt{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}} \\
&= \frac{\Psi^{-1}\delta_+}{\sqrt{1 + \delta_+^{\mathrm{T}}\Psi^{-1}\delta_+}} = \alpha_+
\end{aligned}$$

We can then write

$$\left(D_t^{-1}\mathrm{x}_t \mid Y_{t-1} > 0\right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_+, \alpha_+\right)$$

and the proof is complete. □

Proof of the positivity of the unconditional skewness. We shall prove the theorem for $\mathrm{z}_t = D_t^{-1}\mathrm{x}_t$ only: the proof for x_t is similar. When $\gamma = 0$ the distribution of z_t is multivariate normal, so that it suffices to show that $\gamma \neq 0$ implies that $S(\mathrm{z}_t) > 0$. Let the vectors w_t, u_t, g_1, g_2 and the matrix Γ

be defined as follows:

$$\begin{aligned} &V(\mathrm{z}_t) = \Sigma, && \Gamma = \Sigma^{-1} \otimes \Sigma^{-1} \otimes \Sigma^{-1} \\ &\mathrm{w}_t \sim (\mathrm{z}_t | Y_{t-1} > 0), && \mathrm{g}_1 = E(\mathrm{w}_t \otimes \mathrm{w}_t \otimes \mathrm{w}_t) \\ &\mathrm{u}_t \sim (\mathrm{z}_t | Y_{t-1} < 0), && \mathrm{g}_2 = E(\mathrm{u}_t \otimes \mathrm{u}_t \otimes \mathrm{u}_t) \end{aligned}$$

The distribution of z_t is the mixture, with equal weights, of the distributions of w_t and u_t. Ordinary properties of mixtures lead to

$$\begin{aligned} E(\mathrm{z}_t \otimes \mathrm{z}_t \otimes \mathrm{z}_t) &= \frac{1}{2} E(\mathrm{w}_t \otimes \mathrm{w}_t \otimes \mathrm{w}_t) + \frac{1}{2} E(\mathrm{u}_t \otimes \mathrm{u}_t \otimes \mathrm{u}_t) \\ &= \frac{1}{2} (\mathrm{g}_1 + \mathrm{g}_2) \end{aligned}$$

Hence the multivariate skewness of the vector z_t can be represented as follows:

$$S(z_t) = \frac{1}{4} (\mathrm{g}_1 + \mathrm{g}_2)^{\mathrm{T}} \Gamma (\mathrm{g}_1 + \mathrm{g}_2)$$

The distribution of w_t equals that of $-\mathrm{u}_t$ only when $\gamma = 0$. By assumption $\gamma \neq 0$, so that

$$\mathrm{g}_1 + \mathrm{g}_2 \neq 0 \Rightarrow S(z_t) = \frac{1}{4} (\mathrm{g}_1 + \mathrm{g}_2)^{\mathrm{T}} \Gamma (\mathrm{g}_1 + \mathrm{g}_2) > 0$$

The last inequality follows from Σ being positive definite and from properties of the Kronecker product. The proof is then complete. □

Proof of the asymptotic distribution (7). Let the random variables $\{\bar{\mathrm{S}}_t\}$ and the random vectors $\{\bar{\mathrm{r}}_t\}$ be defined as follows:

$$\bar{S}_t = \bar{\mathrm{r}}_t^{\mathrm{T}} \bar{\mathrm{r}}_t, \quad \bar{\mathrm{r}}_t = \begin{cases} \Omega_+^{-1/2} \left(D_t^{-1} \mathrm{x}_t - \gamma \sqrt{\dfrac{2}{\pi}} \right) & Y_{t-1} > 0 \\ \Omega_-^{-1/2} \left(D_t^{-1} \mathrm{x}_t - \gamma \sqrt{\dfrac{2}{\pi}} \right) & Y_{t-1} < 0 \end{cases}$$

From Section 3 we know that the distributions of $D_t^{-1}\mathrm{x}_t | Y_{t-1} < 0$ and $D_t^{-1}\mathrm{x}_t | Y_{t-1} > 0$ are skew-normal:

$$\left(D_t^{-1}\mathrm{x}_t | Y_{t-1} > 0\right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_+, \alpha_+\right)$$
$$\left(D_t^{-1}\mathrm{x}_t | Y_{t-1} < 0\right) \sim SN_p\left(\gamma\sqrt{\frac{2}{\pi}}, \Omega_-, \alpha_-\right)$$

Apply now linear properties of the *SN* distribution:

$$(\bar{\mathrm{r}}_t | Y_{t-1} > 0) \sim SN_p\left(0, I_p, \lambda_+\right), \quad \lambda_+ = \Omega_+^{1/2}\alpha_+$$
$$(\bar{\mathrm{r}}_t | Y_{t-1} < 0) \sim SN_p\left(0, I_p, \lambda_-\right), \quad \lambda_- = \Omega_-^{1/2}\alpha_-$$

First notice that the vectors $\{\bar{\mathrm{r}}_t\}$ are i.i.d. Moreover, the distribution of $\bar{\mathrm{r}}_t$ is the mixture, with equal weights, of two p-dimensional *SN* distributions with zero location parameter and the identity matrix for the scale parameter, that is:

$$f(\bar{\mathrm{r}}_t) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{\bar{\mathrm{r}}_t^{\mathrm{T}}\bar{\mathrm{r}}_t}{2}\right)\left[\Phi\left(\lambda_-^{\mathrm{T}}\bar{\mathrm{r}}_t\right) + \Phi\left(\lambda_+^{\mathrm{T}}\bar{\mathrm{r}}_t\right)\right]$$

The above density can be represented as follows:

$$f(\bar{\mathrm{r}}_t) = 2\phi_p\left(\bar{\mathrm{r}}_t; I_p\right)\omega(\bar{\mathrm{r}}_t)$$

where $\phi_p(\cdot;\Sigma)$ denotes the density of $N_p(0,\Sigma)$ and $w(\cdot)$ is a function such that $0 \leq w(-\bar{\mathrm{r}}_t) = 1 - w(\bar{\mathrm{r}}_t) \leq 1$. Hence the distribution of $\bar{\mathrm{r}}_t$ is generalized *SN* with the zero location parameter and the identity matrix for the scale parameter (Loperfido, 2004; Genton & Loperfido, 2005). It follows that the pdf of even functions $g(\bar{\mathrm{r}}_t) = g(-\bar{\mathrm{r}}_t)$ does not depend on $w(\cdot)$. As an immediate consequence, we have:

$$\bar{S}_t = \bar{\mathrm{r}}_t^{\mathrm{T}}\bar{\mathrm{r}}_t \sim \chi_p^2 \Rightarrow E\left(\bar{S}_t\right) = p, \quad V\left(\bar{S}_t\right) = 2p$$

A standard application of the Central Limit Theorem leads to

$$\sqrt{\frac{n}{2p}}\left(\frac{1}{n}\sum_{t=1}^{n}\bar{S}_t - p\right) \overset{a}{\sim} N(0,1)$$

When the sample size is large, the distribution of $\{\mathrm{r}_t\}$ approximates the distribution of $\{\bar{\mathrm{r}}_t\}$. Hence the squared norms $\{\mathrm{S}_t\}$ are approximately independent and identically distributed according to a χ_p^2 distribution. Moreover,

$$\sqrt{\frac{n}{2p}}\left(\frac{1}{n}\sum_{t=1}^{n}S_t - p\right) \overset{a}{\sim} N(0,1)$$

and this completes the proof. $\square$

This page intentionally left blank

SEMI-PARAMETRIC MODELLING OF CORRELATION DYNAMICS

Christian M. Hafner, Dick van Dijk and Philip Hans Franses

ABSTRACT

In this paper we develop a new semi-parametric model for conditional correlations, which combines parametric univariate Generalized Auto Regressive Conditional Heteroskedasticity specifications for the individual conditional volatilities with nonparametric kernel regression for the conditional correlations. This approach not only avoids the proliferation of parameters as the number of assets becomes large, which typically happens in conventional multivariate conditional volatility models, but also the rigid structure imposed by more parsimonious models, such as the dynamic conditional correlation model. An empirical application to the 30 Dow Jones stocks demonstrates that the model is able to capture interesting asymmetries in correlations and that it is competitive with standard parametric models in terms of constructing minimum variance portfolios and minimum tracking error portfolios.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 59–103

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20003-8

1. INTRODUCTION

Estimating and forecasting the (conditional) covariance structure of financial asset returns plays a crucial role in important areas of finance such as portfolio construction, asset pricing and risk management. It is therefore not surprising that modelling the dynamics of conditional covariance matrices has received ample attention. This research area was boosted by the introduction of the (Generalized) AutoRegressive Conditional Heteroskedasticity ((G)ARCH) model by Engle (1982) and Bollerslev (1986). Early-stage multivariate GARCH models, including the VEC-model of Bollerslev, Engle, and Wooldridge (1988) and the BEKK-model of Engle and Kroner (1995), describe the dynamics of all elements of the covariance matrix in a flexible way, using many parameters. As such, they suffer from the "curse of dimensionality", meaning to say that the number of parameters to be estimated in these models increases very rapidly as the number of included assets increases.[1] This basically prevented their successful application in empirically relevant settings, where portfolios might consist of 10s or even 100s of assets. More parsimonious models, such as the Constant Conditional Correlation (CCC) model of Bollerslev (1990), are applied more frequently in practice. These models limit the possible dynamic patterns of the covariance matrix in important ways, however. In the CCC model, for example, the covariance between two assets changes over time only because of variation in their conditional volatilities, as their correlation is assumed to be constant.

Recently, several extensions of the CCC model have been proposed that allow for time-variation in the conditional correlations. On the one hand, Engle (2002) and Tse and Tsui (2002) developed Dynamic Conditional Correlation (DCC) models, where the conditional correlations evolve according to a GARCH-type structure. The attractive feature of this approach is that the number of parameters in the conditional correlation model can be limited by using the idea of "correlation targeting", which means that the unconditional correlations implied by the model are restricted to be equal to the unconditional sample correlations. In that case, the basic DCC-model of Engle (2002), for example, involves only two unknown parameters. Cappiello, Engle, and Sheppard (2003) and Hafner and Franses (2003) considered generalizations to allow for richer correlation dynamics, while still keeping the number of parameters within reasonable bounds even for large numbers of assets.[2] On the other hand, Pelletier (2005) and Silvennoinen and Teräsvirta (2005) proposed Regime-Switching Conditional Correlation models, where the correlations switch back and forth between a limited number of different values, according to an unobserved Markov-Switching process or according to the value of observed exogenous variables, respectively. The main disadvantage of these models is that correlation targeting can no longer be applied in a straightforward manner, such that the number of parameters to be estimated again grows rapidly (that is, quadratically) with the number of assets.

In this paper, we put forward a new semi-parametric model for the conditional covariance matrix that is flexible yet easy to estimate even in high dimensions. The model is semi-parametric, in the sense that the conditional variances are described parametrically, for example, using standard univariate GARCH-type models, while the conditional correlations are estimated using nonparametric techniques. The conditional correlations are assumed to depend on an observable exogenous (or pre-determined) variable. The appropriate choice of this conditioning variable depends on the particular application. For example, in Section 4 below, we model the correlations between the 30 individual stocks included in the Dow Jones Industrial Average (DJIA) index, and allow these to depend on the market volatility and market return. These variables are motivated by both theoretical and empirical considerations. First, as shown by Andersen, Bollerslev, Diebold, and Ebens (2001), in case asset returns exhibit a factor structure, the correlations are affected by the conditional volatility of the latent factor. In particular, under quite realistic assumptions concerning the factor loadings, high factor volatility induces high correlations between individual asset returns. In the context of the DJIA stocks, a factor structure is a sensible possibility and, furthermore, the market portfolio is a natural candidate to use as a proxy for the latent common factor. Additional empirical evidence supporting the hypothesis that correlations increase in high volatility states can be found in Longin and Solnik (1995, 2001), and Ramchand and Susmel (1998), but see Loretan and English (2000), and Forbes and Rigobon (2002) for critical discussion. Second, modelling the correlations as a function of the lagged index return is motivated by the literature on "correlation breakdown", which has documented that typically asset correlations increase in bear markets, but are not affected (or to a much lesser extent) in bull markets, see Longin and Solnik (2001), Ang and Chen (2002), Butler and Joaquin (2002), and Campbell, Koedijk, and Kofman (2002), among others.

The idea of allowing the conditional correlations to depend on exogenous factors resembles the regime-switching model of Silvennoinen and Teräsvirta (2005). However, instead of a priori imposing a certain parametric specification for the correlation functions, we suggest a nonparametric estimator.

Basically, a Nadaraya–Watson kernel regression is applied to each conditional correlation individually, using the same bandwidth to guarantee that the resulting estimator of the correlation matrix is positive definite. In order to obtain a genuine correlation matrix in finite samples, we apply the same transformation as in the DCC model of Engle (2002). Our nonparametric estimator only requires the assumption that correlations are smooth functions of the conditioning variable, while the data completely determines the appropriate shape of the functional relationship. An additional advantage is that the individual correlation functions are allowed to exhibit quite different shapes, which is problematic in currently available parametric specifications, as will become clear below.

The paper proceeds as follows: In Section 2, we briefly review the various conditional correlation models that have recently been proposed. We describe our new semi-parametric model in Section 3, including a detailed discussion of the issues involved in the nonparametric part of the estimation procedure. In Section 4, we consider daily returns of the 30 stocks included in the Dow Jones Industrial Average Index over the period 1989–2003, and apply several models to describe the dynamics in the conditional covariance matrix of these stocks. We find that our new semi-parametric model is competitive with rival specifications, in particular, in terms of tracking error minimization. We conclude in Section 5, also pointing out interesting directions for future research. The appendix contains a proof of the consistency and asymptotic normality of the semi-parametric estimator developed in Section 3.

2. DYNAMIC CONDITIONAL CORRELATION MODELS

In this section, we review existing models for describing the dynamics in conditional correlations of asset returns. These models are a specific subgroup within the general class of multivariate GARCH models. We refer to Bauwens, Laurent, and Rombouts (2005) for elaborate discussion of other models in this class.

Let r_t denote an N-dimensional vector time series, such as daily returns on the stocks in the Dow Jones index. Suppose for simplicity (but without loss of generality) that the conditional mean of r_t is constant and equal to zero, but that its conditional covariance matrix is time-varying. That is, we have the generic model

$$\mathrm{E}[r_t|F_{t-1}] = 0 \tag{1}$$

$$\mathrm{V}(r_t|F_{t-1}) = \mathrm{E}\left[r_t r_t'|F_{t-1}\right] = H_t \tag{2}$$

where F_t is the information set that includes all information up to and including time t. The conditional covariance matrix H_t can be decomposed as

$$H_t = D_t(\theta) R_t D_t(\theta) \tag{3}$$

with $D_t(\theta) = \mathrm{diag}\left(\sqrt{h_{1t}}, \ldots, \sqrt{h_{Nt}}\right)$ a diagonal matrix with the square root of the conditional variances h_{it}, parameterized by the vector θ, on the diagonal. For example, h_{it} can be described by the univariate GARCH-model of Bollerslev (1986) or nonlinear extensions such as the GJR-GARCH model of Glosten, Jagannathan and Runkle (1993). The matrix R_t, with the (i,j)-th element denoted as ρ_{ijt}, is the possibly time-varying conditional correlation matrix. As noted by Engle (2002), R_t also is the conditional covariance matrix of the standardized disturbances $\varepsilon_t = D_t^{-1}(\theta) r_t$, that is

$$\mathrm{E}\left[\varepsilon_t \varepsilon_t'|F_{t-1}\right] = R_t \tag{4}$$

Different specifications of R_t give rise to different models. First, the CCC model of Bollerslev (1990) assumes that $R_t = R$ is constant. In that case, estimation is straightforward and can be performed in two steps. First, one estimates univariate GARCH models for the individual series r_{it}, $i = 1, \ldots, N$, using (quasi) maximum likelihood to obtain estimates of the conditional variances h_{it}. Second, R can be consistently estimated by the sample covariance matrix of the standardized residuals $\hat{\varepsilon}_t = \widehat{D}_t^{-1} r_t$, that is $\widehat{R} = T^{-1}\Sigma_{t=1}^{T}\, \hat{\varepsilon}_t\, \hat{\varepsilon}_t'$ with T denoting the sample size.

Second, Engle (2002) and Tse and Tsui (2002) introduced the class of DCC models, where R_t is allowed to vary according to a GARCH-type process, such as

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha \varepsilon_{t-1}\varepsilon_{t-1}' + \beta Q_{t-1} \tag{5}$$

$$R_t = Q_t^{*\ -1} Q_t Q_t^{*\ -1} \tag{6}$$

where $\bar{Q} = \mathrm{E}[\varepsilon_t \varepsilon_t']$ is the unconditional covariance matrix of standardized returns, Q_t^* is a diagonal matrix containing the square roots of the diagonal elements of Q_t, and α and β are scalars. The attractive feature of the DCC model is not only that two-step estimation is still feasible, but also that "correlation targeting" can be used in the second step. That is, after estimating univariate GARCH models for the conditional volatilities h_{it} in the first step, $\bar{Q}$ can be replaced by the sample covariance matrix of the standardized residuals $\hat{\varepsilon}_t$. This imposes the restriction that the unconditional correlations as implied by the model are equal to the unconditional sample correlations, and reduces the number of parameters to be estimated in the second step to two, namely, α and β. Engle and Sheppard (2001) prove consistency and asymptotic normality of this two-step estimator. Note that correlation targeting is the multivariate analogue of variance targeting in the univariate GARCH framework as introduced by Engle and Mezrich (1996).

The DCC model in (5) has two particular features that may render it too restrictive in practice. First, the model imposes all correlations to have the same dynamic pattern as governed by the parameters α and β. It is not difficult to come up with examples of situations where this restriction may be violated. For example, correlations between stocks from the same industry may behave rather differently from correlations between stocks from different industries. Second, the model implies symmetry, in the sense that a pair of positive standardized returns ε_{it} and ε_{jt} has the same effect on the conditional correlation as a pair of negative returns of the same magnitude. Recent empirical evidence in Ang and Chen (2002), among others, indicates the presence of asymmetries in correlations, suggesting for example, that correlations increase for large downward stock price movements but not, or less, for large upward movements.

Several extensions of the basic model in (5) have been developed that alleviate these shortcomings. Hafner and Franses (2003) suggested the semi-generalized DCC (SGDCC) model that allows for asset-specific news impact parameters by replacing (5) with

$$Q_t = \left(1 - \bar{\alpha}^2 - \beta\right)\bar{Q} + \alpha\alpha' \odot \varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1} \tag{7}$$

where $\odot$ denotes the Hadamard product and α now is an $(N \times 1)$ vector, $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)'$, and $\bar{\alpha} = N^{-1}\sum_{i=1}^{N} \alpha_i$ is the average news impact parameter. The memory parameter β can be made asset-specific in identical fashion. Note that the number of parameters in (7) increases linearly with the number of assets N. Cappiello et al. (2003) allowed for nonlinear impact of shocks on the conditional correlations, in particular of simultaneous negative shocks in assets i and j, by considering the asymmetric DCC (ADCC) model

$$Q_t = (1 - \alpha - \beta)\bar{Q} - \gamma\bar{N} + \alpha\varepsilon_{t-1}\varepsilon'_{t-1} + \beta Q_{t-1} + \gamma n_{t-1}n'_{t-1} \tag{8}$$

where $n_t = \mathrm{I}[\varepsilon_t < 0] \odot \varepsilon_t$, with $\mathrm{I}[A]$ being the indicator function for the event A, and $\bar{N} = \mathrm{E}[n_t n_t']$. In estimation of the ADCC model, $\bar{N}$ can be replaced by its sample analogue $T^{-1}\Sigma_{t=1}^{T}\hat{n}_t\hat{n}_t'$, such that the number of parameters is equal to three independent of the number of assets included. Combining the ideas in (7) and (8) is also possible, see Cappiello et al. (2003).

Third, Pelletier (2005) developed a Markov-Switching conditional correlation model that allows the conditional correlations to switch between k distinct values, by specifying the conditional correlation matrix in (3) as

$$R_t = R_{s_t} \tag{9}$$

where s_t is an unobserved first-order Markov process with k states, with transition probabilities $P[s_t = j | s_{t-1} = i] = p_{ij}$, for $i,j = 1, \ldots, k$. Note that in contrast to the DCC-type models discussed above, this model does not necessarily imply any direct dynamic dependence in the conditional correlation matrix, although there is indirect dependence through the Markov-character of s_t. Estimation of the parameters in this model can again be done using a two-step approach, where in the second step, the EM-algorithm can be employed, see Pelletier (2005) for details. Note however that correlation targeting is not possible here. Hence, the elements in the R_{s_t}, $s_t = 1, \ldots, k$ matrices have to be treated as unknown parameters such that the number of parameters to be estimated in the second step equals $k(N(N-1)/2 + (k-1))$.

Fourth, Silvennoinen and Teräsvirta (2005) consider a smooth transition conditional correlation (STCC) model that allows the conditional correlations to vary continuously between two extremes, according to the value of an observed variable x_t,

$$R_t = R_1(1 - G(x_t; \gamma, c)) + R_2 G(x_t; \gamma, c) \tag{10}$$

where the function $G(x_t; \gamma, c)$ is continuous and bounded between 0 and 1. For example, $G(x_t; \gamma, c)$ can be the logistic function

$$G(x_t; \gamma, c) = \frac{1}{1 + \exp(-\gamma(x_t - c))} \tag{11}$$

which tends to 0 (1) for very small (large) values of the transition variable x_t, relative to the threshold parameter c. The parameter γ determines the smoothness of the change from 0 to 1 as x_t increases. The transition variable x_t should of course be included in the information set F_{t-1} to make the model operational and useful for forecasting purposes. Although the parameters in this model can again be estimated using a two-step approach, correlation targeting is not possible, such that the number of parameters to be estimated in the second step equals $2N(N-1)/2+2$. An interesting special case of the STCC model (10) with (11) arises when the parameter $\gamma \to \infty$, such that the change of the correlation matrix from R_1 and R_2 occurs instantaneously at $x_t = c$. In that case, estimation of the parameters is simplified considerably, in the sense that correlation targeting becomes possible again. Note that, if $\gamma = \infty$ such that $G(x_t; \gamma, c) = \mathrm{I}[x_t \geq c]$, $R_1 = \mathrm{E}[n_{1t}n'_{1t}]$ with $n_{1t} = \mathrm{I}[x_t < c] \odot \varepsilon_t$ and R_2 can be similarly defined as $R_2 = \mathrm{E}[n_{2t}n'_{2t}]$ with $n_{2t} = \mathrm{I}[x_t \geq c] \odot \varepsilon_t$. For a fixed value of c, R_1 and R_2 can then be replaced by their sample analogues $\left(\Sigma_{t=1}^{T} \mathrm{I}[x_t < c]\right)^{-1} \Sigma_{t=1}^{T} \hat{n}_{1t}\, \hat{n}'_{1t}$ and $\left(\Sigma_{t=1}^{T} \mathrm{I}[x_t \geq c]\right)^{-1} \Sigma_{t=1}^{T} \hat{n}_{2t}\, \hat{n}'_{2t}$, respectively. This leaves only c to be estimated, which can be achieved by means of grid search.

A crucial ingredient of the STCC model is of course the transition variable x_t. Berben and Jansen (2005) take $x_t = t$ in an application to industry stock returns, effectively modelling a gradual structural change in correlations. Alternatively, x_t could be taken to be a lagged market return or volatility, motivated by the correlation breakdown effect and by the presence of a factor structure, respectively, as discussed in Section 1. Silvennoinen and Teräsvirta (2005) develop an LM-type test for the constancy of the correlation matrix against the STCC-alternative, which also can be used as a guiding tool in the choice of the appropriate transition variable x_t.

3. A SEMI-PARAMETRIC CONDITIONAL CORRELATION MODEL

In this section, we put forward our semi-parametric conditional correlation model. The model can be considered a generalization of the STCC model of Silvennoinen and Teräsvirta (2005) given in (10) and (11). We are sympathetic to the idea that the conditional correlations depend on exogenous factors such as the market return or volatility. However, the STCC model is restrictive, in the sense that it imposes that all correlations depend on x_t in the same way, as determined by the transition function $G(x_t; \gamma, c)$. This assumption may be unrealistic in practice, as one can easily imagine that different correlations respond differently to changes in x_t; see the factor model in Andersen et al. (2001) and the empirical application in Silvennoinen and Teräsvirta (2005) for examples. It might be possible to generalize the STCC model to allow for asset-specific smooth transition in correlations, similar to the SGDCC model of Hafner and Franses (2003) in (7). In that case, however, it would be difficult, if not impossible, to guarantee positive definiteness of the correlation matrix for all values of x_t. In addition, even if it is reasonable to assume that all correlations have the same functional relationship to x_t, the specific form of this dependence, such as the logistic function in (11), is not likely to be known a priori. Hence, we suggest to model the conditional correlation matrix in a nonparametric way.

In sum, we propose the following semi-parametric model:

$$r_t = D_t(\theta)\varepsilon_t \tag{12}$$

with $D_t(\theta)$ as defined before, $\mathrm{E}[\varepsilon_t|F_{t-1}] = 0$ and

$$\mathrm{E}[\varepsilon_t\varepsilon_t'|F_{t-1}, x_t = x] = R(x) \tag{13}$$

where x_t is an observed variable, for example, the market return or volatility at time $t-1$. Assuming that we have a $\sqrt{T}$-consistent estimator of θ, which we denote by $\hat{\theta}$, the standardized residuals are then defined by $\hat{\varepsilon}_t = D_t(\hat{\theta})^{-1} r_t$. These are used in the second stage to estimate $R(x)$ nonparametrically, as

$$\widehat{R}(x) = \widehat{Q}^*(x)^{-1}\widehat{Q}(x)\widehat{Q}^*(x)^{-1} \tag{14}$$

where

$$\widehat{Q}(x) = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)} \tag{15}$$

where $K_h(\cdot) = (1/h)K(\cdot/h)$, K is a kernel function and h a bandwidth parameter. As before, $\widehat{Q}^*(x)$ is a diagonal matrix with the square roots of the diagonal elements of $\widehat{Q}(x)$ on its diagonal. Hence the (i, j)-th element of $\widehat{R}(x)$ in (14) can be written as

$$\widehat{\rho}_{ij}(x) = \frac{\hat{q}_{ij}(x)}{\sqrt{\hat{q}_{ii}(x)\hat{q}_{jj}(x)}}$$

The estimator $\widehat{R}(x)$ in (14) is essentially a transformed Nadaraya–Watson estimator applied to the elements of the matrix $\hat{\varepsilon}_t\,\hat{\varepsilon}_t'$. Of course, other nonparametric estimators such as local polynomials could be used as well, see for example, Fan and Gijbels (1996), Härdle, Lütkepohl, and Chen (1997) and Pagan and Ullah (1999) for reviews, and Fan and Yao (1998) and Ziegelmann (2002) for applications to volatility modelling. Recall that the conditional covariance matrix of r_t can be written as

$$H_t = \mathrm{V}(r_t|F_{t-1}, x_t = x) = D_t(\theta)R(x)D_t(\theta)$$

where the semi-parametric character of the model for H_t now becomes obvious: $D_t(\theta)$ is modelled parametrically using standard univariate

GARCH-type models, for example, while the correlation matrix $R(x)$ is treated in a nonparametric fashion.

Note that (15) essentially boils down to univariate kernel regression. Hence, unlike many of the early-stage multivariate GARCH models and unlike many estimation problems in nonparametric analysis, our semi-parametric model does not suffer from the curse of dimensionality. It is straightforward to see that (14) generalizes the STCC model in (10), as we do not impose that all correlations are related in the same way to x_t, nor do we assume any particular parametric form for this dependence. Also recall that one is likely to encounter difficulties in estimation when applying the STCC model to large systems given that correlation targeting is not possible. Using the non-parametric estimator in (14) and (15) this problem obviously is avoided. Note that this also suggests that our semi-parametric model might be used in exploratory data analysis to examine the shape of the dependence of ρ_{ijt} on x_t.

Apart from the choice of x_t, the crucial decision to be made when implementing the non-parametric estimator of the conditional correlation concerns the bandwidth h. One option would be to use a constant bandwidth, which could be selected using cross-validation or the plug-in method, see for example, Pagan and Ullah (1999). This has the drawback that the variance of $R(x)$ becomes very large near the boundaries of the support of x_t or more generally in areas of the support of x_t where the data is sparse. To alleviate the problem of data sparsity, we suggest to use local bandwidths. In particular, we allow h to depend on the design density $f(x)$ such that $h(x) = bf(x)^{-a}$, where b is a positive constant and $0 \leq a \leq 1$. For positive values of a, this implies that the bandwidth becomes larger when relatively few observations around the point $x_t = x$ are available such that $f(x)$ is small, while $h(x)$ becomes smaller in high-density regions. For $a = 0$, one has the standard kernel smoother with constant bandwidth $h(x) = b$. Jennen-Steinmetz and Gasser (1988) showed that $a = 1/4$ corresponds roughly to spline smoothing and $a = 1$ to generalized nearest-neighbour smoothing. In the empirical application in the next section, we will illustrate the sensitivity of the estimates of $R(x)$ to different values of a.

In Theorem 1, we state the asymptotic properties of the nonparametric estimator of the conditional correlation matrix as given in (14) and (15). Using the notation $\eta_t = \mathrm{vech}(\varepsilon_t \varepsilon_t')$ and $r(x) = \mathrm{vech}(R(x))$, we have

$$r(x) = \mathrm{E}[\eta_t | x_t = x] \tag{16}$$

$$\mathrm{V}(\eta_t | x_t = x) = \mathrm{E}[(\eta_t - r(x))(\eta_t - r(x))' | x_t = x] \tag{17}$$

Also, denote by $\hat{f}(x)$ an estimator of $f(x)$, the density of x_t. Under the assumptions stated in the appendix, the nonparametric estimator of $R(x)$ is consistent and asymptotically normal, as detailed in the following theorem.

Theorem 1. Under Assumptions (A1) to (A6), and $b \to 0$ as T increases such that $Tb^5 \to 0$ and $Tb \to \infty$ it holds that

1. The diagonal elements of the matrix $\widehat{Q}$ in (15) converge in probability to 1:

$$\widehat{q}_{ii}(x) \xrightarrow{p} 1, \quad i = 1, \ldots, N$$

2. The estimator $\widehat{R}(x)$ in (14) is consistent and asymptotically normal:

$$\sqrt{Tb}(\widehat{r}(x) - r(x)) \xrightarrow{\mathscr{L}} N(0, \Sigma(x))$$

where

$$\Sigma(x) = \frac{\int K^2(u)\,du}{f(x)^{1-a}} \mathrm{V}(\eta_t | x_t = x)$$

The proof of the theorem is given in the appendix. Also, an explicit expression for $\mathrm{V}(\eta_t|x_t = x)$ is given there, which depends on $R(x)$ and on the fourth moment characteristics of the innovations ε_t. If these are assumed to be normally distributed, then a consistent estimator of $\mathrm{V}(\eta_t|x_t = x)$ is easily constructed by replacing $R(x)$ by $\hat{R}(x)$. If the distribution is unknown, then fourth moments of ε_t can be estimated using corresponding moments of the standardized residuals $\hat{\varepsilon}_t$, but for consistency we would need the assumption of finite eighth moments.

4. EMPIRICAL APPLICATION

In this section, we explore the potential of our semi-parametric correlation model to describe and forecast the conditional covariance matrix of asset returns. We are mainly interested in empirically relevant situations, where the number of assets N is fairly large, in which it is difficult to apply early-stage multivariate GARCH models and the regime-switching correlation models (9) and (10). We use daily returns of the 30 stocks that constituted the DJIA index between November 1999 and April 2004. Several conditional correlations models are estimated for these stock returns. We evaluate and compare the models not only by means of statistical criteria, but mainly by applying the models for constructing minimum variance portfolios (MVPs) and minimum tracking error volatility portfolios.

4.1. Data

The sample period covers 15 years, running from January 1, 1989 until December 31, 2003. Days on which the stock exchange was closed were removed from the sample, leaving a sample size of $T = 3{,}784$ observations. We use the first decade 1989–1998 (2,528 observations) for in-sample estimation and analysis of competing models, and set aside the final five years 1999–2003 (1,256 observations) for out-of-sample forecasting.

Part of the challenge in this application stems from the fact that the general trend in the stock market was quite different during the initial estimation period and the forecasting period. The 1990s were characterized by a prolonged bull market, which ended with the burst of the internet bubble and was followed by a bearish market around the turn of the millennium. Tables 1 and 2 show annualized means and standard deviations of the daily stock returns, which obviously reflect the overall market sentiment. All 30 stocks had large positive average returns during the first decade of our sample period, while the average return turned negative for 18 stocks during the final five years. Only the stocks of Alcoa Inc. (AA) and 3 M Company (MMM) had higher returns during the period 1999–2003 than during the preceding 10 years. It is also seen that the volatility of stock returns was considerably higher during the last five years of the sample period. For most stocks the standard deviation increased by about 50%. Tables 1 and 2 also contain estimation results from the three-factor regression model developed by Fama and French (1993, 1996), given by

$$r_{it} - r_{f,t} = \alpha + \beta_{i,\mathrm{M}}(r_{\mathrm{M},t} - r_{f,t}) + \beta_{i,\mathrm{SMB}} r_{\mathrm{SMB},t} + \beta_{i,\mathrm{HML}} r_{\mathrm{HML},t} + \varepsilon_t \tag{18}$$

where r_{it} is the daily return for stock i, $r_{f,t}$ is the risk-free rate, $r_{\mathrm{M},t}$ is the market portfolio return, $r_{\mathrm{SMB},t}$ (Small-Minus-Big) and $r_{\mathrm{HML},t}$ (High-Minus-Low) are size and book-to-market factor portfolio returns. The size factor return $r_{\mathrm{SMB},t}$ is computed as the difference between the returns on portfolios of stocks with small and large market capitalization, while $r_{\mathrm{HML},t}$ is obtained similarly using portfolios of stocks with high and low book-to-market values. We employ the daily market and factor returns available on the website of Kenneth French.[3] Most of the DJIA stocks move one-to-one with the market, given that the estimates of $\beta_{\mathrm{i,M}}$ are fairly close to 1. Exceptions include Johnson and Johnson (JNJ) and JP Morgan (JPM), for which we find very low and high market betas, respectively. In general, the difference in $\beta_{\mathrm{i,M}}$ during the two sub-periods is not very large. The negative estimates of $\beta_{i,\mathrm{SMB}}$ are not surprising of course, given the size of the DJIA stocks. For the majority of stocks we find that $\beta_{i,\mathrm{HML}}$ was substantially higher during

Table 1. Characteristics of Daily Returns on DJIA Stocks – 1989–1998.

	Mean	S.D.	$\beta_{i,\mathrm{M}}$	$\beta_{i,\mathrm{SMB}}$	$\beta_{i,\mathrm{HML}}$
AA	9.834	26.273	1.248(0.072)	0.157(0.082)	0.739(0.116)
AXP	14.803	31.448	1.566(0.070)	−0.232(0.090)	0.419(0.111)
BA	9.023	28.416	1.191(0.077)	0.108(0.094)	0.022(0.119)
C	26.579	34.322	1.840(0.077)	−0.168(0.112)	0.489(0.137)
CAT	10.611	28.838	1.292(0.068)	0.030(0.091)	0.597(0.122)
DD	13.064	25.196	1.115(0.065)	−0.419(0.082)	0.425(0.097)
DIS	17.021	26.772	1.101(0.062)	−0.067(0.080)	−0.102(0.105)
EK	7.139	27.269	0.881(0.058)	−0.065(0.086)	0.061(0.107)
GE	22.288	21.788	1.058(0.040)	−0.593(0.060)	−0.180(0.070)
GM	6.112	28.822	1.637(0.066)	−0.116(0.083)	1.498(0.119)
HD	36.930	30.892	1.317(0.061)	−0.095(0.085)	−0.603(0.107)
HON	16.721	27.152	1.240(0.085)	−0.060(0.109)	0.524(0.120)
HPQ	16.566	35.701	1.058(0.080)	−0.293(0.112)	−0.856(0.151)
IBM	11.084	27.831	0.826(0.057)	−0.336(0.078)	−0.592(0.111)
INTC	36.946	38.068	1.076(0.094)	−0.343(0.106)	−1.297(0.173)
IP	6.569	25.008	1.261(0.063)	0.042(0.074)	0.847(0.099)
JNJ	20.843	24.487	0.593(0.052)	−0.840(0.073)	−0.958(0.099)
JPM	15.317	33.808	1.999(0.094)	0.018(0.108)	1.383(0.137)
KO	25.139	24.056	0.763(0.056)	−0.964(0.070)	−0.619(0.086)
MCD	18.858	24.841	0.801(0.062)	−0.436(0.079)	−0.289(0.097)
MMM	8.797	20.664	0.799(0.047)	−0.360(0.063)	0.139(0.080)
MO	18.525	27.321	0.719(0.065)	−0.593(0.074)	−0.545(0.107)
MRK	20.475	25.262	0.620(0.055)	−0.790(0.065)	−0.992(0.092)
MSFT	45.455	35.043	0.889(0.070)	−0.430(0.087)	−1.603(0.129)
PG	21.597	23.243	0.765(0.051)	−0.723(0.069)	−0.413(0.082)
SBC	16.977	23.078	0.840(0.049)	−0.669(0.073)	0.410(0.098)
T	13.573	24.522	0.842(0.061)	−0.466(0.079)	−0.162(0.097)
UTX	16.922	23.731	1.190(0.054)	0.092(0.073)	0.447(0.095)
WMT	23.745	28.419	1.033(0.059)	−0.686(0.082)	−0.496(0.101)
XOM	12.199	19.795	0.589(0.054)	−0.695(0.069)	0.132(0.081)

Note: The table reports the mean and standard deviation (in annualized percentage points) of daily DJIA stock returns over the period January 1, 1989–December 31, 1998, together with coefficient estimates and heteroskedasticity-consistent standard errors from the TF model given in (18).

the period 1999–2003 than during the years 1989–1998. We return to these estimates below.

The increase in volatility mentioned above is also evident from Table 3, which shows summary statistics of the distribution of daily stock return variances, covariances and correlations. It is seen that the average daily stock return variance doubled during the last five years of the sample period,

Table 2. Characteristics of Daily Returns on DJIA Stocks – 1999–2003.

	Mean	S.D.	$\beta_{i,\mathrm{M}}$	$\beta_{i,\mathrm{SMB}}$	$\beta_{i,\mathrm{HML}}$
AA	10.981	42.868	1.513(0.072)	0.073(0.113)	1.426(0.147)
AXP	6.821	41.112	1.424(0.072)	−0.635(0.100)	0.343(0.123)
BA	3.245	38.566	1.183(0.110)	0.107(0.123)	0.967(0.176)
C	14.649	39.029	1.477(0.086)	−0.571(0.100)	0.401(0.146)
CAT	9.088	37.245	1.169(0.068)	−0.243(0.114)	0.870(0.130)
DD	−6.524	34.531	1.112(0.070)	−0.362(0.105)	1.036(0.116)
DIS	−5.558	42.299	1.290(0.112)	0.158(0.122)	0.564(0.175)
EK	−22.289	38.188	0.964(0.088)	0.078(0.145)	0.624(0.155)
GE	−3.011	35.971	1.212(0.066)	−0.556(0.078)	0.069(0.108)
GM	−6.686	38.084	1.471(0.081)	−0.041(0.097)	1.147(0.139)
HD	−1.170	45.755	1.269(0.082)	−0.425(0.125)	0.271(0.154)
HON	−7.353	46.291	1.456(0.103)	−0.232(0.123)	0.815(0.169)
HPQ	−3.961	54.846	1.340(0.088)	0.362(0.130)	−0.509(0.159)
IBM	−0.465	40.505	0.896(0.062)	−0.330(0.099)	−0.496(0.112)
INTC	1.816	58.322	1.419(0.092)	−0.240(0.137)	−0.888(0.165)
IP	−1.423	37.741	1.204(0.069)	−0.167(0.107)	1.211(0.124)
JNJ	4.101	28.501	0.458(0.071)	−0.629(0.088)	0.132(0.119)
JPM	−6.019	44.804	1.678(0.102)	−0.407(0.121)	0.550(0.180)
KO	−7.716	30.837	0.404(0.062)	−0.685(0.104)	0.244(0.107)
MCD	−8.982	33.925	0.575(0.068)	−0.404(0.099)	0.303(0.113)
MMM	15.600	29.008	0.854(0.058)	−0.365(0.082)	0.620(0.115)
MO	−2.553	39.104	0.512(0.089)	−0.355(0.132)	0.603(0.166)
MRK	−9.816	32.048	0.584(0.064)	−0.722(0.095)	0.161(0.121)
MSFT	−6.231	43.732	1.108(0.068)	−0.067(0.136)	−0.584(0.128)
PG	1.997	33.769	0.320(0.077)	−0.647(0.100)	0.120(0.113)
SBC	−16.382	39.224	0.935(0.085)	−0.504(0.115)	0.487(0.139)
T	−28.409	47.886	1.192(0.093)	−0.189(0.142)	0.537(0.171)
UTX	9.378	38.708	1.389(0.174)	−0.083(0.147)	1.087(0.258)
WMT	7.921	37.369	0.811(0.062)	−0.813(0.111)	−0.102(0.120)
XOM	0.154	27.661	0.776(0.059)	−0.304(0.091)	0.757(0.108)

Note: The table reports the mean and standard deviation (in annualized percentage points) of daily DJIA stock returns over the period January 1, 1999–December 31, 2003, together with coefficient estimates and heteroskedasticity-consistent standard errors from the TF model given in (18).

from just over 3% to more than 6%. The same holds for the average covariance between the 30 DJIA stocks, which increased from 0.83 to 1.69. Interestingly, the standard deviation of the variances also doubled, while the standard deviation of the covariances even tripled. This leads to the conclusion that probably the average correlation did not change much, but the spread in the correlations increased considerably. Indeed, as shown in the

Table 3. Distributions of Variances, Covariances and Correlations of Daily Returns on DJIA Stocks.

	Variances	Covariances	Correlations
Panel A: 1989–1998			
Mean	3.045	0.834	0.282
S.D.	1.041	0.279	0.066
Minimum	1.549	0.353	0.157
10th percentile	1.876	0.547	0.209
25th percentile	2.370	0.646	0.235
Median	2.833	0.806	0.271
75th percentile	3.287	0.959	0.316
90th percentile	4.656	1.145	0.369
Maximum	5.728	2.610	0.569
Panel B: 1999–2003			
Mean	6.127	1.686	0.280
S..D.	2.271	0.908	0.119
Minimum	2.947	0.230	0.032
10th percentile	3.247	0.729	0.139
25th percentile	4.591	0.971	0.191
Median	5.744	1.512	0.268
75th percentile	7.139	2.247	0.353
90th percentile	8.260	2.879	0.436
Maximum	13.111	6.228	0.751

final column of Table 3, the average correlation in both subperiods was equal to 0.28, but the standard deviation of the correlations jumped from 0.066 to 0.12.

4.2. Models

For the implementation of all conditional correlation models considered (except RiskMetrics (RM) as discussed below), we use the asymmetric GJR-GARCH model of Glosten et al. (1993) to describe the conditional volatilities of the individual daily stock returns. In this model, the conditional volatility of r_{it} evolves according to

$$h_{it} = \omega_i + \alpha_i r_{i,t-1}^2 + \gamma_i r_{i,t-1}^2 \mathrm{I}[r_{i,t-1} < 0] + \beta_i h_{i,t-1} \tag{19}$$

allowing positive and negative lagged returns to have a different impact on current volatility. Typically, $\gamma_i > 0$ for stock returns, such that negative returns have a larger effect on volatility than positive returns of the same magnitude. Results obtained with symmetric GARCH(1,1) and with asymmetric power GARCH (APGARCH) volatility models are qualitatively similar to the ones reported below, and are available upon request.

We examine two choices for the conditioning variable in the nonparametric correlation estimator (15): the logarithm of the contemporaneous daily market volatility and the one-day lagged weekly market return, denoted as $\log(h_{\mathrm{M},t})$ and $r_{\mathrm{M},w,t\text{-}1}$, respectively. As explained in Section 1, these variables are motivated by the empirical observations that correlations increase when market volatility is high and returns are low. Market returns are measured by the returns on a value-weighted portfolio (VWP) of all NYSE, AMEX and NASDAQ stocks, again taken from Kenneth French's website. We are able to employ the logged contemporaneous market volatility $\log(h_{\mathrm{M},t})$, as conditioning variable x_t and still use the model for forecasting purposes by estimating a GJR-GARCH model (19) for daily market returns and obtaining the fitted value $\hat{h}_{\mathrm{M},t}$. Note that this can be constructed using only information dated $t-1$ and earlier.

We estimate our semi-parametric correlation model for different choices of the tuning parameter a, to examine the sensitivity of the correlation estimates with respect to the use of local bandwidths, as discussed in Section 3. To make the nonparametric estimates comparable for different values of a, we set the constant b equal to 1.50 for $a=0$, and determine the value of b for other values of a such that the average bandwidth $T^{-1}\Sigma_{t=1}^{T}\, h(x_t)$ is the same for all values of a. Throughout we use the quartic kernel function $K(u)=\frac{15}{16}(1-u^2)^2 I(|u|\leq 1)$ and a standard kernel density estimator with plug-in bandwidth for the density of x_t.

For comparison, we also estimate the CCC model with $R_t = R$, the standard DCC model (5) and (6), the asymmetric DCC model (8), and the STCC model (10) also using the logged contemporaneous daily market volatility and one-day lagged weekly market return as transition variable x_t in the logistic function (11). We also consider the RM model, which specifies the conditional covariance matrix as

$$H_t = (1-\lambda)r_{t-1}r_{t-1}' + \lambda H_{t-1}$$

with $\lambda = 0.94$. Finally, following Chan, Karceski, and Lakonishok (1999) we implement single-factor (SF) and three-factor (TF) models. The TF model uses the excess return on the value-weighted market portfolio and the size and book-to-market factor returns, and effectively boils down to implementing (18) with the raw stock return r_{it} as the dependent variable. The conditional covariance matrix of the stock returns is then given by

$$H_t = B\Omega_t B' + S_t \tag{20}$$

where B is the matrix of factor loadings of the stocks, Ω_t is the conditional covariance matrix of the factors, and S_t is a diagonal matrix containing the idiosyncratic conditional variances. For convenience, we assume that the factors are uncorrelated, such that Ω_t also is diagonal, and that the conditional variance of each factor can be adequately described by means of a GJR-GARCH model (19). The same model is also used for the idiosyncratic variances in S_t. The SF model only includes the excess market return, that is β_{SMB} and β_{HML} in (18) are set equal to 0.

4.3. Estimation Results

Figs. 1 and 2 plot the average correlation between the DJIA stocks together with the 10th and 90th percentiles obtained from the for the semi-parametric models with the contemporaneous log-volatility $\log(h_{\text{M},t})$ of the value-weighted market portfolio and with the lagged weekly market return $r_{\text{M},w,t-1}$, respectively, estimated with daily returns over the period 1989–1998. From Fig. 1, it is seen that for all values of a considered, the average correlation behaves similarly and increases with lagged volatility from 0.25 to 0.50, consistent with the SF model as discussed before. The increase in correlations for high volatility is more pronounced for small values of a. Turning to Fig. 2, we observe that the average correlation appears to increase for negative-lagged market returns, while it remains fairly constant for positive values of $r_{\text{M},w,t-1}$, in particular for $a \leq 0.50$. This corresponds quite well with the patterns documented in Longin and Solnik (2001) and Ang and Chen (2002). For $a = 0.75$, the correlations increase for large positive index returns as well, albeit the effect of negative index returns is stronger.

Panels (a) and (b) of Fig. 3 display the average correlation and the 10th and 90th percentiles over time for the semi-parametric models with the two choices of x_t and $a = 0.50$ in both cases. It appears that using market volatility as conditioning variable leads to more sizable changes in the conditional correlations than using the market return. At the same time however, the average conditional correlation based on the model with $x_t = \log(h_{\text{M},t})$ evolves more smoothly, with the average correlation based on the model with $x_t = r_{\text{M},w,t-1}$ showing more erratic short-run behaviour.

Panels (c) and (d) of Fig. 3 show the average correlation and the 10th and 90th percentiles based on the DCC model and the SF model, respectively. For the DCC model, the average correlation varies very little over time. In contrast, the conditional correlations implied by the SF model vary

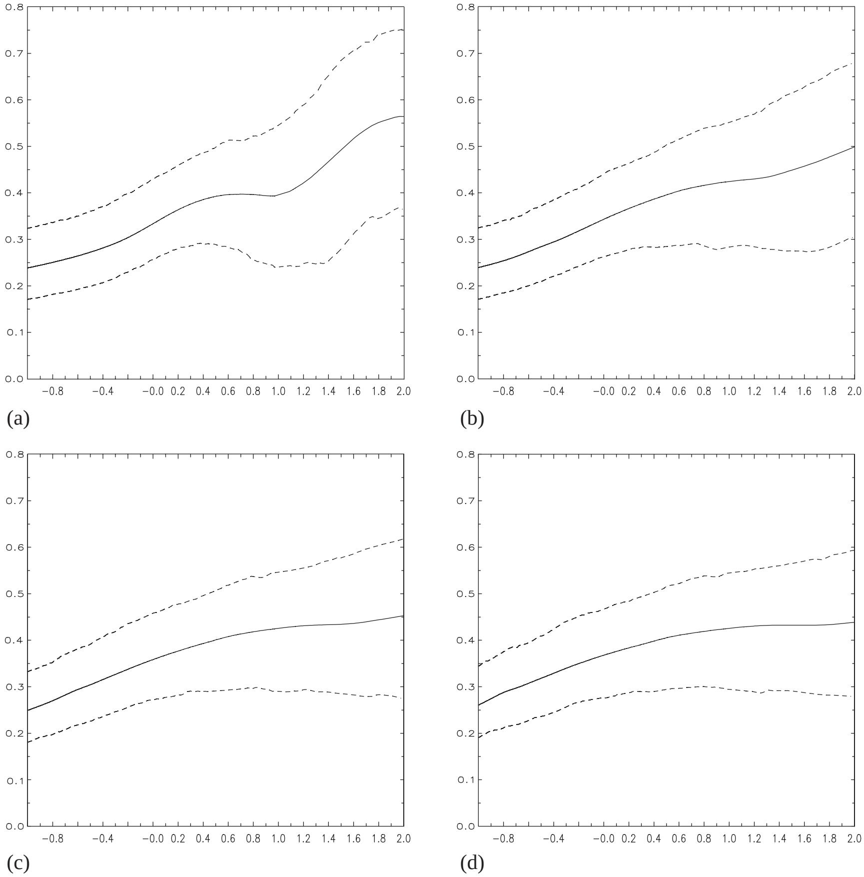


Fig. 1. Average Correlation Between Dow Jones Stocks (solid line), and 10th and 90th Percentiles (Dashed Lines) Obtained from the Nonparametric Estimator (14) with x_t being the Contemporaneous Daily Log-Volatility on the Value-Weighted Market Portfolio $\log(h_{M,t})$, and Bandwidth $h(x) = bf(x)^{-a}$, Estimated Using Daily Returns over the Period January 1, 1989–December 31, 1998, Plotted as a Function of $\log(h_{M,t})$. (a) $a = 0$; (b) $a = 0.25$; (c) $a = 0.50$; (d) $a = 0.75$.

considerably. This is a direct consequence of the SF structure together with the use of a GJR-GARCH model for the conditional variance of the factor. In the SF model with the excess market return as the single factor, B in (20) reduces to a vector consisting of the stocks' betas $\beta_{j,\mathrm{M}}$, while $\Omega_t = h_{\mathrm{M},t}$ is a

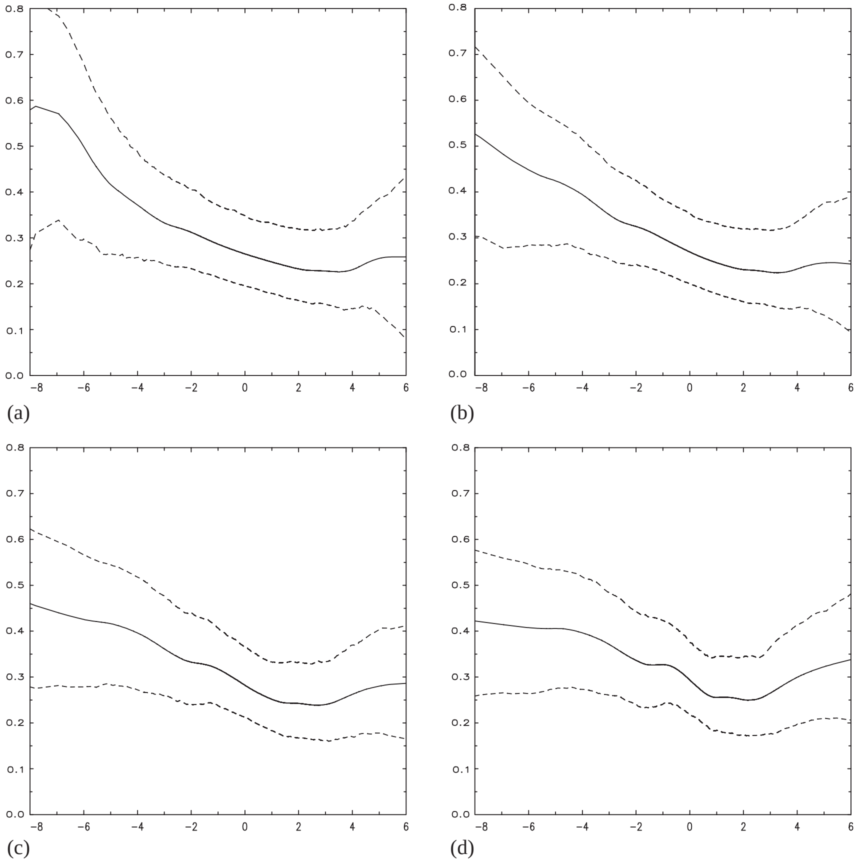
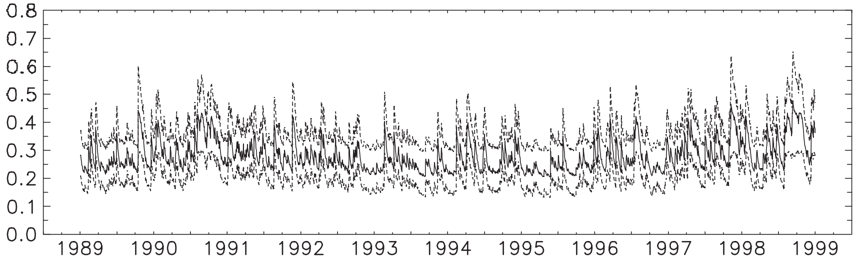


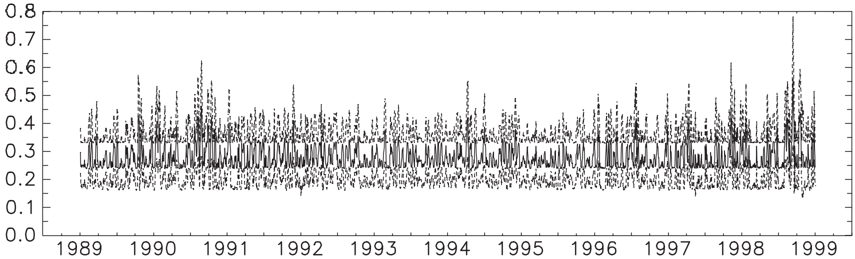
Fig. 2. Average Correlation Between Dow Jones Stocks (Solid Line), and 10th and 90th Percentiles (Dashed Lines) Obtained from the Nonparametric Estimator (14) with x_t being the One-Day Lagged Weekly Return on the Value-Weighted Market Portfolio $r_{\mathrm{M},w,t-1}$ and Bandwidth $h(x) = bf(x)^{-a}$, Estimated Using Daily Returns over the Period January 1, 1989–December 31, 1998, Plotted as a Function of $r_{\mathrm{M},w,t-1}$. (a) $a = 0$; (b) $a = 0.25$; (c) $a = 0.50$; (d) $a = 0.75$.

scalar. The conditional correlation between stocks i and j then is equal to

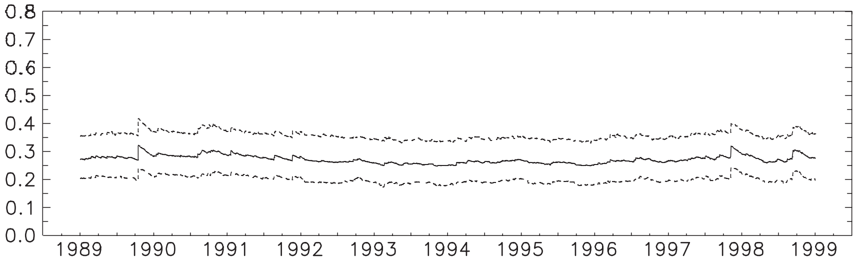
$$\rho_{ijt} = \frac{\beta_{i,\mathrm{M}}\beta_{j,\mathrm{M}}h_{\mathrm{M},t}}{\sqrt{\beta_{i,\mathrm{M}}^2 h_{\mathrm{M},t} + s_{it}^2}\sqrt{\beta_{j,\mathrm{M}}^2 h_{\mathrm{M},t} + s_{jt}^2}}$$


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999

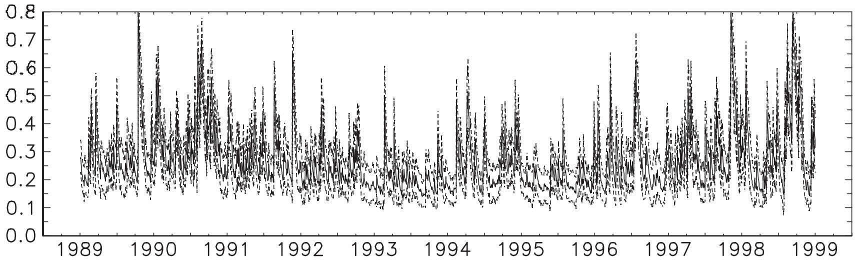
(a)


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999

(b)


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999

(c)


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999

(d)

where s_{it}^2 is the idiosyncratic conditional variance of stock i. As shown by Andersen et al. (2001), it follows that $\partial\rho_{ijt}/\partial h_{\mathrm{M},t} > 0$ if $\beta_{i,\mathrm{M}}, \beta_{j,\mathrm{M}}, h_{\mathrm{M},t} > 0$. Hence, as long as the market betas of two stocks are both positive, an increase in market volatility will lead to an increase in their conditional correlation. Given that all betas for the DJIA stocks are positive and not too different from each other, see Table 1, all pairwise conditional correlations increase when $h_{\mathrm{M},t}$ becomes higher. Additionally, the GJR-GARCH model for the market portfolio returns implies considerable variation in $h_{\mathrm{M},t}$, leading to the substantial changes in the conditional correlations as seen in Fig. 3.

Panel (b) of Fig. 4 explicitly shows how the conditional correlations are related to the (logged) market volatility. Note that the increase in the average correlation is much more pronounced than in the corresponding semi-parametric model, compare Fig. 1. Panel (a) of Fig. 4 shows that the average correlation from the DCC model is not related to the market volatility at all. Similarly, in panel (c) of this figure a rather weak quadratic relation between the average correlation from the DCC model and the lagged weekly market return is visible, but much less pronounced than in the semi-parametric model. For the SF model, we do find a strong relationship between the conditional correlations and the lagged weekly market return, of the same form as found for the semi-parametric model in Fig. 2 and supporting the "correlation breakdown" effect with a much larger increase in correlation for large negative market returns than for large positive ones. At first sight this may seem surprising, but in fact it is a direct consequence of the use of a GJR-GARCH model for the market volatility. In that model, large past returns lead to high current volatility, while the parameter estimates are such that negative returns increase volatility more than positive returns of the same magnitude. The higher volatility in turn increases the correlations between the stocks.

Finally, Figs. 5 and 6 summarize the estimation results for the STCC model with $x_t = \log(h_{\mathrm{M},t})$ and $r_{\mathrm{M},w,t-1}$, respectively. The STCC models also

---

Fig. 3. Average Correlation Between Dow Jones Stocks (Solid Line), and 10th and 90th Percentiles (Dashed Lines) Obtained from the Nonparametric Estimator (14) with x_t being the Contemporaneous Daily log-Volatility on the Value-Weighted Market Portfolio $\log(h_{\mathrm{M},t})$ and the One-Day Lagged Weekly Return on the Value-Weighted Market Portfolio $r_{\mathrm{M},w,\mathrm{t}-1}$, from the DCC Model (5), and from the SF Model, Estimated Using Daily Returns over the Period January 1, 1989–December 31, 1998. (a) SEMI-$\log(h_{M,\mathrm{t}})$; (b) SEMI-$r_{M,\mathrm{w},\mathrm{t}-1}$; (c) DCC; (d) SF.

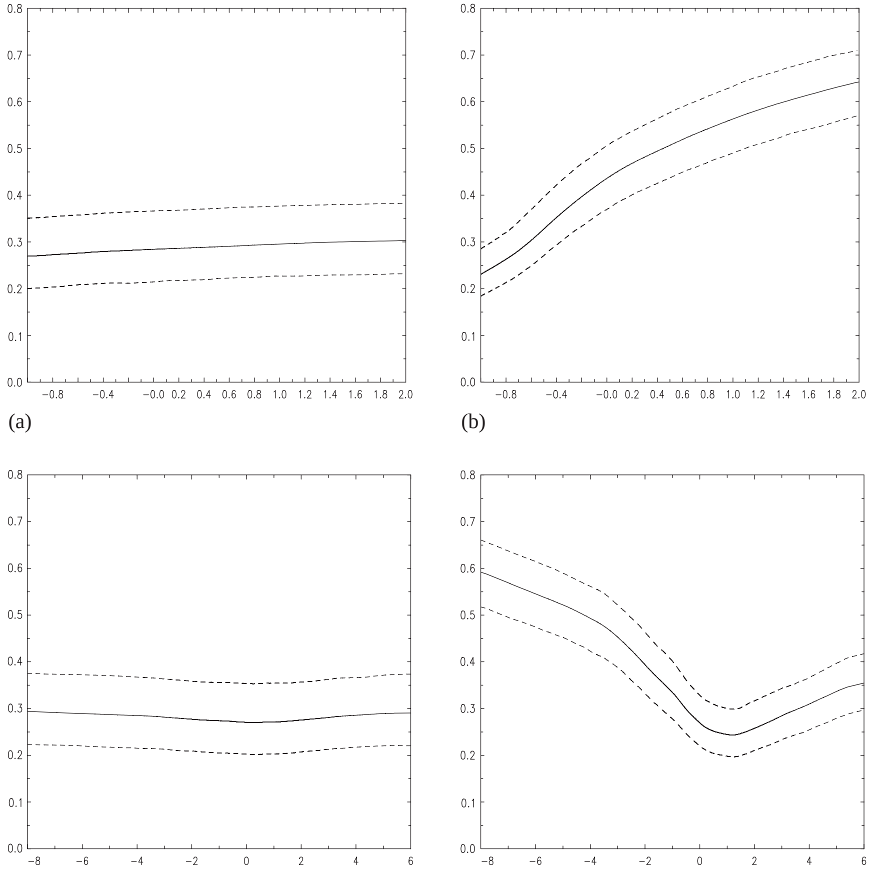


Fig. 4. (Smoothed) Average Correlation Between Dow Jones Stocks (Solid Line), and 10th and 90th Percentiles (Dashed Lines) Obtained from the DCC Model (5) and the SF Model, Estimated Using Daily Returns over the Period January 1, 1989–December 31, 1998, Plotted as a Function of the Contemporaneous Daily Log-Volatility on the Value-Weighted Market Portfolio $\log(h_{M,t})$ and the One-Day Lagged Weekly Return on the Value-Weighted Market Portfolio $r_{M,w,t-1}$. (a) DCC-$\log(h_{M,t})$; (b) SF-$\log(h_{M,t})$; (c) DCC-$r_{M,w,t-1}$; (d) SF-$r_{M,w,t-1}$.

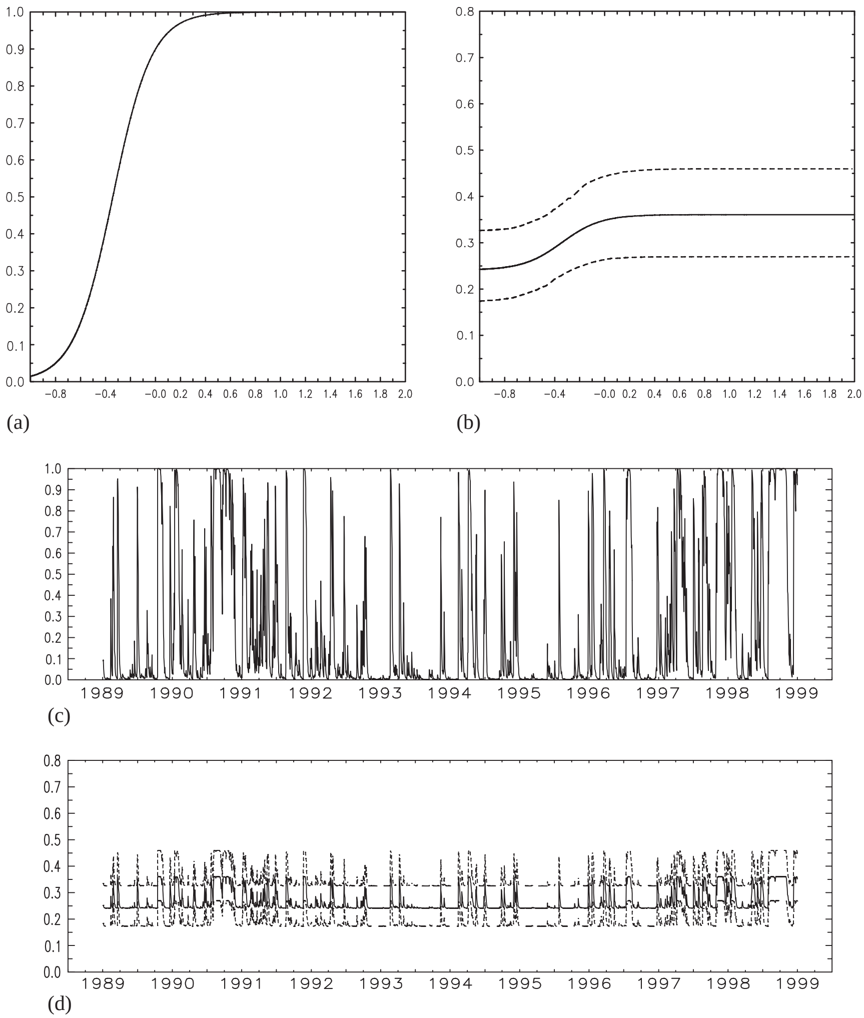


Fig. 5. Logistic Transition Function $G(x_t; \gamma, c)$ as in (11), and Average Correlation Between Dow Jones Stocks (Solid Line), and 10th and 90th Percentiles (Dashed Lines) Obtained from the STCC Model (10) with x_t being the Contemporaneous Daily Log Volatility on the Value-Weighted Market Portfolio $\log(h_{M,t})$, Estimated Using Daily Returns over the Period January 1, 1989–December 31, 1998. (a) Transition function against x_t; (b) Correlation quantiles against x_t; (c) Transition function over time; (d) Correlation quantiles over time.

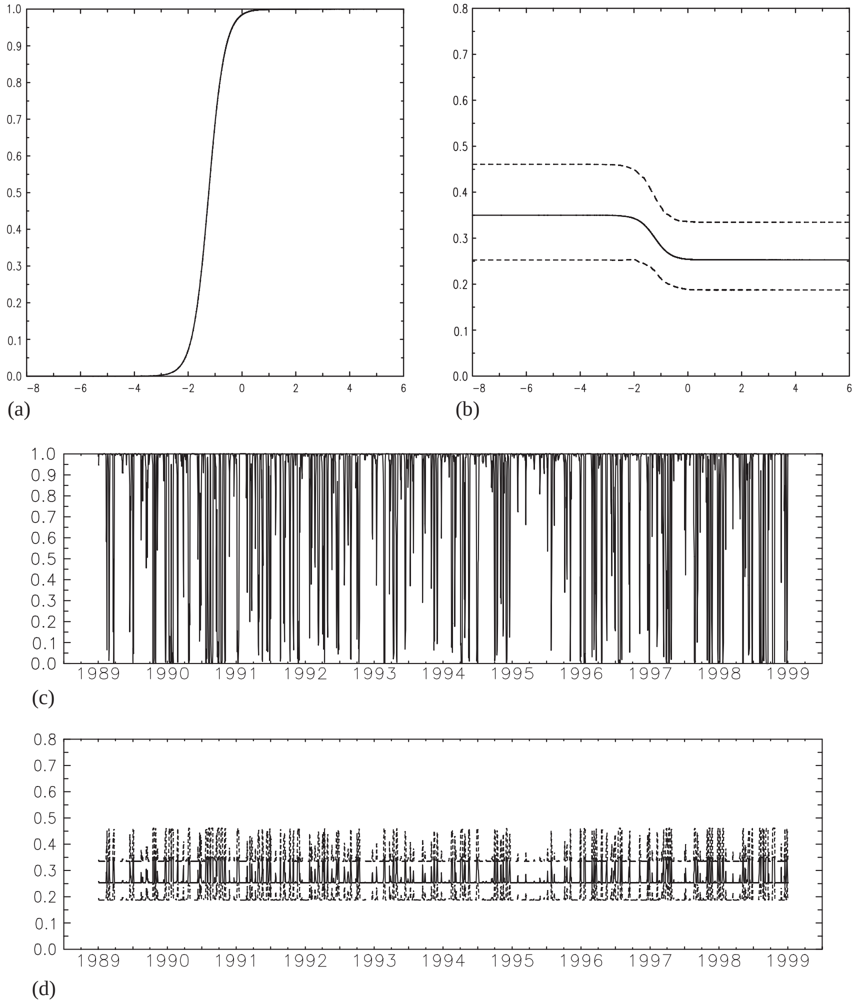


Fig. 6. Logistic Transition Function $G(x_t; \gamma, c)$ as in (11), and Average Correlation Between Dow Jones Stocks (Solid Line), and 10th and 90th Percentiles (Dashed Lines) Obtained from the STCC Model (10) with x_t being the One-Day Lagged Weekly Return on the Value-Weighted Market Portfolio $r_{\mathrm{M,w},t-1}$, Estimated Using Daily Returns over the Period January 1, 1989–December 31, 1998. (a) Transition function against x_t; (b) Correlation quantiles against x_t; (c) Transition function over time; (d) Correlation quantiles over time.

imply that conditional correlations are higher in case of high market volatility and in case of low market returns. However, note that by construction, the model implies that the conditional correlations only change monotonically from one level to another as the logistic transition function (11) changes from 0 to 1. Panel (c) of these figures plot the value of the transition functions, suggesting that the correlations spend most of the time in the low-volatility and positive market return regimes, where the conditional correlations are at their lower level.

4.4. Minimizing Portfolio Variance Using Correlation Forecasts

We follow the recommendation of Engle and Sheppard (2001) to judge the adequacy of the dynamic correlation models by examining certain characteristics of stock portfolios that are constructed based on covariance matrix forecasts from the models. In particular, we consider the global MVP, which is often used for judging the goodness of fit of multivariate volatility models, see for example, Chan et al. (1999). The MVP weights are given by $w_t = \hat{H}_t^{-1}\iota/(\iota'\hat{H}_t^{-1}\iota)$, where $\hat{H}_t$ is the one-step ahead forecast of the conditional covariance matrix constructed at time $t-1$, and ι is a $(N \times 1)$ vector of ones. Note that the MVP weights only depend on forecasts of the conditional covariance, such that forecasting expected returns, which is known to be notoriously difficult, is avoided. For each MVP based on the different covariance models we consider the average return, standard deviation, Sharpe ratio and tracking error relative to the S&P 500 index. The latter is defined as the square root of the mean squared difference between the portfolio's return and the S&P 500 return. For comparison purposes, we also construct equally weighted and value-weighted portfolios and the MVP. For the equally weighted portfolio (EWP), the portfolio weights are constant and equal to $w_t = N^{-1}\iota$, where, the time-varying weights for the VWP are obtained as $w_t = w_{t-1}(1 + r_{t-1})/(1 + r_t)'\iota$, starting with an EWP at t = 0, cf. Engle and Sheppard (2001).

The EWP and VWP provide reasonable benchmarks to assess the extent to which optimization actually helps in reducing portfolio variance. In addition, together with the MVP they can be used to further evaluate the covariance models by considering additional portfolio characteristics. First, we compute the variance of the standardized portfolio returns $r_{p,t} = w_t'r_t/\sqrt{w_t'\widehat{H}_t w_t}$. If the multivariate conditional covariance model is correctly specified, the variance of $r_{p,t}$ should be equal to 1, for any choice of weights

w_t. Second, we compute one-day Value-at-Risk (VaR) forecasts as $\hat{\sigma}_p z_q$, where $\hat{\sigma}_p$ is the one-step ahead forecast of the portfolio standard deviation, $\hat{\sigma}_p = \sqrt{w_t' \widehat{H}_t w_t}$. In contrast to Engle and Sheppard (2001), for example, we do not use the q-th quantile of the standard normal distribution for z_q. Rather, we employ the quantiles of the standardized in-sample portfolio returns obtained using the relevant weights w_t. We consider one-day VaR forecasts at $100 \times q = 1\%$, 5% and 10%. The accuracy of the VaR forecasts is assessed by means of the Dynamic Quantile test of Engle and Manganelli (2004). For that purpose, define the binary variable HIT_t such that $HIT_t = 1$ if the portfolio return is below the VaR forecast and 0 otherwise. Under the null hypothesis of a correctly specified model for the conditional covariance matrix and hence for the VaR forecasts, the HIT_t variable should have mean q and should be independent from all information available at $t-1$, including lagged HIT_t's and the VaR forecast for time t. This can be tested by constructing an F-statistic for the null hypothesis $\delta_0 = \cdots = \delta_{l+1} = 0$ in the auxiliary regression

$$HIT_t - q = \delta_0 + \delta_1 HIT_{t-1} + \cdots + \delta_l HIT_{t-l} + \delta_{l+1} VaR_t + e_t$$

where we set $l = 5$.

Table 4 summarizes the in-sample MVP results over the period January 1, 1989–December 31, 1998, which suggest several conclusions. First, compared to the EWP and VWP, all models except Riskmetrics achieve in producing an MVP with lower standard deviation, although the reduction in volatility is quite modest. The best performing model is our semi-parametric model using the lagged weekly market return as conditioning variable, with an annualized standard deviation of 12.54% compared with 15.16% for the EWP. Second, minimizing portfolio variance comes at the cost of a sharp reduction in the portfolio's average return, such that the Sharpe ratios of the MVPs is considerably lower than the Sharpe ratios of the EWP and VWP. Third, the conditional correlation models outperform the SF and TF models. Hence, it seems worthwhile to model the dynamics in the covariance structure of the individual stocks explicitly, rather than indirectly by means of a factor approach. Fourth, the performance of the different conditional correlation models is very similar, with the largest difference in standard deviation of the MVP returns being less than 0.3%.

The conclusion that it is difficult to distinguish between the competing models based on in-sample measures of fit also emerges from Table 5. The variance of the standardized portfolio returns shows that actually all models perform quite badly in terms of the volatility of the MVP returns, which is

Table 4. In-Sample Performance and Characteristics of Minimum Variance Portfolios Based on Forecasting Models – 1989–1998.

	Mean	S. D.	Sharpe Ratio	Tracking Error	β_{M}	β_{SMB}	β_{HML}
SF	9.24	13.42	0.306	10.89	0.675(0.037)	−0.420(0.051)	0.287(0.057)
TF	8.51	13.80	0.245	11.20	0.806(0.041)	−0.155(0.057)	0.374(0.059)
RM	13.81	17.82	0.487	15.53	0.806(0.044)	−0.155(0.069)	0.256(0.077)
CCC	11.78	12.82	0.519	8.17	0.791(0.023)	−0.308(0.028)	0.170(0.037)
DCC	11.57	12.73	0.506	8.18	0.786(0.023)	−0.301(0.029)	0.172(0.037)
ADCC	11.19	12.72	0.476	8.25	0.785(0.023)	−0.299(0.029)	0.176(0.037)
STCC(*h*)	11.40	12.71	0.492	8.42	0.780(0.024)	−0.305(0.030)	0.195(0.039)
STCC(*r*)	10.19	12.69	0.398	8.39	0.775(0.023)	−0.310(0.030)	0.187(0.039)
SEMI(*h*)	10.53	12.63	0.427	8.71	0.765(0.024)	−0.300(0.030)	0.212(0.039)
SEMI(*r*)	10.11	12.54	0.397	8.57	0.761(0.025)	−0.310(0.030)	0.206(0.038)
EWP	17.99	15.16	0.848	4.16	1.072(0.013)	−0.333(0.017)	−0.053(0.021)
VWP	21.52	17.95	0.913	6.83	1.114(0.017)	−0.410(0.021)	−0.458(0.027)

Note: The table reports in-sample (January 1, 1989–December 31, 1998) summary statistics for the minimum variance portfolios based on the SF , TF , RM, CCC, DCC, ADCC, STCC model with x_t equal to the daily log-volatility for the market return *(h)* or the one-day lagged weekly index return (*r*), and the SEMI model with the same choices of x_t. In addition, summary statistics are reported for the EWPs and VWPs. For each MVP, the table shows the mean and standard deviation (in annualized percentage points) of the portfolio return, the annualized Sharpe ratio, the annualized tracking error relative to the S&P 500 index, and coefficient estimates and standard errors from the Fama–French TF model (18).

consistently underestimated. In contrast, the variances of the standardized EWP and VWP returns are quite close to the theoretical value of one for all models except perhaps RM. In terms of VaR forecasts, results also are similar across models (and by construction, violation frequencies are identical for the EWP). The VaR(q) forecasts are violated too frequently for the MVP for all three levels of q, which aligns with the underestimation of the MVP volatility. The VaR violation frequencies are approximately correct for the EWP and VWP portfolios. The results of the Dynamic Quantile test confirm these observations, except that the small p-values for the test applied to the 10% VaR forecasts for the EWP are perhaps surprising, given that the violation frequency is close to perfect.

Next, we examine the out-of-sample forecasting performance of the models over the period January 1, 1999–December 31, 2003. For this purpose, we re-estimate all models after every 20 observations using a 10-year moving window. We examine the same portfolio characteristics as before, except that now we employ genuine out-of-sample forecasts of the conditional covariance matrix.

Table 5. Correlation Models for Dow Jones Stocks: In-Sample Results.

	SF	TF	RM	CCC	DCC	ADCC				Semi-Parametric Model							
							STCC, $x_t =$			$x_t = \log(h_{M,t})$				$\boldsymbol{x_t = r_{\mathbf{M},w,t\text{-}1}}$			
							$\log(h_{\mathrm{M,t}})$	$r_{M,w,t-1}$	$a =$	0	0.25	0.50	0.75	0	0.25	0.50	0.75
Portfolio standard deviations																	
MVP	1.142	1.078	2.678	1.097	1.086	1.089	1.096	1.097		1.096	1.091	1.086	1.081	1.087	1.091	1.086	1.073
EWP	1.017	0.938	1.073	0.993	0.989	0.988	0.990	0.992		1.000	0.991	0.970	0.952	1.003	0.999	0.982	0.956
VWP	1.024	0.905	1.070	1.007	1.001	1.001	1.005	1.005		1.014	1.005	0.985	0.968	1.017	1.012	0.994	0.968
MVP – VaR violations																	
1%	1.19	1.15	12.78	1.27	1.23	1.31	1.35	1.43		1.35	1.31	1.35	1.42	1.54	1.58	1.54	1.62
5%	5.22	5.26	22.71	5.86	5.82	5.94	6.02	5.82		5.94	5.86	6.13	6.02	5.78	5.86	6.21	6.41
10%	10.29	10.25	27.58	11.24	11.28	11.32	11.36	11.40		11.59	11.59	11.36	11.32	11.52	11.52	11.52	11.32
MVP: *p*-value of dynamic quantile statistic																	
1%	0.267	0.037	0.000	0.041	0.034	0.046	0.037	0.018		0.035	0.044	0.044	0.043	0.048	0.045	0.045	0.019
5%	0.759	0.410	0.000	0.038	0.054	0.061	0.059	0.137		0.084	0.089	0.050	0.047	0.056	0.065	0.041	0.037
10%	0.676	0.878	0.000	0.030	0.048	0.029	0.068	0.038		0.045	0.045	0.065	0.046	0.030	0.039	0.033	0.055
EWP – VaR violations																	
1%	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95		0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
5%	4.94	4.94	4.94	4.94	4.94	4.94	4.94	4.94		4.94	4.94	4.94	4.94	4.94	4.94	4.94	4.94
10%	9.97	9.97	9.97	9.97	9.97	9.97	9.97	9.97		9.97	9.97	9.97	9.97	9.97	9.97	9.97	9.97

EWP: p-value of dynamic quantile statistic																
1%	0.918	0.918	0.085	0.795	0.7999	0.7999	0.708	0.418	0.769	0.7999	0.735	0.758	0.461	0.494	0.484	0.379
5%	0.585	0.585	0.153	0.073	0.072	0.076	0.309	0.098	0.291	0.735	0.601	0.640	0.096	0.131	0.186	0.148
10%	0.672	0.803	0.109	0.024	0.027	0.035	0.069	0.027	0.040	0.053	0.093	0.098	0.020	0.030	0.040	0.046
VWP – VaR violations																
1%	0.87	0.87	1.03	1.19	1.07	1.07	0.99	0.95	0.99	0.95	0.95	0.95	0.87	0.91	0.91	0.91
5%	4.59	4.67	5.14	5.03	5.03	4.9	4.95	5.03	4.91	4.79	4.87	4.83	5.07	5.07	5.10	5.07
10%	10.57	10.61	10.09	10.21	10.37	10.37	10.53	10.41	10.45	10.49	10.37	10.37	10.33	10.37	10.33	10.41
VWP: p-value of dynamic quantile statistic																
1%	0.093	0.089	0.136	0.162	0.766	0.774	0.819	0.658	0.927	0.885	0.660	0.689	0.721	0.844	0.501	0.743
5%	0.528	0.715	0.319	0.241	0.122	0.171	0.500	0.177	0.368	0.459	0.516	0.570	0.176	0.137	0.149	0.123
10%	0.478	0.421	0.146	0.103	0.122	0.122	0.154	0.190	0.196	0.183	0.206	0.211	0.186	0.244	0.192	0.190

Note: The table reports in-sample results for DCC models for Dow Jones stocks, estimated using daily returns over the period January 1, 1989–December 31, 1998. Results are based on the SF, TF, RM, CCC,DCC, ADCC, STCC model with x_t equal to the logged contemporaneous market volatility ($\log(h_{\mathrm{M},t})$) or the one-day lagged weekly market portfolio return ($r_{\mathrm{M},w,t-1}$), and the semi-parametric conditional correlation SEMI model with the same choices of x_t. MVP denotes minimum variance portfolio, EWP and VWP indicate equally weighted and value-weighted portfolios, respectively. Portfolio standard deviations refer to the in-sample standard deviation of standardized portfolio returns. VaR violation percentages and Dynamic Quantile statistics for the corresponding HIT-sequences concern one-day ahead in-sample forecasts.

The MVP results in Table 6 resemble the results in Table 4, in the sense that most conclusions drawn for the in-sample MVP results continue to hold out-of-sample. The only exception is that, in terms of the standard deviation of MVP returns, the DCC-type models now perform slightly better than the STCC and semi-parametric models. In addition, we note that the variance of the MVP during the out-of-sample period is almost double the variance during the in-sample period. From Table 7, we observe that all models continue to underestimate the MVP variance in the out-of-sample period, while the corresponding VaR forecasts are (thus) violated much too frequently. The conditional correlation models perform much better for the EWP and MVP, although the 5% and 10% VaR forecasts for these portfolios are violated too frequently as well.

The most conspicuous finding in the MVP analysis presented above is the similarity in performance of all covariance models for the portfolio's variance. The characteristics of the different MVPs are examined further by estimating the Fama–French TF model as given in (18) using the portfolio's excess return as dependent variable. The resulting estimates, shown in Tables 4 and 6, confirm that the portfolios are comparable in terms of sensitivities to the market return, size and book-to-market factors. In addition, all MVPs (except the ones based on the SF and RM models) must select stocks with low market betas, given that their market betas are close to 0.8 and 0.7 for the in-sample and out-of-sample periods, respectively, compared with market betas close to 1 for the EWP and VWP. The emphasis on stocks with low market betas is also seen in Table 8, which shows the mean and standard deviation of the weights in the MVP based on the semi-parametric conditional correlation model with x_t equal to the daily log-volatility for the market return. During the in-sample period, the MVP is dominated by Exxon Mobil (XOM) with a mean portfolio weight close to 20%. The estimates of $\beta_{i,\mathrm{M}}$ in Table 1 show that indeed this stock had the lowest market beta during this period. Also, note that the stocks of 3M (MMM) and SBC Communications (SBC) have large mean portfolio weights during the in-sample period although their betas are not particularly low compared to the other stocks. The opposite is observed for the stock of Merck & Co (MRK), which has the second-lowest market beta but only an average mean portfolio weight. This illustrates the fact that the portfolio variance cannot only be reduced by putting much weight on low beta stocks, but also by tilting the portfolio towards stocks with low return volatilities. 3M and SBC are typical examples of the latter. Finally, we remark that the regression results of the TF model also shed light on the reason why the MVP variance is so much higher during the out-of-sample period compared

Table 6. Out-of-Sample Performance and Characteristics of Minimum Variance Portfolios Based on Forecasting Models – 1999–2003.

	Mean	S.D.	Sharpe Ratio	Tracking Error	β_{M}	β_{SMB}	β_{HML}
SF	−8.78	19.46	−0.628	23.50	0.509(0.061)	−0.329(0.074)	0.671(0.108)
TF	−12.42	20.46	−0.775	21.75	0.632(0.081)	−0.129(0.086)	0.572(0.137)
RM	−16.39	22.52	−0.881	24.13	0.450(0.056)	−0.243(0.067)	0.170(0.088)
CCC	−7.48	17.51	−0.624	15.97	0.705(0.035)	−0.269(0.048)	0.507(0.064)
DCC	−8.66	17.20	−0.704	15.97	0.677(0.036)	−0.261(0.047)	0.459(0.064)
ADCC	−9.26	17.20	−0.739	16.05	0.674(0.037)	−0.262(0.047)	0.459(0.064)
STCC(*h*)	−8.29	17.69	−0.664	17.09	0.675(0.038)	−60.260(0.049)	0.518(0.067)
STCC(*r*)	−8.87	18.23	−0.712	17.43	0.676(0.037)	−0.261(0.048)	0.487(0.066)
SEMI(*h*)	−8.26	18.11	−0.646	17.38	0.678(0.040)	−60.260(0.051)	0.511(0.070)
SEMI(*r*)	−8.63	18.54	−0.671	16.97	0.677(0.036)	−60.260(0.050)	0.506(0.069)
EWP	−1.98	21.82	−0.249	7.98	1.066(0.024)	−30.307(0.034)	0.425(0.039)
VWP	−4.42	19.67	−0.400	7.53	0.938(0.023)	−0.298(0.029)	0.313(0.037)

Note: The table reports out-of-sample (January 1, 1999–December 31, 2003) summary statistics for the minimum variance portfolios based on various forecasting models. See Table 4 for models definitions and description of the summary statistics shown.

Table 7. Correlation Models for Dow Jones Stocks: Out-of-Sample Results.

	SF	TF	RM	CCC	DCC	ADCC	STCC, $x_t =$			Semi-Parametric Model							
										$x_t = \log(h_{\mathrm{M,t}})$				$x_t = r_{\mathrm{M},w,t-1}$			
							$\log(h_{M,t})$	$rM_{,w,t-1}$	$a =$	0	0.25	0.50	0.75	0	0.25	0.50	0.75
Portfolio standard deviations																	
MVP	1.381	1.195	2.645	1.240	1.208	1.208	1.238	1.554		1.254	1.252	1.252	1.253	1.323	1.308	1.266	1.248
EWP	1.054	0.944	1.034	1.124	1.097	1.095	1.065	1.092		1.060	1.052	1.043	1.037	1.132	1.123	1.107	1.101
VWP	1.054	0.932	1.038	1.135	1.106	1.103	1.072	1.097		1.067	1.059	1.050	1.043	1.107	1.101	1.088	1.081
MVP – VaR violations																	
1%	1.32	0.74	14.32	2.06	2.14	2.06	2.392	2.47		2.39	2.47	2.39	2.39	2.63	2.47	2.31	2.55
5%	6.42	6.58	25.27	9.55	9.63	9.71	9.79	9.79		10.29	10.45	10.45	10.70	9.49	9.79	9.79	10.19
10%	12.10	13.74	30.95	16.71	17.28	17.37	16.95	17.28		17.28	17.04	17.20	17.45	17.28	17.27	17.04	17.68
MVP: p-value of dynamic quantile statistic																	
1%	0.517	0.151	0.001	0.015	0.009	0.007	0.018	0.047		0.028	0.049	0.046	0.046	0.014	0.022	0.055	0.067
5%	0.111	0.071	0.000	0.005	0.002	0.002	0.004	0.014		0.004	0.003	0.002	0.002	0.020	0.012	0.013	0.008
10%	0.061	0.053	0.000	0.006	0.004	0.004	0.005	0.009		0.005	0.006	0.006	0.005	0.006	0.005	0.005	0.004
EWP – VaR violations																	
1%	0.82	0.82	0.74	1.15	0.99	0.99	0.99	1.19		0.91	0.91	0.91	0.91	1.04	1.04	1.11	1.11
5%	4.61	4.61	5.60	7.57	7.49	7.49	7.16	5.81		7.08	6.91	6.75	6.83	5.89	5.97	5.81	5.57
10%	12.35	12.35	11.36	14.98	14.73	14.81	13.91	12.26		14.07	13.99	13.91	14.07	12.58	12.58	12.58	12.42

EWP: *p*-value of dynamic quantile statistic																
1%	0.560	0.546	0.398	0.272	0.498	0.497	0.568	0.565	0.390	0.398	0.408	0.407	0.546	0.573	0.584	0.488
5%	0.375	0.466	0.109	0.111	0.135	0.135	0.193	0.785	0.320	0.332	0.334	0.337	0.518	0.518	0.582	0.676
10%	0.173	0.122	0.107	0.021	0.033	0.026	0.051	0.319	0.039	0.045	0.047	0.041	0.164	0.190	0.208	0.281
VWP – VaR violations																
1%	0.91	0.82	0.58	1.32	1.15	1.19	0.99	0.99	0.99	0.99	0.99	0.99	1.11	1.11	1.11	1.19
5%	5.10	4.86	5.76	7.90	7.41	7.49	7.49	6.05	7.16	7.00	6.83	7.00	6.237	6.29	6.13	5.97
10%	12.51	11.93	11.60	15.47	14.90	14.98	14.07	12.66	14.07	13.99	14.16	14.16	12.66	12.74	12.81	12.50
VWP: *p*-value of dynamic quantile statistic																
1%	0.809	0.657	0.258	0.316	0.173	0.145	0.540	0.772	0.568	0.567	0.563	0.561	0.703	0.713	0.741	0.810
5%	0.266	0.167	0.046	0.051	0.184	0.183	0.232	0.807	0.232	0.249	0.307	0.212	0.559	0.461	0.585	0.680
10%	0.075	0.073	0.087	0.014	0.026	0.028	0.031	0.123	0.043	0.044	0.036	0.036	0.081	0.110	0.116	0.162

Note: The table reports out-of-sample results for dynamic correlation models for Dow Jones stocks over the period January 1, 1999–December 31, 2003. Results are based on the SF, TF, RM, CCC, DCC, ADCC, STCC model with x_t equal to the logged contemporaneous market volatility ($\log(h_{\mathrm{M,t}})$) or the one-day lagged weekly market portfolio return ($r_{\mathrm{M},w,t\text{-}1}$), and the SEMI model with the same choices of x_t. All models are re-estimated every 20 observations using a 10-year moving window of daily return observations. MVP denotes minimum variance portfolio, EWP and VWP indicate equally weighted and value-weighted portfolios, respectively. Portfolio standard deviations are the standard deviation of standardized portfolio return forecasts. VaR violation percentages and Dynamic Quantile statistics for the corresponding HIT-sequences concern one-day ahead out-of-sample forecasts.

Table 8. Summary Statistics for Portfolio Weights.

	Minimum Variance		Minimum Tracking Error	
	1989–1998	1999–2003	1989–1998	1999–2003
AA	4.28(3.57)	0.34(3.67)	2.43(0.79)	1.31(0.74)
AXP	−0.98(3.17)	−1.74(3.83)	1.84(0.69)	2.64(1.20)
BA	3.95(4.43)	3.91(3.94)	2.31(0.63)	1.92(0.70)
CAT	2.94(3.40)	0.51(2.69)	1.96(0.62)	0.86(0.87)
C	−5.15(1.71)	−1.04(4.45)	2.26(0.46)	4.55(1.43)
DD	−0.59(3.82)	−1.12(3.38)	3.01(0.79)	0.93(1.32)
DIS	3.08(3.23)	0.55(3.04)	2.23(0.58)	1.91(0.76)
EK	6.06(3.72)	11.88(8.38)	2.30(0.49)	2.37(1.54)
GE	0.42(8.34)	−6.03(6.29)	6.08(1.61)	5.81(1.74)
GM	1.38(3.05)	4.62(4.64)	2.90(0.77)	3.34(1.44)
HD	−0.03(3.38)	−0.31(4.75)	2.72(0.71)	2.12(1.50)
HON	3.88(4.29)	1.18(5.52)	2.39(0.84)	3.08(2.02)
HPQ	−0.42(1.89)	0.35(2.32)	1.68(0.33)	2.69(0.94)
IBM	7.26(5.77)	6.22(5.24)	4.15(1.56)	4.92(1.41)
INTC	−1.56(1.67)	−2.26(2.06)	2.15(0.37)	4.57(1.38)
IP	4.33(4.26)	4.36(4.97)	2.35(0.66)	2.49(1.39)
JNJ	1.95(4.96)	11.84(8.41)	3.22(1.03)	5.00(1.87)
JPM	2.03(3.93)	−0.28(3.21)	3.91(1.35)	4.57(1.28)
KO	−1.60(4.34)	4.52(6.77)	3.46(1.19)	1.83(1.54)
MCD	5.67(4.37)	5.83(5.04)	2.33(0.82)	1.85(0.96)
MMM	12.46(7.59)	6.30(4.76)	3.94(1.16)	2.65(1.46)
MO	3.12(3.51)	3.97(4.52)	2.94(1.13)	1.69(0.71)
MRK	3.73(4.17)	1.31(5.11)	4.75(1.14)	4.03(1.49)
MSFT	0.09(2.49)	3.31(4.20)	1.88(0.58)	5.36(1.96)
PG	4.19(5.00)	12.45(10.56)	4.07(0.94)	3.68(1.93)
SBC	8.38(6.77)	3.58(5.38)	5.91(1.34)	4.42(1.29)
T	6.41(5.42)	1.85(3.08)	4.58(1.22)	3.46(0.88)
UTX	6.69(4.80)	4.85(7.47)	3.51(0.97)	2.59(1.92)
WMT	−1.48(3.17)	2.11(5.83)	2.80(0.82)	4.90(1.60)
XOM	19.50(6.93)	16.97(7.96)	9.95(1.81)	8.46(2.44)

Note: The table contains means and standard deviations of in-sample (1989–1998) and out-of-sample (1999–2003) portfolio weights in the MVP and minimum tracking error portfolio based on the SEMI model with x_t equal to the daily log-volatility for the market return.

to the in-sample period. This appears largely due to the portfolio's sensitivity to the HML factor, which is close to zero during the 1989–1998 period but substantially different from zero during 1999–2003.

In the context of factor models, Chan et al. (1999) attribute the similarity in performance of different covariance models in terms of MVP construction to the presence of a dominant factor, the market, which is more

Table 9. In-Sample Performance and Characteristics of Minimum Tracking Error Portfolios Based on Forecasting Models – 1989–1998.

	Mean	S. D.	Sharpe Ratio	Tracking Error	β_{M}	β_{SMB}	β_{HML}
SF	17.39	14.36	0.853	3.64	0.998(0.012)	−0.387(0.015)	−0.030(0.017)
TF	16.82	14.39	0.812	3.75	1.020(0.012)	−0.346(0.015)	−0.010(0.018)
RM	17.52	14.86	0.832	4.61	1.010(0.015)	−0.397(0.018)	−0.064(0.023)
CCC	17.57	14.40	0.864	3.37	0.997(0.011)	−0.403(0.014)	−0.045(0.016)
DCC	17.53	14.40	0.861	3.35	1.000(0.011)	−0.402(0.014)	−0.047(0.016)
ADCC	17.49	14.37	0.862	3.32	1.007(0.011)	−0.401(0.014)	−0.046(0.016)
STCC(h_{M})	17.55	14.39	0.863	3.34	1.000(0.011)	−0.398(0.014)	−0.042(0.016)
STCC(h_{SMB})	17.48	14.39	0.857	3.35	0.996(0.011)	−0.403(0.014)	−0.049(0.016)
STCC(h_{HML})	17.58	14.39	0.865	3.32	0.998(0.010)	−0.402(0.014)	−0.046(0.016)
SEMI(h_{M})	17.34	14.37	0.849	3.28	1.004(0.010)	−0.390(0.013)	−0.035(0.015)
SEMI(h_{SMB})	17.37	14.44	0.848	3.28	1.006(0.010)	−0.395(0.014)	−0.044(0.016)
SEMI(h_{HML})	17.65	14.36	0.871	3.26	0.998(0.010)	−0.398(0.014)	−0.044(0.015)
EWP	17.99	15.16	0.848	4.16	1.072(0.013)	−0.333(0.017)	−0.053(0.021)
VWP	21.52	17.95	0.913	6.83	1.114(0.017)	−0.410(0.021)	−0.458(0.027)

Note: The table reports in-sample (January 1, 1989–December 31, 1998) summary statistics for the minimum tracking error portfolios based on various forecasting models. h_{M}, h_{SMB} and h_{HML} indicate that the variable x_t in the STCC and semi-parametric models is taken to be the (contemporaneous) daily log-volatility for the market, SMB and HML factor returns, respectively. See Table 4 for definitions of the other models and description of the summary statistics shown.

Table 10. In-Sample Performance and Characteristics of Minimum Tracking Error Portfolios Based on Forecasting Models – 1999–2003.

	Mean	S. D.	Sharpe Ratio	Tracking Error	β_{M}	β_{SMB}	β_{HML}
SF	−3.16	21.96	−0.301	7.91	1.062(0.025)	−0.330(0.034)	0.406(0.040)
TF	−4.80	22.42	−0.368	7.70	1.070(0.030)	−0.293(0.036)	0.344(0.047)
RM	−5.25	23.25	−0.374	7.88	1.019(0.018)	−0.296(0.028)	0.098(0.035)
CCC	−3.14	21.61	−0.305	6.91	1.032(0.020)	−0.346(0.030)	0.332(0.034)
DCC	−3.69	21.70	−0.329	6.52	1.025(0.020)	−0.342(0.029)	0.284(0.033)
ADCC	−3.87	21.76	−0.339	6.78	1.029(0.020)	−0.344(0.030)	0.315(0.034)
STCC(h_{M})	−3.18	22.03	−0.301	6.60	1.036(0.020)	−0.343(0.029)	0.275(0.035)
STCC(h_{SMB})	−3.23	22.16	−0.316	6.56	1.041(0.020)	−0.314(0.026)	0.285(0.032)
STCC(h_{HML})	−3.31	21.90	−0.321	6.61	1.037(0.020)	−0.341(0.030)	0.223(0.031)
SEMI(h_{M})	−2.40	22.31	−0.262	6.65	1.045(0.020)	−0.343(0.030)	0.265(0.034)
SEMI(h_{SMB})	−3.30	22.06	−0.306	6.50	1.051(0.019)	−0.304(0.026)	0.288(0.032)
SEMI(h_{HML})	−3.47	22.30	−0.310	6.36	1.032(0.019)	−0.347(0.30)	0.219(0.033)
EWP	−1.98	21.82	−0.249	7.98	1.066(0.024)	−0.307(0.034)	0.425(0.039)
VWP	−4.42	19.67	−0.400	7.53	0.938(0.023)	−0.298(0.029)	0.313(0.037)

Note: The table reports out-of-sample (January 1, 1999–December 31, 2003) summary statistics for the minimum tracking error portfolios based on various forecasting models. See Table 9 for models definitions and description of the summary statistics shown.

important than the other influences on returns. Constructing the MVP then effectively boils down to minimizing the portfolio's sensitivity to the market. This is achieved by tilting the portfolio towards stocks with low market betas and explains why all covariance models yield such similar results. Chan et al. (1999) suggest that any differences between the different covariance models may be brought to light more clearly when the dominant factor is removed. This turns out to be equivalent to tracking a benchmark portfolio that resembles the dominant factor, and the relevant portfolio to consider becomes the portfolio that minimizes the tracking error.

4.5. Minimizing Tracking Error Using Correlation Forecasts

As the benchmark portfolio for the tracking error optimization problem, we choose the S&P 500 index. The minimum tracking error portfolio (MTEP) is equivalent to the MVP for the stock returns in excess of the return on the benchmark portfolio. Hence, we apply the various models to obtain forecasts of the conditional covariance matrix of the excess DJIA stock returns relative to the S&P 500 return, denoted $\widehat{H}_t^e$, and compute the MTEP weights as $w_t = \widehat{H}_t^{e-1}\iota/(\iota'\widehat{H}_t^{e-1}\iota)$. In this case, we also apply the STCC and semi-parametric models with conditioning variable x_t being the contemporaneous log-volatility of the size and book-to-market factor returns. This is motivated by the fact that these factors might be of considerable importance for the covariance structure of the excess returns as the dominant market factor is removed. Tables 9 and 10 display the characteristics of the resulting portfolios for the in-sample period 1989–1998 and out-of-sample period 1999–2003, respectively. We find that the semi-parametric model with $x_t = \log h_{\mathrm{HML},t}$ renders the smallest tracking error, both in-sample and out-of-sample, although the differences with other conditional correlation models are not large. Note that the SF and TF models perform quite disappointing for the out-of-sample period, with tracking errors that are comparable to the tracking errors of the EWP and VWP. Finally, we observe that the out-of-sample tracking error is twice as large as the in-sample tracking error. This again is due to the increased sensitivity to the book-to-market factor.

5. CONCLUDING REMARKS

We have developed a new semi-parametric model for conditional correlations, which combines parametric univariate GARCH-type specifications for the individual conditional volatilities with nonparametric kernel regression for the conditional correlations. This approach not only avoids the proliferation of parameters as the number of assets becomes large, which typically happens in conventional multivariate conditional volatility models, but also the rigid structure imposed by more parsimonious models, such as the DCC model. Hence, our model is applicable in empirically relevant settings, where portfolios might consist of 10s or even 100s of assets.

The application to the DJIA stocks illustrates the potential of the semiparametric model. The estimation results suggest that correlation functions are asymmetric when the conditioning variable is the lagged weekly market return. That is, correlations increase stronger in bear markets than they do in bull markets, in agreement with the "correlation breakdown" effect. We also find support for the idea that correlations increase with market volatility, as would be implied by a factor structure for the stock returns. The semi-parametric model performs quite well when compared with standard parametric DCC-type models in terms of constructing MVPs and minimum tracking error portfolios.

A point to note, finally, is that the time required to estimate the semiparametric model turned out to be far below that of even the simple DCC models with two parameters. The reason is that no ill-conditioned likelihood needs to be maximized, but rather a simple data smoother is used. Future research should develop more guidance with respect to bandwidth selection, and find ways to avoid the curse of dimensionality when more than one conditioning variable is considered.

NOTES

1. The number of parameters in an unrestricted VEC-model for N assets equals $(N(N+1)/2) \times (1+2N(N+1)/2$. The BEKK-model contains $2N^2+N(N+1)/2$ parameters.
2. As discussed in Section 2 below, the number of parameters in these models grows linearly with the number of assets.
3. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data–library.html.

ACKNOWLEDGMENTS

We thank participants of the 3rd Annual Advances in Econometrics Conference, Baton Rouge (November 2004), and the 55th Session of the International Statistical Institute (Sydney, April 2005), as well as an anonymous referee for useful comments and suggestions. Any remaining errors are ours.

REFERENCES

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*, 43–76.

Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, *63*, 443–494.

Bauwens, L., Laurent, S., & Rombouts, J.V.K. (2005). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, forthcoming.

Berben, R.-P., & Jansen, W. J. (2005). Comovement in international equity markets: A sectoral view. *Journal of International Money and Finance*, *24*, 832–857.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, *72*, 498–505.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time varying covariances. *Journal of Political Economy*, *96*, 116–131.

Butler, K. C., & Joaquin, D. C. (2002). Are the gains from international portfolio diversification exaggerated? The influence of downside risk in bear markets. *Journal of International Money and Finance*, *21*, 981–1011.

Campbell, R., Koedijk, K., & Kofman, P. (2002). Increased correlation in bear markets. *Financial Analyst Journal*, *58*, 87–94.

Cappiello, L., Engle, R.F., & Sheppard, K. (2003). *Asymmetric dynamics in the correlations of global equity and bond returns*. European Central Bank Working paper No. 204.

Chan, L. K. C., Karceski, J., & Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, *12*, 674–937.

Davydov, Y. A. (1973). Mixing conditions for Markov chains. *Theory of Probability and its Applications*, *18*, 312–328.

Doukhan, P. (1994). *Mixing: Properties and examples, lecture notes in statistics 85*. Berlin: Springer.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Engle, R. F. (2002). Dynamic conditional correlation – A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics*, *20*, 339–350.

Engle, R. F., & Kroner, K. F. (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory*, *11*, 122–150.

Engle, R. F., & Manganelli, S. (2004). CaViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*, 367–381.

Engle, R.F., & Sheppard, K. (2001). *Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH*. UCSD Discussion paper No. 2001–15.

Engle, R. F., & Mezrich, J. (1996). GARCH for groups. *Risk*, *9*, 36–40.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3–56.

Fama, E. F., & French, K. R. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, *51*, 55–84.
Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. New York: Chapmann and Hall.
Fan, J., & Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, *85*, 645–660.
Forbes, K. J., & Rigobon, R. (2002). No contagion, only interdependence: Measuring stock market comovements. *Journal of Finance*, *57*, 2223–2261.
Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*, 1779–1801.
Hafner, C.M., & Franses, P.H. (2003). *A generalized dynamic conditional correlation model for many assets*. Econometric Institute Report No. 2003–18, Erasmus University, Rotterdam.
Härdle, W., Lütkepohl, H., & Chen, R. (1997). A review of nonparametric time series analysis. *International Statistical Review*, *65*, 49–72.
Härdle, W., Tsybakov, A., & Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, *68*, 221–245.
Jennen-Steinmetz, C., & Gasser, T. (1988). A unifying approach to nonparametric regression estimation. *Journal of the American Statistical Association*, *83*, 1084–1089.
Liptser, R. S., & Shirjaev, A. N. (1980). A functional central limit theorem for martingales. *Theory of Probability and its Applications*, *25*, 667–688.
Longin, F., & Solnik, B. (1995). Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance*, *14*, 3–26.
Longin, F., & Solnik, B. (2001). Extreme correlation of international equity markets. *Journal of Finance*, *56*, 651–676.
Loretan, M., & English, W.B. (2000). Evaluating changes in correlations during periods of high market volatility. *Bank of International Settlements Quarterly Review*, (June), 29–36.
Pagan, A. R., & Ullah, A. (1999). *Nonparametric econometrics*. Cambridge: Cambridge University Press.
Pelletier, D. (2005). Regime switching for dynamic correlations. *Journal of Econometrics*, forthcoming.
Pötscher, B. M., & Prucha, I. R. (1997). *Dynamic nonlinear econometric modelling: Asymptotic theory*. Berlin: Springer.
Ramchand, L., & Susmel, R. (1998). Volatility and cross-correlation across major stock markets. *Journal of Empirical Finance*, *5*, 397–416.
Robinson, P. M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, *4*, 185–207.
Rodriguez-Poo, J. M., & Linton, O. (2001). Nonparametric factor analysis of residual time series. *TEST, 10, 161–182*.
Silvennoinen, A., & Teräsvirta, T. (2005). *Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations*. SSE/EFI Working Paper Series in Economics and Finance No. 577, Stockholm School of Economics.
Tse, Y. K., & Tsui, A. K. C. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics*, *20*, 351–362.
Ziegelmann, F. A. (2002). Nonparametric estimation of volatility functions: The local exponential estimator. *Econometric Theory*, *18*, 985–991.

APPENDIX: ASSUMPTIONS AND PROOF OF THEOREM 1

We use the following notation. First and second derivatives of a function $f:\mathbb{R}\to\mathbb{R}$ are denoted by $f'(\cdot)$ and $f''(\cdot)$, respectively. Let x be an N-vector and X an $N\times N$ symmetric matrix. The Euclidean norm of x is denoted as $|x|$. The operator diag builds a diagonal matrix from a vector argument, i.e.,

$$\mathrm{diag}(x)=\begin{pmatrix} x_1 & 0 & \cdots & 0 \\ 0 & x_2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & x_N \end{pmatrix}$$

The operator dg stacks the diagonal of a matrix into a vector, i.e., $\mathrm{dg}(X)=(X_{11},X_{22},\ldots,X_{NN})'$. The operator vec stacks the columns of a matrix into a column vector, and vech stacks only the lower triangular part including the diagonal into a vector. Denote by D_N the $N^2\times(N(N+1)/2$ duplication matrix and by D_N^+ its generalized inverse, of size $(N(N+1)/2\times N^2$, with the properties $\mathrm{vech}(X)=D_N^+\mathrm{vec}(X)$ and $\mathrm{vec}(X)=D_N\mathrm{vech}(X)$. Further, denote by L_N the $N\times N^2$ matrix with the property $L_N\mathrm{vec}(X)=\mathrm{dg}(X)$.

We work under the following assumptions:

Assumption 1.

(A1) The true parameter vector θ_0 governing the conditional variance matrix D_t exists uniquely and lies in the interior of a compact set.

(A2) The first-stage estimator of θ is $\sqrt{T}$-consistent and asymptotically normal.

(A3) The process $\{r_t\}$ is strictly stationary.

(A4) The second-stage kernel function satisfies $\int uK(u)\,\mathrm{d}u=0$, $\int K^2(u)\,\mathrm{d}u=||K||_2^2<\infty$, $\mu_2(K)=\int u^2K(u)\,\mathrm{d}u<\infty$, $\int K(u)\,\mathrm{d}u=1$, $\sup|K(u)|=B<\infty$, and $|u||K(u)|\to 0$ as $|u|\to\infty$.

(A5) The off-diagonal elements of $R(x)$ are strictly between -1 and 1, uniformly in x, and are at least twice continuously differentiable on the support of x_t.

(A6) The process $\xi_t=R^{-1/2}(x_t)\varepsilon_t$ is i.i.d. The density $p()$ of ξ_1 exists and is positive on any compact subset of $\mathbb{R}^N$. Furthermore, $\mathrm{E}[\xi_1]=0$, $\mathrm{E}[\xi_1\xi_1']=I_N$, and there exist constants $\delta,B>0$ such that $\mathrm{E}[\xi_{it}^{4+\delta}]\le B<\infty$, $i=1\ \ldots,N$.

Note that (A6) implies that fourth moments of ε_t are finite, because $R()$ is bounded. For notational convenience, define $\tilde{\xi}_t = \mathrm{vec}(\xi_t \xi_t' - I_N)$, which by (A6) is i.i.d. with mean zero and $E[\tilde{\xi}_t \tilde{\xi}'_t] = M_4$, say.

In the proof of Theorem 1, we make use of the following lemmata:

Lemma 1. Let

$$\widetilde{R}(x) = \widetilde{Q}^*(x)^{-1} \widetilde{Q}(x) \widetilde{Q}^*(x)^{-1}$$

where

$$\widetilde{Q}(x) = \frac{\sum_{t=1}^{T} \tilde{\varepsilon}_t \tilde{\varepsilon}_t' K_h(x_t - x)}{\sum_{t=1}^{T} K_h(x_t - x)}$$

$\widetilde{Q}^*(x)$ is a diagonal matrix with the square roots of the diagonal elements of $\widetilde{Q}(x)$ on its diagonal, and $\tilde{\varepsilon}_t = D_t(\theta_0)^{-1} r_t$. Then, under assumptions (A1) to (A6)

$$\tilde{r}(x) - \hat{r}(x) = O_p\left(\frac{1}{Tb}\right)$$

uniformly in x, where $\tilde{r}(x) = \mathrm{vech}(\widetilde{R}(x))$.

Proof. Very similar to the proof of Lemma A.2 of Rodriguez-Poo and Linton (2001). □

Lemma 2. Assume that the diagonal elements of D_t^2 are specified as

$$h_{it} = \omega_i + \alpha_i r_{i,t-1}^2 + \beta_i h_{i,t-1}$$

with $\alpha_i + \beta_i < 1, \quad i = 1, \ldots, N$. Under Assumptions (A1) to (A6) and $R_t = R(r_{t-1})$, the process $(r_t)_{t=0}^{\infty}$ is geometrically ergodic. Furthermore, if r_0 is drawn from the stationary distribution, then $(r_t)_{t=0}^{\infty}$ is geometrically α-mixing.

Proof. Let $\omega = (\omega_1, \ldots, \omega_N)'$, $\alpha = (\alpha_1, \ldots, \alpha_N)'$, and $\beta = (\beta_1, \ldots, \beta_N)'$. Define the processes $\mathrm{V}_t = (V_{1t}, \ldots, V_{Nt})'$ with $V_{it} = \sum_{j=0}^{\infty} \beta_i^j r_{i,t-j}^2$, and the $2N$-vector process $Z_t = \left(r_t', V_t'\right)'$. By construction, Z_t is a Markov process which can be represented as $Z_t = m(Z_{t-1}) + g(Z_{t-1})\zeta_t$ with $\zeta_t = (\xi_t', \tilde{\xi}'_t L_N')'$ and where $m : \mathbb{R}^{2N} \to \mathbb{R}^{2N}$ and $g : \mathbb{R}^{2N} \to \mathbb{R}^{2N} \times \mathbb{R}^{2N}$, given by

$$m(Z_{t-1}) = \begin{pmatrix} 0_{N \times 1} \\ \mathrm{diag}(\beta) V_{t-1} + \mathrm{diag}\left(\tilde{V}_{t-1}\right) R_t \end{pmatrix}$$

and with $\tilde{V}_t = (I_N - \mathrm{diag}(\beta))^{-1}\omega + \mathrm{diag}(\alpha)V_t$,

$$g(Z_{t-1}) = \begin{pmatrix} \mathrm{diag}\left(\tilde{V}_{t-1}\right)^{1/2} R_t^{1/2} & 0_{N\times N} \\ 0_{N\times N} & \mathrm{diag}\left(\tilde{V}_{t-1}\right) L_N\left(R_t^{1/2}\otimes R_t^{1/2}\right) \end{pmatrix}$$

Using assumption (A6), $\mathrm{E}[\zeta_t\zeta_t']$ is finite and given by W, say. The condition of Doukhan (1994),

$$\lim_{|z|\to\infty} \sup \frac{|m(z)|^2 + \mathrm{Tr}(Wg'(z)g(z))}{|z|^2} < 1$$

is easily checked, as $R(\cdot)$ is bounded. By proposition 6 of Doukhan (1994, p. 107), geometric ergodicity of (Z_t) follows. Finally, by Davydov (1973), a geometrically ergodic Markov chain with initial value drawn from the stationary distribution is geometrically α-mixing. This proves the lemma. $\square$

Proof of Theorem 1. Let us only consider the case of constant bandwidths, $a = 0$, so that $h = b$. The proof for the general case $a > 0$ is analogous to Jennen-Steinmetz and Gasser (1988). Also, we only sketch the proof for the factor $x_t = w'r_{t-1}$, where w is a fixed N-vector. First, by definition $\rho_{ii}(x) = \mathrm{E}[\varepsilon_{it}^2|x_{t-1} = x] = 1$, $i = 1, \ldots, N$. The geometric mixing property of $(x_t)_{t=0}^{\infty}$ as implied by Lemma 2 ensures consistency and asymptotic normality of the standard kernel smoother as shown by Robinson (1983) under weaker conditions. Thus, $\widehat{Q}(x)$ is a consistent estimator for $R(x)$. Because $\widehat{Q}^*(x) \xrightarrow{p} I_N$ and using Slutsky's Theorem, $\widehat{R}(x)$ is also a consistent estimator for $R(x)$ (it has the additional advantage of being a correlation matrix for finite samples).
Turning to the asymptotic distribution, note that

$$\sqrt{Th}(\hat{r}(x) - r(x)) = \sqrt{Th}(\hat{r}(x) - \tilde{r}(x)) + \sqrt{Th}(\tilde{r}(x) - r(x)) \qquad \text{(A.1)}$$

where the first term on the right-hand side of (A.1) is $O_p((Th)^{-1/2})$ using Lemma 1, and therefore converges to zero in probability under the assumption, $Th \to \infty$. Thus, the asymptotic distribution of the left-hand side of (A.1) is the same as that of the second term on the right-hand side of (A.1). This term however has the same asymptotic distribution as

$\sqrt{Th}(\tilde{q}(x) - r(x))$, where $\tilde{q}(x) = \text{vech}(\tilde{Q}(x))$. We can write

$$\sqrt{Th}(\tilde{q}(x) - r(x)) = \frac{1}{\hat{f}(x)}(Th)^{-1/2}\sum_{t=1}^{T} K(\frac{x_t - x}{h})\{\eta_t - r(x)\} \qquad \text{(A.2)}$$

where $\hat{f}(x) = (Th)^{-1}\Sigma_{t=1}^{T} K(\frac{x_t-x}{h})$ is the kernel estimator of $f(x)$, the density of x_t, and $\eta_t = \text{vech}(\varepsilon_t\varepsilon_t')$. As $\hat{f}$ converges to f in probability, the asymptotic distribution is determined by that of $(Th)^{-1/2}\Sigma_{t=1}^{T} K(\frac{x_t-x}{h})\{\eta_t - r(x)\}$, which we can decompose as

$$(Th)^{-1/2}\sum_{t=1}^{T} K\Big(\frac{x_t - x}{h}\Big)\big\{\eta_t - r(x)\big\} = I_1 + I_2$$

where, denoting $u_t = \eta_t - r(x_t)$,

$$I_1 = (Th)^{-1/2}\sum_{t=1}^{T} K\Big(\frac{x_t - x}{h}\Big)u_t$$

and

$$I_2 = (Th)^{-1/2}\sum_{t=1}^{T} K\Big(\frac{x_t - x}{h}\Big)\{r(x_t) - r(x)\}$$

Since $r(x_t) = \text{E}[\eta_t|x_t]$, u_t is a martingale difference sequence. By Lemma 2 it is a function of geometrically α-mixing processes and thus itself geometrically α-mixing. Note that it can be written as $u_t = D_N^+(R_t^{1/2} \otimes R_t^{1/2})\tilde{\xi}_t$ with $R_t = R(x_t)$ and

$$\text{V}(u_t|x_t) = D_N^+(R_t^{1/2} \otimes R_t^{1/2})M_4(R_t^{1/2} \otimes R_t^{1/2})D_N < \infty$$

by (A6). Using the Cramér-Wold device, $I_1 \xrightarrow{\mathscr{L}} N(0, f(x)^2\ \Sigma(x)) \Leftrightarrow c'I_1 \xrightarrow{\mathscr{L}} N(0, f(x)^2 c'\Sigma(x)c)$ for any fixed vector c of length $N(N+1)/2$. Thus, it suffices to show that a central limit theorem holds for $c'I_1$ with the required asymptotic variance. Using Lemma 5.2 of Härdle, Tsybakov, and Yang (1998), one obtains $\text{V}(c'I_1) = f(x)^2 c'\Sigma(x)c + o(1)$ with $\Sigma(x) = ||K||_2^2\text{V}(u_t|x_t = x)/f(x)$. This shows that the asymptotic variance of $\hat{f}(x)^{-1}I_1$ is given by $\Sigma(x)$. To prove that the asymptotic distribution is normal, the conditions of the central limit theorem for square integrable martingale differences of Liptser and Shirjaev (1980, Corollary 6) can be shown to hold along the lines of Härdle et al. (1998).

For the second term, note first that

$$\begin{aligned}
\mathrm{E}[I_2] &= (Th)^{-1/2}\sum_{t=1}^{T} \mathrm{E}\Big[K(\frac{x_t - x}{h})\{r(x_t) - r(x)\}\Big] \\
&= (Th)^{1/2}\int K(\psi)\{r(x + h\psi) - r(x)\}f(x + h\psi)\,\mathrm{d}\psi \\
&= (Th)^{1/2}\Big\{h^2\mu_2(K)\Big(\frac{f(x)r''(x)}{2} + f'(x)r'(x)\Big) + o(h^2)\Big\} \\
&= (Th)^{1/2}\big\{O(h^2) + o(h^2)\big\} = O\Big\{(Th)^{1/2}h^2\Big\}
\end{aligned}$$

which converges to zero under the assumption $h = o(T^{-1/5})$. Now $Y_t = K(\psi)\{r(x_t) - r(x)\}$, $\psi = (x_t - x)/h$, is a mixing process as it is a measurable function of a mixing process. Hence, by Theorem 6.3 of Pötscher and Prucha (1997), a weak law of large numbers holds for Y_t if $\sup_T T^{-1}\Sigma_t^T\ \mathrm{E}\big[|Y_t|^{1+\varepsilon}\big] < \infty$ for some $\varepsilon > 0$. Using (A5), $|r(x_t) - r(x)| \leq 2|r(x_t)| \leq 2$ a.s. Furthermore, $K(\cdot)$ is bounded by (A4). Thus, $\mathrm{E}[|Y_t|^{1+\varepsilon}] \leq \mathrm{E}[|K(\psi)|^{2(1+\varepsilon)}]^{1/2}\mathrm{E}[|r(x_t) - r(x)|^{2(1+\varepsilon)}]^{1/2} \leq |2B|^{1+\varepsilon} < \infty$. Thus, $I_2 = o_p(1)$, which completes the proof. □

This page intentionally left blank

A MULTIVARIATE HEAVY-TAILED DISTRIBUTION FOR ARCH/GARCH RESIDUALS☆

Dimitris N. Politis

ABSTRACT

A new multivariate heavy-tailed distribution is proposed as an extension of the univariate distribution of Politis (2004). The properties of the new distribution are discussed, as well as its effectiveness in modeling ARCH/GARCH residuals. A practical procedure for multi-parameter numerical maximum likelihood is also given, and a real data example is worked out.

1. INTRODUCTION

Consider univariate data $X_1, \ldots, X_n$ arising as an observed stretch from a financial returns time series $\{X_t, \ t = 0, \pm 1, \pm 2, \ldots\}$ such as the percentage returns of a stock price, stock index or foreign exchange rate. The returns series $\{X_t\}$ are assumed strictly stationary with mean zero which – from a practical point of view – implies that trends and other nonstationarities have

☆Research partially supported by NSF Grant SES-04-18136 funded jointly by the Economics and Statistics Sections of the National Science Foundation.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 105–124

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20004-X

been successfully removed. Furthermore, for simplicity the returns series $\{X_t\}$ will be assumed to have a symmetric[1] distribution around zero.

The celebrated ARCH models of Engle (1982) were designed to capture the phenomenon of volatility clustering in the returns series. An ARCH(p) model can be described by the following equation:

$$X_t = Z_t \sqrt{a + \sum_{i=1}^{p} a_i X_{t-i}^2} \tag{1}$$

where a, a_1, …, a_p are nonnegative parameters. Since the residuals from a fitted ARCH(p) model do not appear to be normally distributed, practitioners have typically been employing ARCH models with heavy-tailed errors. A popular assumption for the distribution of the numbers of Z_t is the t-distribution with degrees of freedom empirically chosen to match the apparent degree of heavy tails as measured by higher-order moments such as the kurtosis; see e.g., Bollerslev, Chou, and Kroner (1992) or Shephard (1996) and the references therein.

Nevertheless, the choice of a t-distribution is quite arbitrary, and the same is true for other popular heavy-tailed distributions, e.g., the double exponential. For this reason, Politis (2004) developed an implicit ARCH model that helps suggest a more natural heavy-tailed distribution for ARCH and/or GARCH residuals; see Bollerslev (1986) for the definition of GARCH models. The motivation for this new distribution is given in Section 2, together with a review of some of its properties. In Section 3, the issue of multivariate data is brought up, and a multivariate version of the new heavy-tailed distribution is proposed. A practical algorithm for multiparameter numerical maximum likelihood is given in Section 4 based on an approximate steepest-ascent idea. Section 5 deals in detail with the bivariate paradigm. Finally, a real data example is worked out in Section 6.

2. THE IMPLICIT ARCH MODEL IN THE UNIVARIATE CASE: A BRIEF REVIEW

Engle's (1982) original assumption was that the errors Z_t in model (1) are i.i.d. $N(0,1)$. Under that assumption, the residuals

$$\hat{Z}_t = \frac{X_t}{\sqrt{\hat{a} + \sum_{i=1}^{p} \hat{a}_i X_{t-i}^2}} \tag{2}$$

ought to appear normally distributed; here, $\hat{a}, \hat{a}_1, \hat{a}_2, \ldots$ are estimates of the nonnegative parameters a, a_1, a_2, … . Nevertheless, as previously mentioned, the residuals typically exhibit a degree of heavy tails. In other words, typically there is not a specification for $\hat{a}, \hat{a}_1, \hat{a}_2, \ldots$ that will render the residuals $\hat{Z}_t$ normal-looking.

A way to achieve ARCH residuals that appear normal was given in the implicit ARCH model of Politis (2004) that is defined as

$$X_t = W_t \sqrt{a + a_0 X_t^2 + \sum_{i=1}^{p} a_i X_{t-i}^2} \tag{3}$$

When the implicit ARCH model (3) is fitted to the data, the following residuals ensue:

$$\hat{W}_t = \frac{X_t}{\sqrt{\hat{a} + \hat{a}_0 X_t^2 + \sum_{i=1}^{p} \hat{a}_i X_{t-i}^2}} \tag{4}$$

As argued in Politis (2004), by contrast to the standard ARCH model (1), there typically always exists here a specification for $\hat{a}$, $\hat{a}_0$, $\hat{a}_1, \ldots$ that will render the residuals $\hat{W}_t$ normal-looking.[2] Hence, it is not unreasonable to assume that the W_t errors in model (3) are i.i.d. $N(0, 1)$.

Technically speaking, however, this is not entirely correct since it is easy to see that the W_ts are actually bounded (in absolute value) by $1/\sqrt{a_0}$. A natural way to model the situation where the W_ts are thought to be close to $N(0,1)$ but happen to be bounded is to use a truncated standard normal distribution, i.e., to assume that the W_ts are i.i.d. with probability density given by

$$\frac{\phi(x)1\{|x| \leq C_0\}}{\int_{-C_0}^{C_0} \phi(y)\,\mathrm{d}y} \quad \text{for all} \quad x \in \mathbf{R} \tag{5}$$

where ϕ denotes the standard normal density, $1\{S\}$ the indicator of set S, and $C_0 = 1/\sqrt{a_0}$. If/when a_0 is small enough, the boundedness of W_t is effectively not noticeable.

Note that the only difference between the implicit ARCH model (3) and the standard ARCH model (1) is the introduction of the term X_t paired with a nonzero coefficient a_0 in the right-hand side of Eq. (3). It is precisely this X_t term appearing in both sides of (3) that gives the characterization 'implicit' to model (3).

Nevertheless, we can solve Eq. (3) for X_t, to obtain

$$X_t = U_t \sqrt{a + \sum_{i=1}^{p} a_i X_{t-i}^2} \tag{6}$$

where

$$U_t = \frac{W_t}{\sqrt{1 - a_0 W_t^2}} \tag{7}$$

Thus, the implicit ARCH model (3) is seen to be tantamount to the regular ARCH(p) model (6) associated with the new innovation term U_t.

However, if W_t is assumed to follow the truncated standard normal distribution (5), then the change of variable (7) implies that the innovation term U_t appearing in the ARCH model (6) has the density $f(u; a_0, 1)$ defined as

$$f(u;\ a_0,\ 1) = \frac{(1 + a_0 u^2)^{-3/2} \exp\left(-\frac{u^2}{2(1+a_0 u^2)}\right)}{\sqrt{2\pi}\left(\Phi(1/\sqrt{a_0}) - \Phi(-1/\sqrt{a_0})\right)} \quad \text{for all} \quad u \in \mathbf{R} \tag{8}$$

where Φ denotes the standard normal distribution function. Eq. (8) gives the univariate heavy-tailed density for ARCH residuals proposed in Politis (2004).

Note that the density $f(u;\ a_0,\ 1)$ can be scaled to create a two-parameter family of densities with a typical member given by

$$f(x; a_0, c) = \left(\frac{1}{c}\right) f\left(\left(\frac{x}{c}\right); a_0, 1\right) \text{ for } \quad \text{all } x \in \mathbf{R} \tag{9}$$

The nonnegative parameter a_0 is a shape parameter having to do with the degree of heavy tails, while the positive parameter c represents scale; note that $f(u;\ a_0,\ 1) \to \phi(u)$ as $a_0 \to 0$.

It is apparent that $f(u;\ a_0,\ c)$ has heavy tails. Except for the extreme case when $a_0 = 0$ and all moments are finite, in general moments are finite only up to (almost) order two. In other words, if a random variable U follows the density $f(u; a_0, c)$ with $a_0, c > 0$, then it is easy to see that

$$E|U|^d < \infty \text{ for all } \quad d \in [0, 2) \text{ but } \quad E|U|^d = \infty \text{ for all } \quad d \in [2, \infty)$$

Thus, working with the heavy-tailed density $f(u; a_0, c)$ is associated with the conviction that financial returns have higher-order moments that are infinite; Politis (2003, 2004) gives some empirical evidence pointing to that fact.[3]

3. A MULTIVARIATE HEAVY-TAILED DISTRIBUTION FOR ARCH/GARCH

Here, and throughout the remainder of the paper, we will consider the case where the data $X_1, \ldots, X_n$ are an observed stretch from a *multivariate*, strictly stationary and zero mean, series of returns $\{X_t,\ t = 0, \pm 1, \pm 2, \ldots\}$, where $\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)}, \ldots, X_t^{(m)})'$.

By the intuitive arguments of Section 2, each of the coordinates of $\mathbf{X}_t$ may be thought to satisfy an implicit ARCH model such as (3) with residuals following the truncated standard normal distribution (5). Nevertheless, it is possible in general that the coordinates of the vector $\mathbf{X}_t$ are dependent, and this can be due to their respective residuals being correlated.

Focusing on the jth coordinate, we can write the implicit ARCH model

$$X_t^{(j)} = W_t^{(j)} \sqrt{a^{(j)} + \sum_{i=0}^{p} a_i^{(j)} (X_{t-i}^{(j)})^2} \quad \text{for} \quad j = 1, \ldots, m \tag{10}$$

and the corresponding regular ARCH model

$$X_t^{(j)} = U_t^{(j)} \sqrt{a^{(j)} + \sum_{i=1}^{p} a_i^{(j)} (X_{t-i}^{(j)})^2} \quad \text{for} \quad j = 1, \ldots, m \tag{11}$$

with residuals given by

$$U_t^{(j)} = \frac{W_t^{(j)}}{\sqrt{1 - a_0^{(j)} (W_t^{(j)})^2}} \quad \text{for} \quad j = 1, \ldots, m \tag{12}$$

As usual, we may define the jth volatility by

$$s_t^{(j)} = \sqrt{a^{(j)} + \sum_{i=1}^{p} a_i^{(j)} (X_{t-i}^{(j)})^2}$$

With sufficiently large p, an ARCH model such as (11) can approximate an arbitrary stationary univariate GARCH model; see Bollerslev (1986) or Gouriéroux (1997). For example, consider the multivariate GARCH (1,1) model given by the equation

$$X_t^{(j)} = s_t^{(j)} U_t^{(j)} \tag{13}$$

with $s_t^{(j)} = \sqrt{C^{(j)} + A^{(j)} (X_{t-1}^{(j)})^2 + B^{(j)} (s_{t-1}^{(j)})^2}$ for $j = 1, \ldots, m$.

It is easy to see that the multivariate GARCH model (13) is tantamount to the multivariate ARCH model (11) with $p = \infty$ and the following identifications:

$$a^{(j)} = \frac{C^{(j)}}{1 - B^{(j)}} \quad \text{and} \quad a_k^{(j)} = A^{(j)}(B^{(j)})^{k-1} \quad \text{for} \quad k = 1, 2, \ldots \tag{14}$$

The above multivariate ARCH/GARCH models (11) and (13) are in the spirit of the constant (with respect to time) conditional correlation multivariate models of Bollerslev (1990). For instance, note that the volatility equation for coordinate j involves past values and volatilities from the same jth coordinate only. More general ARCH/GARCH models could also be entertained where $s_t^{(j)}$ is allowed to depend on past values and volatilities of other coordinates as well; see e.g., Vrontos, Dellaportas, and Politis (2003) and the references therein. For reasons of concreteness and parsimony, however, we focus here on the simple models (11) and (13).

To complete the specification of either the multivariate ARCH model (11) or the multivariate GARCH model (13), we need to specify a distribution for the vector residuals $(U_t^{(1)}, U_t^{(2)}, \ldots, U_t^{(m)})'$. For fixed t, those residuals are expected to be heavy-tailed and generally dependent; this assumption will lead to a constant (with respect to time) conditional *dependence* model since correlation alone cannot capture the dependence in nonnormal situations.

As before, each $W_t^{(j)}$ can be assumed (quasi)normal but bounded in absolute value by

$$C_0^{(j)} = 1/\sqrt{a_0^{(j)}}$$

So, the vector $(W_t^{(1)}, W_t^{(2)}, \ldots, W_t^{(m)})'$ may be thought to follow a truncated version of the m-variate normal distribution $N_m(0, \mathrm{P})$, where P is a positive definite *correlation* matrix, i.e., its diagonal elements are unities. As well known, the distribution $N_m(0, \mathrm{P})$ has density

$$\phi_{\mathbf{P}}(\mathbf{y}) = (2\pi)^{-m/2}|\mathbf{P}|^{-1/2}\exp(-\mathbf{y}'\mathbf{P}^{-1}\mathbf{y}/2) \quad \text{for all} \quad \mathbf{y} \in \mathbf{R}^m \tag{15}$$

Thus, the vector $(W_t^{(1)}, W_t^{(2)}, \ldots, W_t^{(m)})'$ would follow the truncated density

$$\frac{\phi_{\mathbf{P}}(\mathbf{w})1\{\mathbf{w} \in E_0\}}{\int_{E_0} \phi_{\mathbf{P}}(\mathbf{y})\, d\mathbf{y}} \quad \text{for all} \quad \mathbf{w} \in \mathbf{R}^m \tag{16}$$

where E_0 is the rectangle $[-C_0^{(1)}, C_0^{(1)}] \times [-C_0^{(2)}, C_0^{(2)}] \times \ldots \times [-C_0^{(m)}, C_0^{(m)}]$.

Consequently, by the change of variables (12) it is implied that the vector innovation term $(U_t^{(1)}, U_t^{(2)}, \ldots, U_t^{(m)})'$ appearing in the multivariate ARCH model (11) or the GARCH model (13) has the multivariate density $f_{\mathbf{p}}(\mathbf{u}; \mathbf{a}_0, 1)$

defined as

$$f_{\mathbf{P}}(\mathbf{u};\mathbf{a}_0,1)=\frac{\phi_{\mathbf{P}}(\mathbf{g}(\mathbf{u}))\prod_{j=1}^{m}\,(1+a_0^{(j)}u_j^2)^{-3/2}}{\int_{E_0}\,\phi_{\mathbf{P}}(\mathbf{y})\,\mathrm{d}\mathbf{y}}\quad\text{for all}\quad\mathbf{u}\in\mathbf{R}^m \tag{17}$$

where $\mathbf{u}=(u_1,\ldots,u_m)'$, $\mathbf{a}_0=(a_0^{(1)},\ldots,a_0^{(m)})'$, and $\mathbf{g}(\mathbf{u})=(g_1(\mathbf{u}),\ldots,g_m(\mathbf{u}))'$ with

$$g_j(\mathbf{u})=u_j/\sqrt{1+a_0^{(j)}u_j^2}\ \text{ for } j=1,\ldots,m$$

note that, by construction, $\mathbf{g}(\mathbf{u})$ is in E_0 for all $\mathbf{u}\in\mathbf{R}^m$.

Therefore, our constant (with respect to time) conditional dependence ARCH/GARCH model is given by (11) or (13) respectively, together with the assumption that the vector residuals $(U_t^{(1)},U_t^{(2)},\ldots,U_t^{(m)})'$ follow the heavy-tailed multivariate density $f_{\mathbf{p}}(\mathbf{u};\,\mathbf{a}_0,\,1)$ defined in (17). Because the dependence in the new density $f_{\mathbf{p}}(\mathbf{u};\,\mathbf{a}_0,\,1)$ is driven by the underlying (truncated) normal distribution (16), the number of parameters involved in $f_{\mathbf{p}}(\mathbf{u};\,\mathbf{a}_0,\,1)$ is $m+m(m-1)/2$, i.e., the length of the $\mathbf{a}_0$ vector plus the size of the lower triangular part of $\mathbf{P}$; this number is comparable to the number of parameters found in Bollerslev's (1990) constant conditional correlation models.

The vector $\mathbf{a}_0$ is a shape parameter capturing the degree of heavy tails in each direction, while the entries of correlation matrix $\mathbf{P}$ capture the multivariate dependence. Again note that the density $f_{\mathbf{p}}(\mathbf{u};\,\mathbf{a}_0,\,1)$ can be scaled to create a more general family of densities with typical a member given by

$$f_{\mathbf{P}}(\mathbf{u};\mathbf{a}_0,\mathbf{c})=\frac{1}{\prod_{j=1}^{m}c_j}f_{\mathbf{P}}\left(\left(\frac{u_1}{c_1},\ldots,\frac{u_m}{c_m}\right)';\mathbf{a}_0,1\right)\quad\text{for all}\quad\mathbf{u}\in\mathbf{R}^m \tag{18}$$

where $\mathbf{c}=(c_1,\ldots,c_m)'$ is the scaling vector.

4. NUMERICAL MAXIMUM LIKELIHOOD

The general density $f_{\mathbf{p}}(\mathbf{u};\,\mathbf{a}_0,\,\mathbf{c})$ will be useful for the purpose of Maximum Likelihood Estimation (MLE) of the unknown parameters. Recall that the (pseudo)likelihood, i.e., conditional likelihood, in the case of the multivariate ARCH model (11) with residuals following the density $f_{\mathbf{p}}(\mathbf{u};\,\mathbf{a}_0,\,1)$ is given by the expression

$$\mathbf{L}=\prod_{t=p+1}^{n}f_{\mathbf{P}}(\mathbf{X}_t;\mathbf{a}_0,\mathbf{s}_t) \tag{19}$$

where $S_t=(s_t{}^{(1)},\ldots,s_t{}^{(\mathrm{m})})'$ is the vector of volatilities. The (pseudo)likelihood in the GARCH model (13) can be approximated by that of an ARCH model with order p that is sufficiently large; see also Politis (2004) for more details. As Hall and Yao (2003) recently showed, maximum (pseudo)likelihood estimation will be consistent even in the presence of heavy-tailed errors albeit possibly at a slower rate.

As usual, to find the MLEs one has to resort to numerical optimization.[4] Note that in the case of fitting the GARCH(1,1) model (13), one has to optimize over $4m+m(m-1)/2$ free parameters; the $4m$ factor comes from the parameters $A^{(j)}, B^{(j)}, C^{(j)}$, and $a_0^{(j)}$ for $j=1,\ldots,m$, while the $m+m(m-1)/2$ factor is the number of parameters in $\mathbf{P}$. As in all numerical search procedures, it is crucial to get good starting values.[5] In this case, it is recommended to get starting values for $A^{(j)}, B^{(j)}, C^{(j)}$, and $a_0^{(j)}$ by fitting a univariate GARCH(1,1) model to the jth coordinate (see Politis (2004) for details). A reasonable starting value for the correlation matrix $\mathbf{P}$ is the identity matrix. An alternative – and more informative – starting value for the j, k element of matrix $\mathbf{P}$ is the value of the cross-correlation at lag 0 of series $\{\mathrm{X}_t^{(j)}\}$ to series $\{\mathrm{X}_t^{(k)}\}$.

Nevertheless, even for m as low as 2, the number of parameters is high enough to make a straightforward numerical search slow and impractical. Furthermore, the gradient (and Hessian) of $\log \mathbf{L}$ is very cumbersome to compute, inhibiting the implementation of a 'steepest-ascent' algorithm.

For this reason, we now propose a numerical MLE algorithm based on an approximate 'steepest ascent' together with a 'divide-and-conquer' idea. To easily describe it, let

$$\theta^{(j)} = (A^{(j)}, B^{(j)}, C^{(j)}, a_0^{(j)})' \text{ for } \quad j = 1, \ldots, m$$

In addition, divide the parameters in the (lower triangular part of) matrix $\mathbf{P}$ in M groups labeled $\theta^{(m+1)}, \theta^{(m+2)}, \ldots \theta^{(m+M)}$, where M is such that each of the above groups does not contain more than (say) three or four parameters; for example, as long as $m \leq 3$, M can be taken to equal 1.

Our basic assumption is that we have at our disposal a numerical optimization routine that will maximize $\log \mathbf{L}$ over $\theta^{(j)}$, for every j, when all other $\theta^{(i)}$ with $i \neq j$ are held fixed. The proposed DC–profile MLE algorithm is outlined below.[6]

4.1. DC–Profile MLE Algorithm

Step 0. Let $\theta_0^{(j)}$ for $j = 1, \ldots, m+M$ denote starting values for $\theta^{(j)}$.

Step k + 1. At this step, we assume we have at our disposal the kth approximate MLE solution $\theta_k^{(j)}$ for $j = 1, \ldots, m+M$, and want to update it to obtain the $(k+1)$th approximation $\theta_{k+1}^{(j)}$. To do this, maximize $\log \mathbf{L}$ over $\theta^{(j)}$, for every j, when all other $\theta^{(i)}$ with $i \neq j$ are held fixed at $\theta_k^{(i)}$; denoted by $\tilde{\theta}_k^{(j)}$ for $j = 1, \ldots, m+M$ the results of the $m+M$ 'profile' maximizations. We then define the updated value $\theta_{k+1}^{(j)}$ for $j = 1, \ldots, m+M$ as a 'move in the right direction', i.e.,

$$\theta_{k+1}^{(j)} = \lambda \tilde{\theta}_k^{(j)} + (1-\lambda)\theta_k^{(j)} \tag{20}$$

i.e., a convex combination of $\tilde{\theta}_k^{(j)}$ and $\theta_k^{(j)}$. The parameter $\lambda \in (0,1)$ is chosen by the practitioner; it controls the speed of convergence – as well as the risk of nonconvergence – of the algorithm. A value λ close to one intuitively corresponds to maximum potential speed of convergence but also maximum risk of nonconvergence.

Stop. The algorithm stops when relative convergence has been achieved, i.e., when $||\theta_{k+1} - \theta_k||/||\theta_k||$ is smaller than some prespecified small number; here $|| \cdot ||$ can be the Euclidean or other norm.
The aforementioned maximizations involve numerical searches over a given range of parameter values; those ranges must also be carefully chosen. A concrete recommendation that has been found to speed up convergence of the profile MLE algorithm is to let the range in searching for $\theta_{k+1}^{(J)}$ be centered around the previous value $\theta_k^{(J)}$, i.e., to be of the type $[\theta_k^{(J)} - b_k, \theta_k^{(J)} + B_k]$ for two positive sequences b_k, B_k that tend to zero for large k reflecting the 'narrowing-down' of the searches. For positive parameters, e.g., for $\theta_k^{(J)}$ with $J \leq m$ in our context, the range can be taken as $[d_k\ \theta_k^{(J)}, D_k\ \theta_k^{(J)}]$ for two sequences d_k, D_k. The equence d_k is increasing, tending to 1 from below, while D_k is a decreasing sequence tending to 1 from above; the values $d_1 = 0.7$, $\mathrm{D}_1 = 1.5$ are reasonable starting points.

Remark. The DC–profile optimization algorithm may be applicable to general multiparameter optimization problems characterized by a difficulty in computing the gradient and Hessian of a general objective function $\mathbf{L}$. The idea is to 'divide' the arguments of $\mathbf{L}$ into (say) the m groups $\theta^{(j)}$ for $j = 1, \ldots, m$. The number of parameters within each group must be small enough such that a direct numerical search can be performed to accomplish ('conquer') the optimization of $\mathbf{L}$ over $\theta^{(j)}$, for every j, when all other $\theta^{(i)}$ with $i \neq j$ are held fixed. The m individual ('profile') optimizations are then combined as in Eq.(20) to point to a direction of steep (if not steepest) ascent.

5. THE BIVARIATE PARADIGM

For illustration, we now focus on a random vector $\mathbf{U} = (U^{(1)}, U^{(2)})$ that is distributed according to the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$ in the bivariate case $m = 2$. In this case, the correlation matrix

$$\mathrm{P} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

where ρ is a correlation coefficient. Interestingly, as in the case of normal data, the strength of dependence between $U^{(1)}$ and $U^{(2)}$ is measured by this single number ρ. In particular, $U^{(1)}$ and $U^{(2)}$ are independent if and only if $\rho = 0$.

As mentioned before, if $\mathrm{a}_0 \to 0$, then $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1 \to \phi_{\mathbf{p}}(\mathbf{u})$ In addition, $f_{\mathbf{p}}(\mathbf{u}; \mathbf{a}_0, 1)$ is unimodal, and has even symmetry around the origin. Nevertheless, when $\mathbf{a}_0 \neq 0$, the density $f_{\mathbf{p}}(\mathbf{u}; \mathbf{a}_0, 1)$ is heavy-tailed and nonnormal. To see this, note that the contours of the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$ with $\mathbf{a}_0 = (0.1, 0.1)'$ and $\rho = 0$ are not circular as in the normal, uncorrelated case; see Fig. 1(i).

Parts (ii), (iii), and (iv) of Fig. 1 deal with the case $\rho = 0.5$, and investigate the effect of the shape/tail parameter $\mathbf{a}_0$. Fig. 1 (iv) where $\mathbf{a}_0 = (0{,}0)'$ is the normal case with elliptical contours; in Fig. 1 (ii) we have $\mathbf{a}_0 = (0.1{,}0.1)'$, and the contours are 'boxy'-looking, and far from elliptical. The case of Fig. 1 (iii) is somewhere in-between since $a_0^{(1)} = 0$ but $a_0^{(2)} \neq 0$. By comparing the contours of $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$ to that of the normal $\phi_{\mathbf{P}}(\mathbf{u})$, we come to an interesting conclusion: the density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$ looks very much like the $\phi_{\mathbf{P}}(\mathbf{u})$ for small values of the argument $\mathbf{u}$. As expected, the similarity breaks down in the tails; in some sense, $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$ is like a normal density deformed in a way that its tails are drawn out.

Fig. 2 further investigates the role of the parameters $\mathbf{a}_0$ and ρ. Increasing the degree of heavy tails by increasing the value of $\mathbf{a}_0$ leads to some interesting contour plots, and this is especially true out in the tails.[7] Similarly, increasing the degree of dependence by increasing the value of ρ leads to unusual contour plots especially in the tails. As mentioned before, the contours of $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$ are close to elliptical for small values of $\mathbf{u}$, but acquire unusual shapes in the tails.

Finally, Fig. 2(iii) shows an interesting interplay between the values of $\mathbf{a}_0$ and the correlation coefficient ρ. To interpret it, imagine one has n i.i.d. samples from density $f_{\mathbf{P}}(\mathbf{u}; \mathbf{a}_0, 1)$. If n is small, in which case the heavy tails of $U^{(2)}$ have not had a chance to fully manifest themselves in the sample, a fitted regression line of $U^{(2)}$ on $U^{(1)}$ will have a small slope dictated by the

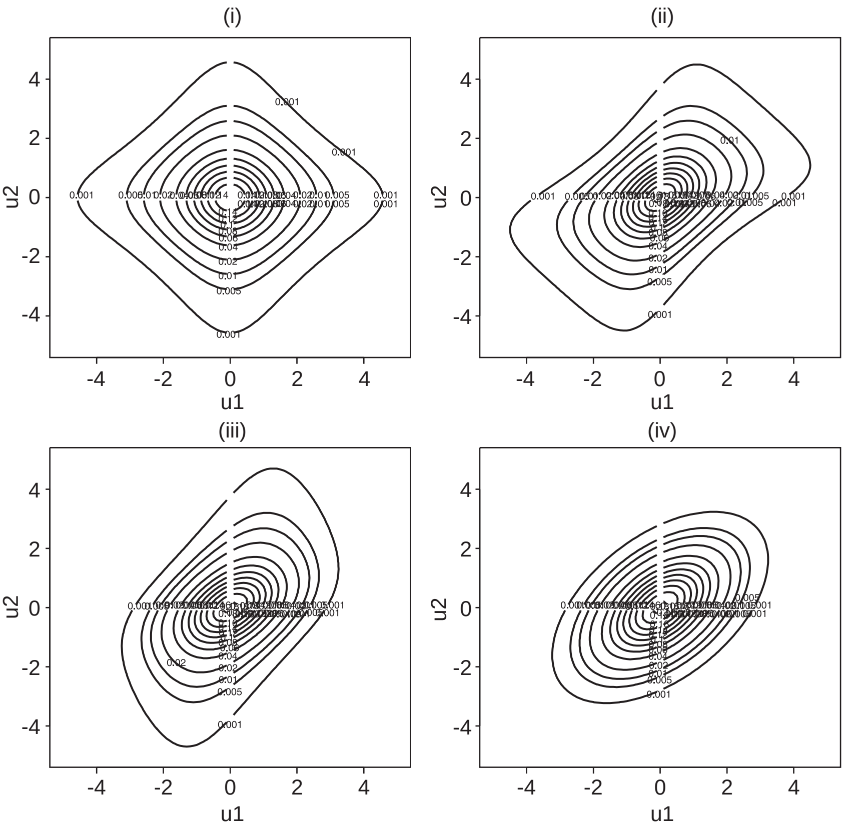


Fig. 1. Contour (Level) Plots of the Bivariate Density $f_{\mathbf{P}}(\mathbf{u};\, \mathbf{a}_0,1)$: (i) $a_0 = (0.1, 0.1)'$, $\rho = 0$; (ii) $a_0 = (0.1, 0.1)'$, $\rho = 0.5$; (iii) $a_0 = (0, 0.1)'$, $\rho = 0.5$; (iv) $a_0 = (0, 0)'$, $\rho = 0.5$.

elliptical contours around the origin. By contrast, when n is large, the slope of the fitted line will be much larger due to the influence of the tails of $f_{\mathbf{P}}(\mathbf{u};\, \mathbf{a}_0,\, 1)$ in the direction of $U^{(2)}$.

6. A REAL DATA EXAMPLE

We now focus on a real data example. The data consist of daily returns of the IBM and Hewlett-Packard stock prices from February 1, 1984 to

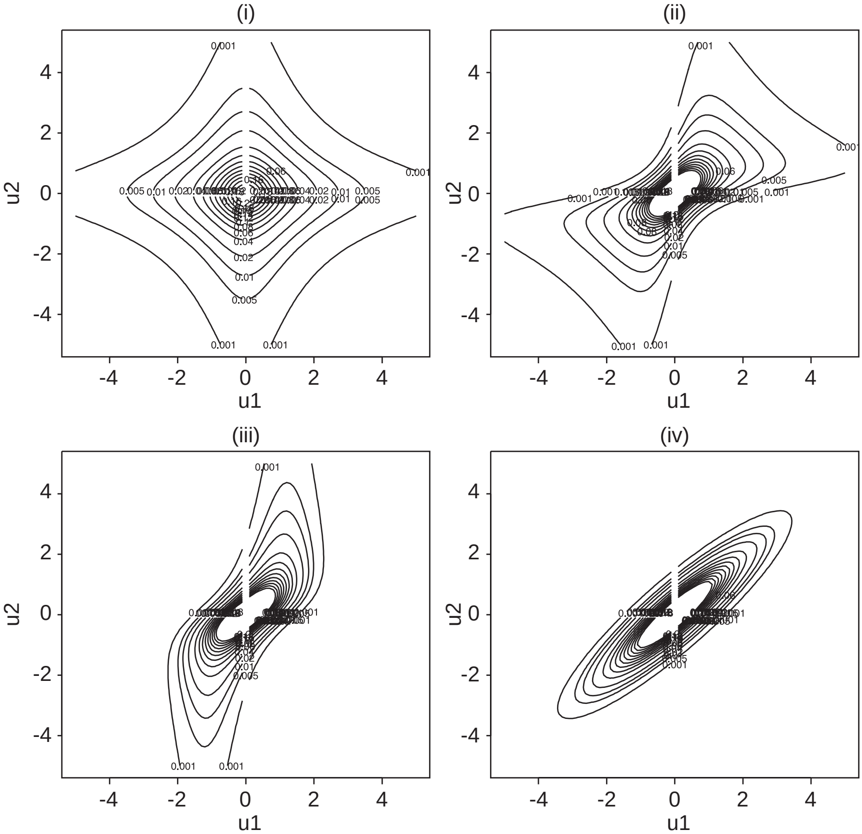


Fig. 2. Contour (Level) Plots of the Bivariate Density $f_{\mathbf{P}}(\mathbf{u};\, \mathbf{a}_0, 1)$: (i) $a_0 = (0.5, 0.5)'$, $\rho = 0$; (ii) $a_0 = (0.5, 0.5)'$, $\rho = 0.9$; (iii) $a_0 = (0, 0.5)'$, $\rho = 0.9$; (iv) $a_0 = (0, 0)'$, $\rho = 0.9$.

December 31, 1991; the sample size is $n = 2000$. This bivariate dataset is available as part of the garch module of the statistical language S+.

A full plot of the two time series is given in Fig. 3; the crash of October 1987 is prominent around the middle of those plots. Fig. 4 focuses on the behavior of the two time series over a narrow time window spanning two months before to two months after the crash of 1987. In particular, Fig. 4 gives some visual evidence that the two time series are not independent; indeed, they seem to be positively dependent, i.e., to have a tendency to 'move together'. This fact is confirmed by the autocorrelation/cross-correlation plot

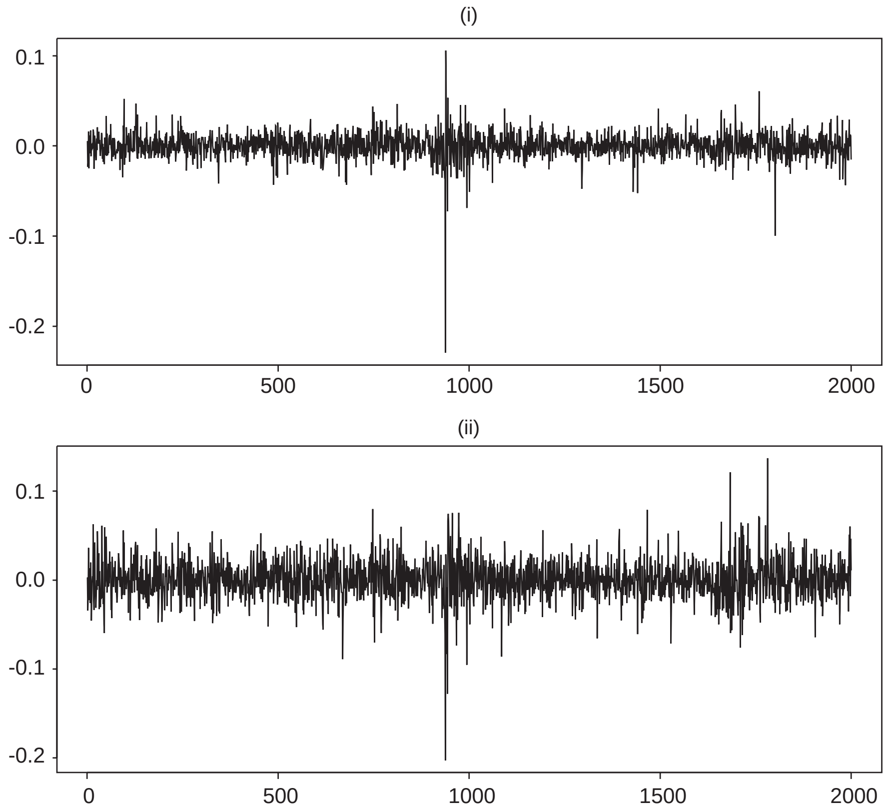


Fig. 3. Daily Returns from February 1, 1984 to December 31, 1991: (i) IBM and (ii) H-P Stock Returns.

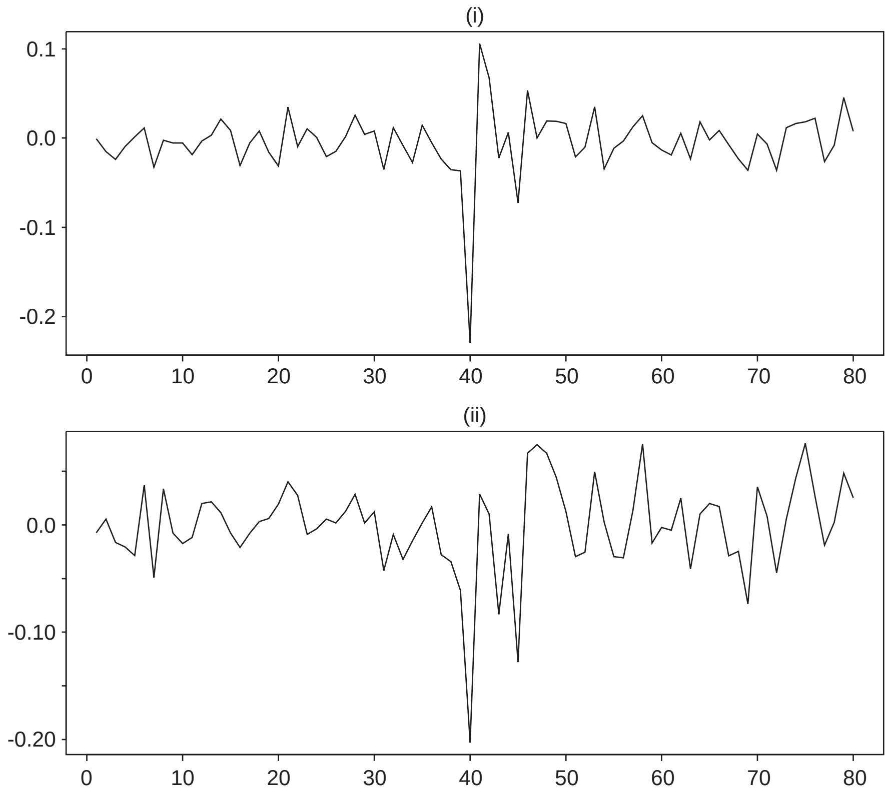


Fig. 4. Daily Returns from Two Months Before to Two Months After the Crash of 1987: (i) IBM and (ii) H-P Stock Returns.

given in Fig. 5; note that the cross-correlation of the two series at lag zero is large and positive (equal to 0.56).

Table 1 shows the MLEs of parameters in univariate GARCH(1,1) modeling of each returns series, while Table 2 contains the approximate MLEs of parameters in bivariate GARCH(1,1) modeling, i.e., entries of Table 2 were calculated using the DC-profile MLE algorithm of Section 4. Although more work on the numerical computability of multivariate MLEs in our context is in order – including gauging convergence and accuracy – it is interesting to note that the entries of Table 2 are markedly different from those of Table 1 indicating that the bivariate modeling is informative. For example, while the sum $\hat{A} + \hat{B}$ was found less than one for both series in the univariate modeling of Table 1, the same sum was estimated to be slightly over one for the IBM series and significantly over one for the H-P series in Table 2; this finding reinforces the viewpoint that financial returns may be lacking a finite second moment – see the discussion in Section 2 and Politis (2004). In addition, the high value of $\hat{\rho}$ in Table 2 confirms the intuition that the two series are strongly (and positively) dependent.

Table 1 shows the MLEs of parameters in univariate GARCH(1,1) modeling of each returns series, while Table 2 contains the approximate MLEs of parameters in bivariate GARCH(1,1) modeling, i.e., entries of Table 2 were calculated using the DC-profile MLE algorithm of Section 4. Although more work on the numerical computability of estimators in this multivariate context is in order – including gauging convergence and accuracy – it is interesting to note that the entries of Table 2 are markedly different from those of Table 1 indicating that the bivariate modeling is informative. For example, while the sum $\hat{A} + \hat{B}$ was found less than one for both series in the univariate modeling of Table 1, the same sum was estimated to be slightly over one for the IBM series and significantly over one for the H-P series in Table 2; this finding reinforces the viewpoint that financial returns may be lacking a finite second moment – see the discussion in Section 2 and Politis (2004). In addition, the high value of $\hat{\rho}$ in Table 2 confirms the intuition that the two series are strongly (and positively) dependent.

Fig. 6 shows diagnostic plots associated with the DC-profile MLE algorithm. The first nine plots show the evolution of the nine parameter estimates over iterations; the nine parameters are: a_0, A, B, C for the IBM series; a_0, A, B, C for the H-P series, and ρ. Those nine plots show convergence of the DC–profile MLE procedure; this is confirmed by the 10th plot that shows the evolution of the value of the Log-Likelihood corresponding to those nine parameters. As expected, the 10th plot shows an increasing graph

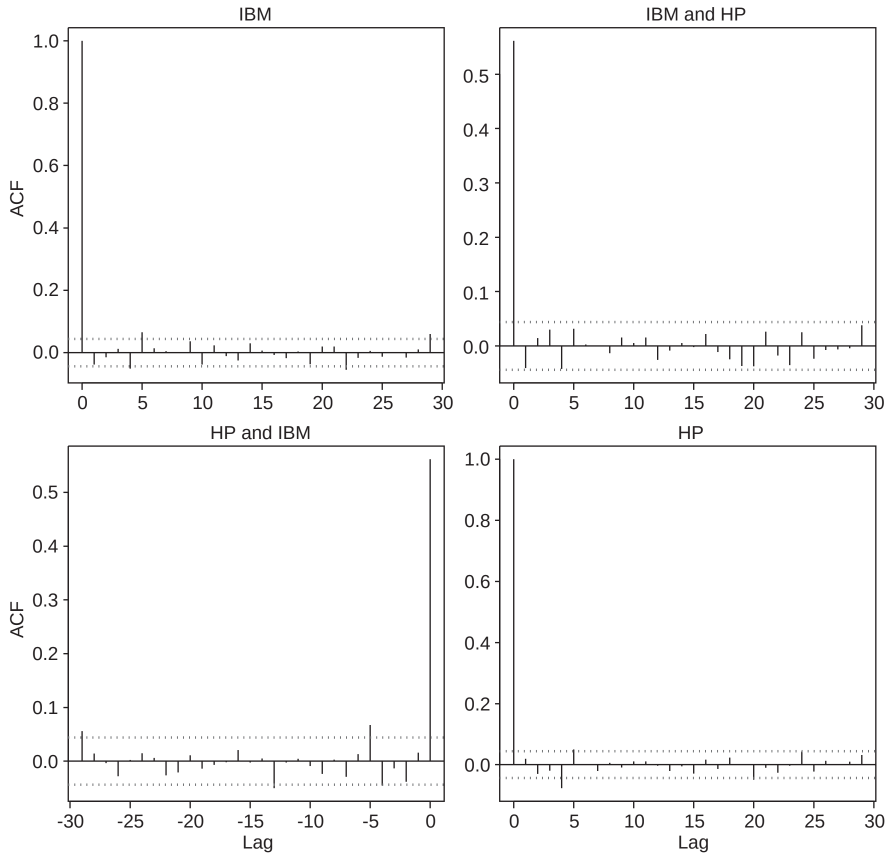


Fig. 5. Autocorrelation/Cross-Correlation Plot of IBM Vs. H-P Daily Returns from February 1, 1984 to December 31, 1991.

Table 1. MLEs of Parameters in *Univariate* GARCH(1,1) Modeling of Each Returns Series.

	$\hat{a}_0$	$\hat{A}$	$\hat{B}$	$\hat{C}$
IBM	0.066	0.029	0.912	6.32e-06
H-P	0.059	0.052	0.867	2.54e-05

Table 2. MLEs of Parameters in *Bivariate* GARCH(1,1) Modeling of the Two Series; Entries of Table 2 Were Calculated Using the DC-Profile MLE Algorithm of Section 4.

	$\hat{a}_0$	$\hat{A}$	$\hat{B}$	$\hat{C}$	$\hat{\rho}$
IBM	0.105	0.055	0.993	7.20e-05	0.53
H-P	0.062	0.258	0.975	3.85e-04	

that tapers off toward the right end, showing the maximization is working; only about 30–40 iteration steps were required for the convergence of the algorithm.

Finally, note that few people would doubt that a $\hat{\rho}$ value of 0.53, arising from a dataset with sample size 2000, might not be significantly different from zero; nonetheless, it is important to have a hypothesis testing procedure for use in other, less clear, cases. A natural way to address this is via the subsampling methodology – see Politis, Romano, and Wolf (1999) for details. There are two complications, however:

(i) Getting $\hat{\rho}$ is already quite computer-intensive, involving a numerical optimization algorithm; subsampling entails re-calculating $\hat{\rho}$ over subsets (subseries) of the data, and can thus be too computationally expensive.
(ii) The estimators figuring in Tables 1 and 2 are not $\sqrt{n}$–consistent as shown in Hall and Yao (2003); subsampling with estimated rate of convergence could then be employed as in Chapter 8 of Politis et al. (1999).

Future work may focus on the above two important issues as well as on the possibility of deriving likelihood-based standard errors and confidence intervals in this nonstandard setting.

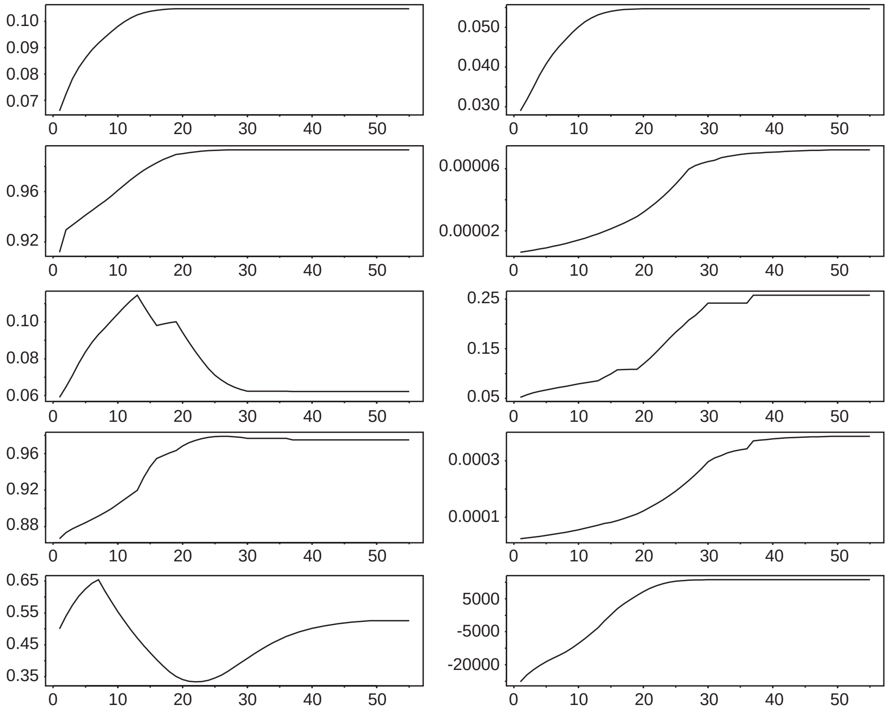


Fig. 6. Diagnostic Plots Associated with the Profile MLE Algorithm; the First Nine Plots Show the Evolution of the Nine Parameter Estimates over Iterations, While the 10th Plot Shows the Evolution of the Value of the Log-Likelihood.

NOTES

1. Some authors have raised the question of the existence of skewness in financial returns; see e.g. Patton (2002) and the references therein. Nevertheless, at least as a first approximation, the assumption of symmetry is very useful for model building.
2. To see why, note that the specification $\hat{a} = 0$, $\hat{a}_0 = 1$, and $\hat{a}_i = 0$ for $i > 0$, corresponds to $\hat{W}_t = sign(X_t)$ that has kurtosis equal to one. As previously mentioned, specifications with $\hat{a}_0 = 0$ typically yield residuals $\hat{W}_t$ with heavier tails than the normal, i.e., kurtosis bigger than three. Thus, by the intermediate value theorem, a specification for $\hat{a}, \hat{a}_0, \hat{a}_1, \ldots$ should exist that would render the kurtosis of $\hat{W}_t$ attaining the intermediate value of three.
3. Although strictly speaking the second moment is also infinite, since the moment of order 2-ε is finite for any $\varepsilon > 0$, one can proceed in practical work as if the second moment were indeed finite; for example, no finite-sample test can ever reject the hypothesis that data following the density $f(u; a_0, 1)$ with $a_0 > 0$ have a finite second moment.
4. An alternative avenue was recently suggested by Bose and Mukherjee (2003) but it is unclear if/how their method applies to a multivariate ARCH model without finite fourth moments.
5. As with all numerical optimization procedures, the algorithm must run with a few different starting values to avoid getting stuck at a local – but not global – optimum.
6. The initials DC stand for divide-and-conquer; some simple S+ functions for GARCH(1,1) numerical MLE in the univariate case $m = 1$ and the bivariate case $m = 2$ are available at: www.math.ucsd.edu/~politis.
7. In both Fig. 1 and 2 the same level values were used corresponding to the density taking the values 0.001, 0.005, 0.010, 0.020, 0.040, 0.060, 0.080, 0.100, 0.120, 0.140, 0.160, 0.180, and 0.200.

REFERENCES

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, *72*, 498–505.

Bollerslev, T., Chou, R., & Kroner, K. (1992). ARCH modelling in finance: a review of theory and empirical evidence. *Journal of Econometrics*, *52*, 5–60.

Bose, A., & Mukherjee, K. (2003). Estimating the ARCH parameters by solving linear equations. *Journal of Time Series Analls*, *24*(2), 127–136.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*, *50*, 987–1008.

Gouriéroux, C. (1997). *ARCH models and financial applications*. New York: Springer.

Hall, P., & Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, *71*, 285–317.

Patton, A. J. (2002). Skewness, asymmetric dependence and portfolios. Preprint (can be downloaded from: http://fmg.lse.ac.uk/~patton).

Politis, D. N. (2003). Model-free volatility prediction. Discussion Paper 2003-16, Department of Economics, UCSD.

Politis, D. N. (2004). A heavy-tailed distribution for ARCH residuals with application to volatility prediction. *Annals of Economics and Finance*, *5*, 283–298.

Politis, D. N., Romano, J. P., & Wolf, M. (1999). *Subsampling*. New York: Springer.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: D. R. Cox, D. V. Hinkley & O. E. Barndorff-Nielsen (Eds), *Time series models in econometrics, finance and other fields* (pp. 1–67). London: Chapman & Hall.

Vrontos, I. D., Dellaportas, P., & Politis, D. N. (2003). A full-factor multivariate GARCH model. *Econometrics Journal*, *6*, 312–334.

A PORTMANTEAU TEST FOR MULTIVARIATE GARCH WHEN THE CONDITIONAL MEAN IS AN ECM: THEORY AND EMPIRICAL APPLICATIONS

Chor-yiu Sin

ABSTRACT

Macroeconomic or financial data are often modelled with cointegration and GARCH (Generalized Auto-Regressive Conditional Heteroskedasticity). Noticeable examples include those studies of price discovery in which stock prices of the same underlying asset are cointegrated and they exhibit multivariate GARCH. It was not until recently that Li, Ling, and Wong's (2001) Biometrika, 88, 1135–1152, paper formally derived the asymptotic distribution of the estimators for the error-correction model (ECM) parameters, in the presence of conditional heteroskedasticity. As far as ECM parameters are concerned, the efficiency gain may be huge even when the deflated error is symmetrically distributed. Taking into consideration the different rates of convergence, this paper first shows that the standard distribution applies to a portmanteau test, even when the conditional mean is an ECM. Assuming the usual null of no multivariate GARCH, the performance of this test in finite samples is examined

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 125–151

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20005-1

through Monte Carlo experiments. We then apply the test for GARCH to the yearly or quarterly (extended) Nelson–Plosser data, embedded with some prototype multivariate models. We also apply the test to the intra-daily HSI (Hang Seng Index) and its derivatives, with the spread as the ECT (error-correction term). The empirical results throw doubt on the efficiency of the usual estimation of the ECM parameters, and more importantly, on the validity of the significance tests of an ECM.

1. INTRODUCTION

Throughout this paper, we consider an m-dimensional autoregressive (AR) process of $\{Y_t\}$, which is generated by

$$Y_t = \Phi_1 Y_{t-1} + \cdots + \Phi_s Y_{t-s} + \varepsilon_t \tag{1.1}$$

$$\varepsilon_t = V_{t-1}^{1/2} \eta_t \tag{1.2}$$

where ϕ_j's are constant matrices, $\eta_t = (\eta_{1t}, \ldots, \eta_{mt})'$ is a sequence of independently and identically distributed (i.i.d.) random $m \times 1$-vectors with zero mean and identity covariance matrix, V_{t-1} is measurable F_{t-1}, where $F_t = \sigma\{\eta_s, s = t, t-1, \ldots\}$. As a result, $E(\varepsilon_t | F_{t-1}) = 0$ and $E(\varepsilon_t \varepsilon_t' | F_{t-1}) = V_{t-1}$.

Assuming the ε_t's are i.i.d., under further conditions on Φ_j's (see Assumptions 2.1 and 2.2), Ahn and Reinsel (1990) (see also Johansen, 1996) show that, although some component series of $\{Y_t\}$ exhibit nonstationary behaviour, there are r linear combinations of $\{Y_t\}$ that are stationary. This phenomenon, which is called cointegration in the literature of economics, was first investigated in Granger (1983) (see also Engle & Granger, 1987). There are numerous economics models such as consumption function, purchasing-power parity, money-demand function, hedging ratio of spot and future exchange rates and yield curves of different terms of maturities. The partially nonstationary multivariate AR model or cointegrating time series models assuming constant V_{t-1} have been extensively discussed over the past 20 years. Other noticeable examples include Phillips and Durlauf (1986) and Stock and Watson (1993). An equivalent representation is called an error-correction model (ECM). See (2.3).

In this paper, we adopt the constant-correlation multivariate GARCH first suggested by Bollerslev (1990). More precisely, we assume that $V_{t-1} = D_{t-1} \Gamma D_{t-1}$, where Γ is a symmetric square matrix of constant correlations and $D_{t-1} = \mathrm{diag}(\sqrt{h_{1t-1}}, \ldots, \sqrt{h_{mt-1}})$ is a diagonal matrix of conditional standard deviations, where:

$$h_{it-1} = a_{i0} + \sum_{j=1}^{q} a_{ij}\varepsilon_{it-j}^2 + \sum_{k=1}^{p} b_{ik}h_{it-1-k}, i = 1, \ldots, m \tag{1.3}$$

Despite the fact that many papers, such as Engle and Patton (2004) and Hasbrouck (2003), consider a model of this kind, few if not none of them estimate model (1.1) with the consideration of GARCH. As suggested by Li, Ling, and Wong (2001), the efficiency gain in estimating the ECM parameters may be huge, even when the deflated error (η_t) is symmetrically distributed.

Using the probability theories developed in Chan and Wei (1988) and Ling and Li (1998), we extend Li, Ling and Wong's (2001) results to testing for multivariate GARCH. The null can be no multivariate GARCH, or it can be a specific multivariate GARCH. This paper differs from others on testing for multivariate time-varying covariance such as Li and McLeod (1981); Ling and Li (1997); Tse (2002) in a number of ways. First, we do not restrict Y_t to be stationary. As one can see in Section 3, that prohibits us from using the usual delta method in deriving the asymptotic distribution of the test statistic. Second, as we are not using the Taylor's expansion, we only assume finite fourth-order moments. Nevertheless, we maintain the assumption that η_t is symmetrically distributed. We leave the asymmetric case to further research.

This paper proceeds along the lines in Ling and Li (1997). Section 2 discusses the structure of models (1.1)–(1.3). An introduction on the test statistic for multivariate GARCH is also found there. Section 3 is the main focus of the paper, which contains a rigorous derivation of the asymptotic distribution of the test statistics. We report the Monte Carlo experiments and the empirical examples of yearly, quarterly as well as intra-daily data in Sections 4 and 5, respectively. Conclusions can be found in Section 6.

Throughout, $\rightarrow_{\mathscr{L}}$ denotes convergence in distribution, $O_p(1)$ denotes a series of random numbers that are bounded in probability and $o_p(1)$ denotes a series of random numbers converging to zero in probability.

2. BASIC PROPERTIES OF THE MODELS AND THE TEST STATISTIC

Denote L as the lag operator. Refer to (1.1) and define $\Phi(L) = I_m - \Sigma_{j=1}^{s}\Phi_j L^j$. We first make the following assumption:

Assumption 2.1. $|\phi(z)| = 0$ implies that either $|z| > 1$ or $z = 1$.

Define $W_t = Y_t - Y_{t-1}$, $\Phi_j^* = -\Sigma_{k=j+1}^{s}\Phi_k$ and $C = -\Phi(1) = -(I_m - \Sigma_{j=1}^{s}\Phi_j)$. By Taylor's formula, $\Phi(L)$ can be decomposed as:

$$\Phi(z) = (1-z)I_m - Cz - \sum_{j=1}^{s-1}\Phi_j^*(1-z)z^j \tag{2.1}$$

Thus, we can re-parameterize process (1.1) as the following full-rank form:

$$W_t = CY_{t-1} + \sum_{j=1}^{s-1}\Phi_j^* W_{t-j} + \varepsilon_t \tag{2.2}$$

Following Ahn and Reinsel (1990) and Johansen (1996), we can decompose $C = AB$, where A and B are $m \times r$, and $r \times m$ matrices of rank r respectively. Thus (2.2) can be written as the following reduced-rank form:

$$W_t = ABY_{t-1} + \sum_{j=1}^{s-1}\Phi_j^* W_{t-j} + \varepsilon_t \tag{2.3}$$

Define $d = m - r$. Denote $B_\perp$ as a $d \times m$ matrix of full rank such that $BB'_\perp = 0_{r\times d}$, $\bar{B} = (BB')^{-1}B$ and $\bar{B}_\perp = (B_\perp B'_\perp)^{-1}B_\perp$, and $A_\perp$ as an $m \times d$ matrix of full rank such that $A'A_\perp = 0_{r\times d}$, $\bar{A} = A(A'A)^{-1}$ and $\bar{A}_\perp = A_\perp(A'_\perp A_\perp)^{-1}$. We impose the following conditions:

Assumption 2.2. $|A'_\perp(I_m - \Sigma_{j=1}^{s-1}\Phi_j^*)B'_\perp| \neq 0$.

Assumption 2.3. For $i = 1, \ldots, m$, $a_{i0} > 0$, $a_{i1}, \ldots, a_{iq}, b_{i1}, \ldots, b_{ip} \geq 0$ and $\Sigma_{j=1}^{q} a_{ij} + \Sigma_{k=1}^{p} b_{ik} < 1$.

Assumption 2.4. For $i = 1, \ldots, m$, all eigenvalues of $E(A_{it} \times A_{it})$ lie inside the unit circle, where $\times$ denotes the Kronecker product and

$$A_{it} = \begin{pmatrix} a_{i1}\eta_{it}^2 & \ldots & a_{iq}\eta_{it}^2 & b_{i1}\eta_{it}^2 & \ldots & b_{ip}\eta_{it}^2 \\ & I_{q-1} & 0_{(q-1)\times 1} & & 0_{(q-1)\times p} & \\ a_{i1} & \ldots & a_{iq} & b_{i1} & \ldots & b_{ip} \\ & 0_{(p-1)\times q} & & & I_{p-1} & 0_{(p-1)\times 1} \end{pmatrix}$$

Assumption 2.5. η_t is symmetrically distributed.

Assumption 2.6. For $i \neq j$, $E[\eta_{it}^k \eta_{jt}^l] = E[\eta_{it}^k]E[\eta_{jt}^l]$, $k, l = 1, 2, 3$.

Assumptions 2.3–2.4 are the necessary and sufficient conditions for $E(\text{vec}[\varepsilon_t\varepsilon_t']\text{vec}[\varepsilon_t\varepsilon_t']') < \infty$. Assumption 2.5 allows the parameters in (1.1) and those in (1.2)–(1.3) to be estimated separately without altering the asymptotic distributions. As a result, the asymptotic distribution of the estimator in (1.2)–(1.3) is unaltered, no matter we consider the full-rank estimation (estimating the parameters in (2.2)) or the reduced-rank estimation (estimating the parameters in (2.3)). In the balance of this paper, *without loss of generality*, we confine our attention to the full-rank estimation. On the other hand, as one can see in Lemmas 3.1 and 3.2, Assumption 2.6 simplifies the asymptotic distribution of our portmanteau test.

Refer to Processes (2.2) and (1.2)–(1.3). Consider the full-rank estimators for $\varphi \equiv \mathrm{vec}[C, \Phi_1^*, \ldots, \Phi_{s-1}^*]$ and $\delta \equiv [\delta_1', \delta_2']'$. $\delta_1 \equiv [a_0', a_1', \ldots, a_q', b_1', \ldots, b_p']'$, $a_j \equiv [a_{1j}, \ldots, a_{mj}]'$, $b_k \equiv [b_{1k}, \ldots, b_{mk}]'$, $j = 0, 1, \ldots, q, k = 1, \ldots, p$ and $\delta_2 \equiv \tilde{v}(\Gamma)$, which is obtained from $\mathrm{vec}(\Gamma)$ by eliminating the supradiagonal and the diagonal elements of Γ (see Magnus, 1988, p. 27). Given $\{Y_t : t = 1, \ldots, n\}$, conditional on the initial values $Y_s = 0$ for $s \le 0$, the log-likelihood function, as a function of the *true* parameter, can be written as:

$$l(\varphi, \delta) = \sum_{t=1}^{n} l_t \quad \text{and} \quad l_t = -\frac{1}{2}\varepsilon_t' V_{t-1}^{-1} \varepsilon_t - \frac{1}{2}\ln|V_{t-1}| \tag{2.4}$$

where $V_{t-1} = D_{t-1}\Gamma D_{t-1}$, $D_{t-1} = \mathrm{diag}(\sqrt{h_{1t-1}}, \ldots, \sqrt{h_{mt-1}})$. Further denote $h_{t-1} = (h_{1t-1}, \ldots, h_{mt-1})'$ and $H_{t-1} = (h_{1t-1}^{-1}, \ldots, h_{mt-1}^{-1})'$.

As proved in Theorem 3.1(c) of Sin and Ling (2004), the asymptotic distribution of $\hat{\delta}$, which is the QMLE (quasi-maximum likelihood estimator) or the one-step-iteration estimator for δ, is independent of that of $\hat{\varphi}$, the QMLE or the one-step iteration estimator for φ. More precisely:

$$\sqrt{n}\left(\hat{\delta} - \delta\right) = \left(n^{-1}\sum_{t=1}^{n} S_t\right)^{-1}\left(n^{-1/2}\sum_{t=1}^{n} \nabla_\delta l_t\right) + o_p(1) \to_{\mathscr{L}} N\left(0, \Omega_\delta^{-1}\Omega_\delta^*\Omega_\delta^{-1}\right) \tag{2.5}$$

where $\Omega_\delta^* = E(\nabla_\delta l_t \nabla_\delta' l_t)$, $\Omega_\delta = -E(S_t)$, $S_t = (S_{ijt})_{2\times 2}$,

$$\nabla_\delta l_t = \begin{pmatrix} -\frac{1}{2}\nabla_{\delta_1} h_{t-1}\left(\iota - w\left(\varepsilon_t\varepsilon_t' V_{t-1}^{-1}\right)\right) \cdot H_{t-1} \\ -\tilde{v}\left(\Gamma^{-1} - \Gamma^{-1} D_{t-1}^{-1}\varepsilon_t\varepsilon_t' D_{t-1}^{-1}\Gamma^{-1}\right) \end{pmatrix} \tag{2.6}$$

$$S_{11t} = -\left(\nabla_{\delta_1} h_{t-1}\right) D_{t-1}^{-2}\left(\Gamma^{-1} \cdot \Gamma + I_m\right) D_{t-1}^{-2}\left(\nabla_{\delta_1}' h_{t-1}\right)/4 \tag{2.7}$$

$$S_{12t} = -\left(\nabla_{\delta_1} h_{t-1}\right) D_{t-1}^{-2}\Psi_m\left(I_m \times \Gamma^{-1}\right) N_m \tilde{L}_m' \tag{2.8}$$

$$S_{22t} = -2\tilde{L}_m N_m\left[\Gamma^{-1} \times \Gamma^{-1}\right] N_m \tilde{L}_m' \tag{2.9}$$

$\iota = (1,1, \ldots ,1)'$ and $w(\chi)$ is a vector containing the diagonal elements of the square matrix χ. Ψ_m, N_m and $\tilde{L}_m$ are constant matrices (see Magnus, 1988, pp. 109, 48 & 96).

Let $\hat{\varepsilon}_t$ be the residual when the *true* parameter (φ,δ) is replaced by the QMLE or the one-step-iteration estimator $(\hat{\varphi}, \hat{\delta})$. Similarly, define $\hat{V}_{t-1}$. The lag l sum of squared (standardized) residual autocorrelation is thus defined as:

$$\hat{R}_l = n^{-1/2} \sum_{t=l+1}^{n} \left(\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - m\right)\left(\hat{\varepsilon}_{t-l}' \hat{V}_{t-1-l}^{-1} \hat{\varepsilon}_{t-l} - m\right) / n^{-1} \sum_{t=l+1}^{n} \left(\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - m\right)^2 \quad (2.10)$$

In the next section, we will consider the asymptotic distribution of the test statistic $\hat{R}$, defined as $(\hat{R}_1, \hat{R}_2, \ldots , \hat{R}_L)'$, where L is a fixed integer.

3. DISTRIBUTION OF THE SUM OF SQUARED RESIDUAL AUTOCORRELATIONS

We first note that for $l = 1, \ldots ,L$:

$$\begin{aligned}
\hat{P}_l &\equiv n^{-1/2} \sum_{t=1}^{n} (\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - m)(\hat{\varepsilon}_{t-l}' \hat{V}_{t-1-l}^{-1} \hat{\varepsilon}_{t-l} - m) \\
&= n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t' V_{t-1}^{-1} \varepsilon_t - m)(\varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l} - m) \\
&\quad + n^{-1/2} \sum_{t=1}^{n} (\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - \varepsilon_t' V_{t-1}^{-1} \varepsilon_t)(\varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l} - m) \\
&\quad + n^{-1/2} \sum_{t=1}^{n} (\varepsilon_t' V_{t-1}^{-1} \varepsilon_t - m)(\hat{\varepsilon}_{t-l}' \hat{V}_{t-1-l}^{-1} \hat{\varepsilon}_{t-l} - \varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l}) \\
&\quad + n^{-1/2} \sum_{t=1}^{n} (\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - \varepsilon_t' V_{t-1}^{-1} \varepsilon_t)(\hat{\varepsilon}_{t-l}' \hat{V}_{t-1-l}^{-1} \hat{\varepsilon}_{t-l} - \varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l}) \\
&\equiv P_l + I_1 + I_2 + I_3 \qquad (3.1)
\end{aligned}$$

Consider I_1. Observe that:

$$\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - \varepsilon_t' V_{t-1}^{-1} \varepsilon_t = \varepsilon_t' \left(\hat{V}_{t-1}^{-1} - V_{t-1}^{-1}\right) \varepsilon_t + (\hat{\varepsilon}_t - \varepsilon_t)' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t + \varepsilon_t' \hat{V}_{t-1}^{-1} (\hat{\varepsilon}_t - \varepsilon_t).$$

In view of the fact that $\hat{V}_{t-1}^{-1} = O(1)$ and $\hat{\varepsilon}_t - \varepsilon_t = O_p(\frac{1}{\sqrt{n}}) - O_p(\frac{1}{\sqrt{n}}) U_{t-1}$, where $U_{t-1} = [(BY_{t-1})', W_{t-1}', \ldots , W_{t-s+1}']'$ (see (A.1) in the appendix).

Further, $E[\varepsilon_t|F_{t-1}] = 0$. It is not difficult to see that:

$$n^{-1/2}\sum_{t=1}^{n}\left[(\hat{\varepsilon}_t-\varepsilon_t)'\hat{V}_{t-1}^{-1}\hat{\varepsilon}_t+\varepsilon_t'\hat{V}_{t-1}^{-1}(\hat{\varepsilon}_t-\varepsilon_t)\right]\left[\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right]=o_p(1)$$

In other words,

$$I_1 = n^{-1/2}\sum_{t=1}^{n}\left(\varepsilon_t'\left(\hat{V}_{t-1}^{-1}-V_{t-1}^{-1}\right)\varepsilon_t\right)\left(\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right)+o_p(1) \tag{3.2}$$

Next, note that:

$$\begin{aligned}\hat{V}_{t-1}^{-1}-V_{t-1}^{-1} = &-V_{t-1}^{-1}\left(\hat{V}_{t-1}-V_{t-1}\right)V_{t-1}^{-1}\\ &+V_{t-1}^{-1}\left(\hat{V}_{t-1}-V_{t-1}\right)V_{t-1}^{-1}\left(\hat{V}_{t-1}-V_{t-1}\right)\hat{V}_{t-1}^{-1}\end{aligned} \tag{3.3}$$

In view of the fact that $\hat{V}_{t-1}^{-1}=O(1)$, $V_{t-1}^{-1}=O(1)$ and $(\hat{V}_{t-1}-V_{t-1})=O_p(n^{-1/2})+O_p(\rho^t)$ (see Lemma A.3 in the Appendix), it is not difficult to see that:

$$\begin{aligned}&n^{-1/2}\sum_{t=1}^{n}\varepsilon_t'V_{t-1}^{-1}\left(\hat{V}_{t-1}-V_{t-1}\right)V_{t-1}^{-1}\left(\hat{V}_{t-1}-V_{t-1}\right)\\ &\times\hat{V}_{t-1}^{-1}\varepsilon_t\left(\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right)=o_p(1)\end{aligned} \tag{3.4}$$

Now we consider the first term on the right hand side of (3.3):

$$\begin{aligned}\hat{V}_{t-1}-V_{t-1}=&\left(\hat{D}_{t-1}-D_{t-1}\right)\Gamma D_{t-1}+D_{t-1}\Gamma\left(\hat{D}_{t-1}-D_{t-1}\right)\\ &+D_{t-1}\left(\hat{\Gamma}-\Gamma\right)D_{t-1}+I_{1t}\end{aligned}$$

where

$$\begin{aligned}I_{1t}\equiv&\left(\hat{D}_{t-1}-D_{t-1}\right)\left(\hat{\Gamma}-\Gamma\right)D_{t-1}+\left(\hat{D}_{t-1}-D_{t-1}\right)\Gamma\left(\hat{D}_{t-1}-D_{t-1}\right)\\ &+\hat{D}_{t-1}\left(\hat{\Gamma}-\Gamma\right)\left(\hat{D}_{t-1}-D_{t-1}\right)\end{aligned}$$

As $\hat{\Gamma}-\Gamma=O_p(n^{-1/2})$ and $\hat{D}_{t-1}-D_{t-1}=O_p(n^{-1/2})+O_p(\rho^t)$ (see Lemma A.2 in the Appendix), it is not difficult to see that:

$$n^{-1/2}\sum_{t=1}^{n}\varepsilon_t'V_{t-1}^{-1}I_{1t}V_{t-1}^{-1}\varepsilon_t\left(\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right)=o_p(1) \tag{3.5}$$

On the other hand, $\hat{D}_{t-1}-D_{t-1}=\frac{1}{2}D_{t-1}^{-1}[(\hat{D}_{t-1}^2-D_{t-1}^2)-(\hat{D}_{t-1}-D_{t-1})^2]$. Since $D_{t-1}^{-1}=O(1)$ and $\hat{D}_{t-1}-D_{t-1}=O_p(n^{-1/2})+O_p(\rho^t)$, similar to (3.5), it is not difficult to see that:

$$n^{-1/2}\sum_{t=1}^{n} \varepsilon_t' V_{t-1}^{-1} D_{t-1}^{-1}\left(\hat{D}_{t-1}-D_{t-1}\right)^2 V_{t-1}^{-1}\varepsilon_t\left(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right)=o_p(1) \tag{3.6}$$

All in all, by (3.2), (3.4), (3.5) and (3.6),

$$I_1 = I_{11}+I_{12}+I_{13}+o_p(1) \tag{3.7}$$

where

$$\begin{aligned}
I_{11} \equiv\ & -n^{-1/2}\sum_{t=1}^{n}\frac{1}{2}\operatorname{tr}\left[\varepsilon_t' V_{t-1}^{-1} D_{t-1}^{-1}\left(\hat{D}_{t-1}^2-D_{t-1}^2\right)\Gamma D_{t-1} V_{t-1}^{-1}\varepsilon_t\right]\\
& \times\left(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right);\\
I_{12} \equiv\ & -n^{-1/2}\sum_{t=1}^{n}\frac{1}{2}\operatorname{tr}\left[\varepsilon_t' V_{t-1}^{-1} D_{t-1}\Gamma D_{t-1}^{-1}\left(\hat{D}_{t-1}^2-D_{t-1}^2\right) V_{t-1}^{-1}\varepsilon_t\right]\\
& \times\left(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right);\ \text{and}\\
I_{13} \equiv\ & -n^{-1/2}\sum_{t=1}^{n}\frac{1}{2}\operatorname{tr}\left[\varepsilon_t' V_{t-1}^{-1} D_{t-1}\left(\hat{\Gamma}-\Gamma\right) D_{t-1} V_{t-1}^{-1}\varepsilon_t\right]\\
& \times\left(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l}-m\right).
\end{aligned}$$

We first consider I_{11}. For two matrices of the same order M_1 and M_2, $(\text{vec}M_1)'\text{vec}M_2 = \text{tr}M_1{}'M_2$ (see Magnus & Neudecker, 1988, p. 30). For a diagonal $m\times m$ matrix χ, $\text{vec}(\chi)=\Psi_m'\omega(\chi)$ (see Magnus, 1988, p. 109). Given Lemma A.1,

$$\begin{aligned}
&\operatorname{tr}\left[\varepsilon_t' V_{t-1}^{-1} D_{t-1}^{-1}\left(\hat{D}_{t-1}^2-D_{t-1}^2\right)\Gamma D_{t-1} V_{t-1}^{-1}\varepsilon_t\right]\\
&=\operatorname{tr}\left[\eta_t'\Gamma^{1/2} D_{t-1} D_{t-1}^{-1}\Gamma^{-1} D_{t-1}^{-1} D_{t-1}^{-1}\left(\hat{D}_{t-1}^2-D_{t-1}^2\right)\right.\\
&\left.\times\Gamma D_{t-1} D_{t-1}^{-1}\Gamma^{-1} D_{t-1}^{-1} D_{t-1}\Gamma^{1/2}\eta_t\right]\\
&=\operatorname{tr}\left[\Gamma^{1/2}\eta_t\eta_t'\Gamma^{-1/2} D_{t-1}^{-2}\left(\hat{D}_{t-1}^2-D_{t-1}^2\right)\right]\\
&=\left(\text{vec}D_{t-1}^{-2}\Gamma^{-1/2}\eta_t\eta_t'\Gamma^{1/2}\right)'\text{vec}\left(\hat{D}_{t-1}^2-D_{t-1}^2\right)\\
&=\left(\text{vec}D_{t-1}^{-2}\Gamma^{-1/2}\eta_t\eta_t'\Gamma^{1/2}\right)'\Psi_m'\left(\hat{h}_{t-1}-h_{t-1}\right)\\
&=\left[\Psi_m\text{vec}D_{t-1}^{-2}\Gamma^{-1/2}\eta_t\eta_t'\Gamma^{1/2}\right]'\left[\nabla_{\delta_1}' h_{t-1}\left(\hat{\delta}_1-\delta_1\right)+O_p\left(\rho^t\right)+A_{1t-1}\right]
\end{aligned} \tag{3.8}$$

Recall that $\Psi_m \mathrm{vec}(\cdot) = w(\cdot)$ (see Magnus, 1988, p. 110). From Section 2 (see around (2.4)), $H_{t-1} = (h_{1t-1}^{-1}, \ldots, h_{mt-1}^{-1})'$. Taking the conditional expectation of the first term on the right hand side of (3.8):

$$\Psi_m \mathrm{vec} D_{t-1}^{-2} \Gamma^{-1/2} E\left[\eta_t \eta_t' | F_{t-1}\right] \Gamma^{1/2} = \Psi_m \mathrm{vec} D_{t-1}^{-2} = H_{t-1} \tag{3.9}$$

Applying an ergodic theorem to $\{(\varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l} - m)(\Psi_m \mathrm{vec}\, D_{t-1}^{-2} \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{1/2})' \nabla_{\delta_1}' h_{t-1}\}$, by (3.8), (3.9) and Lemma A.1,

$$I_{11} = -\frac{1}{2} E\left[\left(\varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l} - m\right)\left(\nabla_{\delta_1} h_{t-1} H_{t-1}\right)'\right] \sqrt{n}\left(\hat{\delta}_1 - \delta_1\right) + o_p(1) \tag{3.10}$$

Next we turn to I_{12}. Refer to a term in I_{12}:

$$\begin{aligned}
&\mathrm{tr}\left[\varepsilon_t' V_{t-1}^{-1} D_{t-1} \Gamma D_{t-1}^{-1} \left(\hat{D}_{t-1}^2 - D_{t-1}^2\right) V_{t-1}^{-1} \varepsilon_t\right] \\
&\quad = \mathrm{tr}\left[D_{t-1}^{-1} \Gamma^{-1} D_{t-1}^{-1} D_{t-1} \Gamma^{1/2} \eta_t \eta_t' \Gamma^{1/2} D_{t-1} D_{t-1}^{-1} \Gamma^{-1} D_{t-1}^{-1} D_{t-1} \Gamma D_{t-1}^{-1} \right. \\
&\quad \left. \times \left(\hat{D}_{t-1}^2 - D_{t-1}^2\right)\right] \\
&\quad = \mathrm{tr}\left[D_{t-1}^{-2} \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{1/2} \left(\hat{D}_{t-1}^2 - D_{t-1}^2\right)\right]
\end{aligned}$$

In a similar token, it can be shown that:

$$\begin{aligned}
I_{12} = &-\frac{1}{2} E\left[\left(\varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l} - m\right)\left(\nabla_{\delta_1} h_{t-1} H_{t-1}\right)'\right] \sqrt{n}\left(\hat{\delta}_1 - \delta_1\right) \\
&+ o_p(1) = I_{11} + o_p(1)
\end{aligned} \tag{3.11}$$

Next we turn to I_{13}. Note that $\mathrm{vec}(\hat{\Gamma} - \Gamma) = 2N_m \tilde{L}'_m (\hat{\delta}_2 - \delta_2)$ (see Magnus, 1988, pp. 48, 96). Consider a term in I_{13}:

$$\begin{aligned}
&\mathrm{tr}\left[\varepsilon_t' V_{t-1}^{-1} D_{t-1} \left(\hat{\Gamma} - \Gamma\right) D_{t-1} V_{t-1}^{-1} \varepsilon_t\right] \\
&= \mathrm{tr}\left[D_{t-1} V_{t-1}^{-1} \varepsilon_t \varepsilon_t' V_{t-1}^{-1} D_{t-1} \left(\hat{\Gamma} - \Gamma\right)\right] \\
&= \mathrm{tr}\left[D_{t-1} D_{t-1}^{-1} \Gamma^{-1} D_{t-1}^{-1} D_{t-1} \Gamma^{1/2} \eta_t \eta_t' \Gamma^{1/2} D_{t-1} D_{t-1}^{-1} \Gamma^{-1} D_{t-1}^{-1} D_{t-1} \left(\hat{\Gamma} - \Gamma\right)\right] \\
&= \mathrm{tr}\left[\Gamma^{-1/2} \eta_t \eta_t' \Gamma^{-1/2} \left(\hat{\Gamma} - \Gamma\right)\right] \\
&= \left(\mathrm{vec} \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{-1/2}\right)' \mathrm{vec}\left(\hat{\Gamma} - \Gamma\right) \\
&= 2\left(\mathrm{vec} \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{-1/2}\right)' N_m \tilde{L}_m' \left(\hat{\delta}_2 - \delta_2\right) \\
&= 2\left(\tilde{L}_m N_m \mathrm{vec} \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{-1/2}\right)' \left(\hat{\delta}_2 - \delta_2\right)
\end{aligned} \tag{3.12}$$

Consider a term on the right hand side of (3.12):

$$E\left[\eta_t\eta_t'|F_{t-1}\right] = I_m \tag{3.13}$$

Applying an ergodic theorem to $\{(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l} - m)(\tilde{L}_m N_m \text{vec}\Gamma^{-1/2}\eta_t\eta_t'\Gamma^{-1/2})'\}$, by (3.12) and (3.13) and the fact that $E[\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l} - m] = 0$,

$$\begin{aligned} I_{13} = & -2E\left[\left(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l} - m\right)\left(\tilde{L}_m N_m \text{vec}\Gamma^{-1}\right)'\right]\sqrt{n}\left(\hat{\delta}_2 - \delta_2\right) \\ & + o_p(1) = o_p(1) \end{aligned} \tag{3.14}$$

Therefore, by (3.7), (3.10), (3.11) and (3.14),

$$\begin{aligned} I_1 = & -E\left[\left(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l} - m\right)\left(\left(\nabla_{\delta_1} h_{t-1} H_{t-1}\right)', 0_{1xm(m-1)/2}\right)\right]\sqrt{n}\left(\hat{\delta} - \delta\right) \\ & + o_p(1) \end{aligned} \tag{3.15}$$

Similarly, repeating the same steps for I_2, we would be able to show that:

$$\begin{aligned} I_2 = & -E\left[(\varepsilon_t' V_{t-1}^{-1}\varepsilon_t - m)\left(\left(\nabla_{\delta_1} h_{t-1-l} H_{t-1-l}\right)', 0_{1xm(m-1)/2}\right)\right]\sqrt{n}\left(\hat{\delta} - \delta\right) \\ & + o_p(1) = o_p(1) \end{aligned} \tag{3.16}$$

since $E[(\varepsilon_t' V_{t-1}^{-1}\varepsilon_t - m)|F_{t-1}] = 0$. Lastly, it can be shown in a similar way that:

$$I_3 = o_p(1) \tag{3.17}$$

Define $X_l' \equiv E[(\varepsilon_{t-l}' V_{t-1-l}^{-1}\varepsilon_{t-l} - m)((\nabla_{\delta_1} h_{t-1} H_{t-1})', 0_{1xm(m-1)/2})]$. By (3.1), (3.15), (3.16) and (3.17),

$$\hat{P}_l = P_l - X_l'\sqrt{n}\left(\hat{\delta} - \delta\right) + o_p(1) \tag{3.18}$$

In the balance of this paper, we let $\hat{P} = (\hat{P}_1, \hat{P}_2, \ldots, \hat{P}_L)'$, $P = (P_1, P_2, \ldots, P_L)'$ and $X = (X_1, X_2, \ldots, X_L)'$. By (3.18),

$$\hat{P} = P - X\sqrt{n}\left(\hat{\delta} - \delta\right) + o_p(1) \tag{3.19}$$

On the other hand, it is not difficult to show that:

$$n^{-1}\sum_{t=1}^{n}\left(\hat{\varepsilon}_t'\hat{V}_{t-1}^{-1}\hat{\varepsilon}_t - m\right)^2 = n^{-1}\sum_{t=1}^{n}\left(\varepsilon_t' V_{t-1}^{-1}\varepsilon_t - m\right)^2 + o_p(1)$$

Apply an ergodic theorem to $\{(\varepsilon_t V_{t-1}^{-1} \varepsilon_t - m)^2\}$. By the arguments in Ling and Li (1997, pp. 450–451),

$$n^{-1} \sum_{t=1}^{n} \left(\hat{\varepsilon}_t' \hat{V}_{t-1}^{-1} \hat{\varepsilon}_t - m \right)^2 = E\left[\left(\varepsilon_t' V_{t-1}^{-1} \varepsilon_t - m \right)^2 \right] + o_p(1) = cm + o_p(1) \tag{3.20}$$

where $c = E[\eta_{it}^4] - 1$.

The following lemma resembles Lemma 3.1 of Ling and Li (1997), which proof is straightforward and is omitted.

Lemma 3.1. For any constant vector $Z = (Z_1, Z_2, \ldots, Z_L)'$

$$Z'P = n^{-1/2} \sum_{t=L+1}^{n} x_t + o_p(1) \tag{3.21}$$

$$\text{where} \quad x_t = \sum_{l=1}^{L} Z_l \left(\varepsilon_t' V_{t-1}^{-1} \varepsilon_t - m \right) \left(\varepsilon_{t-l}' V_{t-1-l}^{-1} \varepsilon_{t-l} - m \right)$$

$$\text{and} \quad Ex_t^2 = (cm)^2 Z'Z < \infty. \ \square \tag{3.22}$$

Lemma 3.2. $E[\nabla_\delta l_t x_t] = cX'Z/2$, x_t is given in (3.22).

Proof. From (2.6) above and in view of the fact that $\varepsilon_t = D_{t-1} \Gamma^{1/2} \eta_t$,

$$\begin{aligned} \nabla_\delta l_t x_t = \Big[&-\left(\tfrac{1}{2} \nabla_{\delta_1} h_{t-1} \left(\iota - w\left(\Gamma^{1/2} \eta_t \eta_t' \Gamma^{-1/2} \right) \right) \odot H_{t-1} \right)', \\ &-\left(\tilde{v}\left(\Gamma^{-1} - \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{-1/2} \right) \right)' \Big] \\ &\left[\sum_{l=1}^{L} Z_l \left(\eta_t' \eta_t - m \right) \left(\eta_{t-l}' \eta_{t-l} - m \right) \right] \end{aligned}$$

Firstly, we note that $E[(\eta_t' \eta_t - m) | F_{t-1}] = 0$. Thus, as far as expectation is concerned, we can ignore the term ι in $(\iota - w(\Gamma^{1/2} \eta_t \eta_t' \Gamma^{-1/2}))$ and the term Γ^{-1} in $(\Gamma^{-1} - \Gamma^{-1/2} \eta_t \eta_t' \Gamma^{-1/2})$. Secondly, by Assumption 2.6, for $i \neq j$,

$$
\begin{aligned}
&E\left[\eta_{it}\eta_{jt}\left(\eta_t'\eta_t - m\right)|F_{t-1}\right] \\
&= E\left[\eta_{it}\eta_{jt}\left(\eta_t'\eta_t - m\right)\right] \\
&= E\left[\eta_{it}^3\eta_{jt}\right] + E\left[\eta_{it}\eta_{jt}^3\right] + E\left[\eta_{it}\eta_{jt}\sum_{k=1,k\neq i,k\neq j}^{m}\eta_{kt}^2\right] - mE\left[\eta_{it}\eta_{jt}\right] \\
&= E\left[\eta_{it}^3\right]E\left[\eta_{jt}\right] + E\left[\eta_{it}\right]E\left[\eta_{jt}^3\right] + E\left[\eta_{it}\right]E\left[\eta_{jt}\right]\sum_{k=1,k\neq i,k\neq j}^{m}E\left[\eta_{kt}^2\right] \\
&- mE\left[\eta_{it}\right]E\left[\eta_{jt}\right] \\
&= 0
\end{aligned}
$$

Consequently, following the lines around (3.8) of Ling and Li (1997), we are able to show that:

$$E[\eta_t\eta_t'(\eta_t'\eta_t - m)|F_{t-1}] = cI_m$$

Therefore, as $E\left[\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l} - m\right] = 0$,

$$
\begin{aligned}
E[\nabla_\delta l_t x_t] &= E\Big\{\Big[\Big(\tfrac{1}{2}\nabla_{\delta_1}h_{t-1}\Big(w(c\Gamma^{1/2}\Gamma^{-1/2}\Big)\Big)\cdot H_{t-1})', \Big(\tilde{v}\Big(c\Gamma^{-1/2}\Gamma^{-1/2}\Big)\Big)'\Big] \\
&\quad \left[\sum_{l=1}^{L}Z_l\left(\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l} - m\right)\right]\Big\} \\
&= \frac{c}{2}\sum_{l=1}^{L}Z_l E\left[\left(\varepsilon_{t-l}'V_{t-1-l}^{-1}\varepsilon_{t-l} - m\right)\left(\left(\nabla_{\delta_1}h_{t-1}H_{t-1}\right)', 0_{1xm(m-1)/2}\right)\right] \\
&= \frac{c}{2}\sum_{l=1}^{L}Z_l X_l' \\
&= \frac{c}{2}X'Z \ \square
\end{aligned}
$$

Lemma 3.3. The asymptotic joint distribution of P and $\sqrt{n}(\hat{\delta} - \delta)$ is normal with mean zero and covariance:

$$\begin{pmatrix} (cm)^2 I_L & cX\Omega_\delta^{-1}/2 \\ c\Omega_\delta^{-1}X'/2 & \Omega_\delta^{-1}\Omega_\delta^{*}\Omega_\delta^{-1} \end{pmatrix} \square$$

Proof. Given (2.5), Lemma 3.1 and Lemma 3.2, the proof is exactly the same as that of Lemma 3.3 in Ling and Li (1997). □

The asymptotic distribution of $\hat{R}$ defined around (2.10) is summarized in the following theorem.

Theorem 3.1. Define $\Omega \equiv I_L - X(c\Omega_{\delta}^{-1} - \Omega_{\delta}^{-1}\Omega_{\delta}^{*}\Omega_{\delta}^{-1})X'/(cm)^2$.

$$\hat{P} \to_{\mathscr{L}} N(0, (cm)^2\Omega)$$

$$\hat{R} \to_{\mathscr{L}} \mathrm{N}(0, \Omega) \;\square$$

Proof. This follows from (3.19), (3.20) and Lemma 3.3. $\square$

The asymptotic distribution of $\hat{R}$ in Theorem 3.1 is exactly the same as that in Theorem of Ling and Li (1997). On the one hand, we show that their *general* distribution applies to the case that the conditional mean is an ECM. On the other hand, we work out the *specific* distribution when the heteroskedasticity is the constant-correlation multivariate GARCH first suggested by Bollerslev (1990). Under the null of no multivariate GARCH, which is the usual null in practice, as $\nabla_{\delta_1} h_{t-1} H_{t-1} = 0$ (see (3.15)), $\Omega = I_L$ and the computation is much simplified. This is the case considered in the balance of the paper.

4. MONTE CARLO EXPERIMENTS

This section examines the performance of the portmanteau test statistic $\hat{R}$ in finite samples through Monte Carlo experiments. Throughout, we consider a bivariate series Y_t, which first-difference, similar to Section 2, is denoted as W_t.

The benchmark DGP (data generation process) is an ECM with homoskedasticity, denoted as ECM-HOMO. More precisely, the **ECM-HOMO** series is generated as:

$$\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix} = \begin{pmatrix} -0.4 \\ 0.12 \end{pmatrix} (1, -2.5) \begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{4.1}$$

where

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim \text{i.i.d. } N(0, V_*), \quad V_* = \begin{pmatrix} 25.0 & 5.4 \\ 5.4 & 9.0 \end{pmatrix} \tag{4.2}$$

For comparison with the usual stationary case, we also consider a stationary Y_t with homoskedasticity, denoted as **STAT-HOMO**, which is generated as:

$$\begin{pmatrix} W_{1t} \\ W_{2t} \end{pmatrix} = \begin{pmatrix} -0.7 & 0.3 \\ 0.3 & -0.7 \end{pmatrix} \begin{pmatrix} Y_{1t-1} \\ Y_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \tag{4.3}$$

where $(\varepsilon_{1t}, \varepsilon_{2t})'$ is distributed as in (4.2).

To investigate the power of the portmanteau test, we consider an ECM with heteroskedasticity, denoted as ECM-HETE, the mean of **ECM-HETE** is represented by (4.1) but $\varepsilon_t = D_{t-1}\Gamma^{1/2}\eta_t$ (see Section 1), where $\eta_t \sim$ i.i.d. $N(0, I_2)$, $D_{t-1} = \mathrm{diag}(\sqrt{h_{1t-1}}, \sqrt{h_{2t-1}})$,

$$\Gamma = \begin{pmatrix} 1 & 0.36 \\ 0.36 & 1 \end{pmatrix} \tag{4.4}$$

$$h_{1t} = 2.5 + 0.05\varepsilon_{1t-1}^2 + 0.85h_{1t-1} \tag{4.5}$$

$$h_{2t} = 0.9 + 0.05\varepsilon_{2t-1}^2 + 0.85h_{2t-1} \tag{4.6}$$

Similarly, we consider a stationary series with heteroskedasticity, denoted as **STAT-HETE**, the mean of which is represented by (4.3) but similar to ECM-HETE, the corresponding ε_t is generated by (4.4)–(4.6).

In order to investigate the importance of normality, we consider two other models **ECM-HOMO-t**, **ECM-HOMO-χ^2**, which are essentially generated as ECM-HOMO, but ε_t is generated from a t_5 distribution and from a de-meaned χ_1^2 distribution, respectively.

Similarly, we also consider the models **ECM-HETE-t**, **ECM-HETE-χ^2**, which are essentially generated as ECM-HETE, but η_t is generated from a t_5 distribution, and from a de-meaned χ_1^2 distribution, respectively.

For each of the 8 DGPs, we generate 50,000 replications of data with the sample size $n = 200$, 400, 800. The residuals from the estimations using Johansen's (1996) method are used to compute the $\hat{R}$ test statistic, where full-rank estimation ($r = 2$), reduced-rank estimation ($r = 1$) and zero-rank estimation ($r = 0$) are considered. The number of lags in each test statistic, L, ranges from 1 to 6. The rejection percentage, based on the 5% significance level, are reported in Tables 1 to 8.

Comparing Table 1 with Table 2, one can see that under the null of homoskedasticity and the correct specification of rank, the portmanteau test renders an appropriate size, though a bit under-rejects. Somewhat surprisingly, the under-rejection is more or less the same, no matter the DGP is an ECM, or is a stationary model, despite the fact that arguments for asymptotic approximation to normality are different. Both Tables 1, 2

Table 1. Rejection Percentage of an ECM-HOMO.

r	n	L					
		1	2	3	4	5	6
2	200	4.5960	4.3220	4.2820	4.2920	4.4640	4.4840
	400	4.6980	4.6320	4.5900	4.6580	4.7220	4.6620
	800	4.7880	4.8640	4.8960	4.9240	4.8860	4.7900
1	200	4.5600	4.2840	4.2920	4.2680	4.4800	4.4480
	400	4.6960	4.6420	4.6360	4.6640	4.7080	4.6840
	800	4.7740	4.8740	4.9100	4.8960	4.8980	4.8200
0	200	20.4320	16.1740	14.0320	12.7440	11.9060	11.1960
	400	37.0600	29.7980	25.5120	22.9900	21.1160	19.5520
	800	63.9720	54.6100	48.3660	44.1360	40.9620	38.1900

Number of replications = 50,000

Table 2. Rejection Percentage of a STAT-HOMO.

r	n	L					
		1	2	3	4	5	6
2	200	4.5500	4.3600	4.3660	4.3880	4.4660	4.4680
	400	4.6920	4.5780	4.5760	4.6680	4.6880	4.7360
	800	4.7600	4.7960	4.8800	4.8640	4.9120	4.7960
1	200	5.1280	5.2140	5.0700	5.0980	4.9780	4.9280
	400	6.0480	6.0520	5.8380	5.6460	5.4960	5.3980
	800	7.5600	7.3640	6.9080	6.8620	6.5520	6.3740
0	200	55.0140	45.0960	39.4660	35.7420	32.8580	30.5800
	400	85.5140	78.1440	72.6000	68.0940	64.4000	61.2220
	800	99.0780	97.8200	96.7260	95.2820	94.0180	92.7040

Number of replications = 50,000.

show that our portmanteau test has power against insufficient rank. Thus, to be conservative, one may want to estimate a full-rank model before testing for multivariate GARCH.

Next we consider Table 3 and Table 4. For $r = 2$, the power when the DGP is ECM-HETE is close to that when the DGP is STAT-HETE. Surprisingly, when $r = 1$, the power for ECM-HETE is slightly higher than that for STAT-HETE, although the latter is supposed to have power against insufficient rank. Further, it should be noted that there is not much variation in power across different L.

Table 3. Rejection Percentage of an ECM-HETE.

r	n	L					
		1	2	3	4	5	6
2	200	11.5460	14.1820	15.4020	16.2580	16.6720	16.7020
	400	20.6720	25.9420	28.7920	30.5160	31.2880	31.9160
	800	36.6200	46.8640	52.3960	55.4160	56.7440	57.7340
1	200	13.8620	16.4380	17.3380	17.9740	18.2500	18.1480
	400	25.5360	31.1600	33.6040	34.6480	34.8480	34.9540
	800	46.5840	56.4080	60.7780	62.2380	63.0320	63.2560
0	200	38.6980	35.9380	34.4340	33.1380	32.0920	31.3960
	400	66.9780	64.5820	62.9160	61.5300	60.0580	58.7940
	800	92.3920	91.7440	90.8700	90.1020	89.4520	88.6960

Number of replications = 50,000

Table 4. Rejection Percentage of a STAT-HETE.

r	n	L					
		1	2	3	4	5	6
2	200	10.6560	13.4260	14.8160	15.6180	16.1760	16.2580
	400	19.5860	25.1500	27.9820	29.7360	30.6420	31.0660
	800	35.7680	46.0260	51.5840	54.6000	56.0720	56.9620
1	200	11.5800	14.3740	15.6420	16.5060	16.9460	16.9400
	400	20.7460	26.0940	28.8800	30.6340	31.4580	31.9940
	800	36.7160	47.0180	52.4280	55.5200	56.8660	57.7060
0	200	66.9580	59.6700	55.7380	52.8500	50.4040	48.5240
	400	93.3040	90.1220	87.8480	85.8320	84.1380	82.3780
	800	99.8360	99.7100	99.5620	99.3980	99.2260	99.0440

Number of replications = 50,000.

Next we consider Tables 5 and 6. Consider $r = 2$ and $r = 1$ first. In spite that t_5 is a symmetric distribution while the de-meaned χ_1^2 is not, the slight over-rejection is more or less the same. In fact, when $n = 200$, the rejection percentage for ECM-HOMO-t is higher, though it is not the case when $n = 800$. When $r = 0$ with insufficient rank, our test has higher power against ECM-HOMO-χ^2.

Lastly, we consider Tables 7 and 8. We first consider $r = 2$ and $r = 1$. Unsurprisingly, there is not much variation in power across different r. Moreover, our test has higher power against ECM-HETE-t, the case with

Table 5. Rejection Percentage of an ECM-HOMO-t.

r	n	L					
		1	2	3	4	5	6
2	200	5.2940	6.5980	7.0880	7.5200	7.7900	7.9920
	400	4.5020	5.5460	6.0500	6.4940	6.8780	7.0140
	800	4.2740	5.2140	5.6580	6.1700	6.3940	6.5100
1	200	5.2840	6.6080	7.1240	7.5700	7.8420	7.9760
	400	4.5040	5.5800	6.0740	6.4680	6.9140	7.0340
	800	4.2720	5.2100	5.6600	6.1820	6.4140	6.5540
0	200	65.8920	58.1160	52.5940	48.6480	45.3660	42.8420
	400	86.4420	80.2620	75.1720	71.2960	68.1200	65.2180
	800	98.2520	96.5080	94.6800	93.0180	91.4140	89.9280

Number of replications = 50,000.

Table 6. Rejection Percentage of an ECM-HOMO-χ^2.

r	n	L					
		1	2	3	4	5	6
2	200	3.6900	5.1420	5.9180	6.4180	6.8640	7.0440
	400	3.5380	4.8520	5.8100	6.4120	6.9860	7.2580
	800	3.7420	4.9060	5.8820	6.4860	6.9680	7.3940
1	200	3.6200	5.0700	5.8800	6.3300	6.7980	6.9840
	400	3.5060	4.8240	5.8120	6.3720	6.9940	7.2880
	800	3.7180	4.9180	5.9160	6.5020	6.9580	7.3920
0	200	78.3560	68.5380	62.0940	57.2480	53.4800	50.3060
	400	96.6360	93.0680	89.6980	86.5640	83.5600	81.0420
	800	99.9320	99.7260	99.4640	99.0400	98.7300	98.2700

Number of replications = 50,000.

both heteroskedasticity and heavy tail. When $r = 0$ with insufficient rank, our test has higher power against ECM-HOMO-χ^2.

5. EMPIRICAL EXAMPLES

In this section, we apply our portmanteau test to three data sets, which are often estimated with an ECM, without the consideration of multivariate GARCH. The first data set is the *yearly* Nelson–Plosser data (see Nelson & Plosser, 1982). The second data set is the *quarterly* extended Nelson–Plosser

Table 7. Rejection Percentage of an ECM-HETE-t.

r	n	L					
		1	2	3	4	5	6
2	200	26.3960	34.9780	39.9920	42.8700	44.4020	45.3320
	400	42.2580	54.7840	61.6880	65.3880	67.5860	68.6660
	800	65.6700	79.6660	85.5620	88.4300	89.9900	90.9080
1	200	26.5300	35.1460	40.1600	43.0800	44.6240	45.5260
	400	42.2700	54.8880	61.7880	65.5860	67.6260	68.7520
	800	65.6880	79.6760	85.5240	88.4580	90.0160	90.9460
0	200	84.8420	84.7160	83.8580	83.1000	82.2040	81.3340
	400	97.6380	97.8660	97.5560	97.3420	97.1460	96.8680
	800	99.9100	99.9440	99.9640	99.9620	99.9500	99.9380

Number of replications = 50,000.

Table 8. Rejection Percentage of an ECM-HETE-χ^2.

r	n	L					
		1	2	3	4	5	6
2	200	15.8220	21.8860	25.7100	28.1700	29.9800	30.8400
	400	29.0340	40.5600	47.4020	51.6440	54.8200	56.5240
	800	51.7240	67.8700	76.0020	80.5460	83.1420	84.7520
1	200	15.9940	22.1140	25.8900	28.3520	30.1300	30.9900
	400	29.3840	41.0320	47.8700	52.1520	55.2120	56.9980
	800	52.1300	68.2660	76.4260	80.9080	83.5760	85.0640
0	200	94.8800	92.9540	90.8140	88.9260	87.4240	86.2360
	400	99.7920	99.7460	99.6100	99.5020	99.3780	99.2080
	800	99.9980	99.9960	99.9980	100.0000	100.0000	99.9980

Number of replications = 50,000.

data (see Schotman & van Dijk, 1991). The third data set is the 30-min intra-daily HSI (Hang Seng Index), HSIF (HSI Futures) and TraHK (Tracker Fund of Hong Kong), an ETF (exchange traded fund) for HSI, which are provided by the *Hong Kong Stock Exchange and Clearing Limited*. For each model, the residuals from the rank estimation using Johansen's (1996) method are used to compute the $\hat{R}$ test statistic, where full-rank estimation ($r = m$, where m is the number of variables in the model), reduced-rank estimation ($r<m, r = m-1, \ldots, 1$) and zero-rank estimation ($r = 0$) are considered, all of which with VAR order ranging from 1 to 6. The number of lags in each test statistic, L, also ranges from 1 to 6.

For the yearly Nelson–Plosser data or the quarterly extended Nelson–Plosser data, we consider the following models:

Model (1a)/(2a): Y = (Unemployment Rate, CPI, Nominal Wage), all are in logs.
Model (1b)/(2b): Y = (Money Stock, GNP Deflator, Real GNP, Interest Rate), all except Interest Rate are in logs.
Model (1c)/(2c): Y = (Stock Price, Interest Rate), Stock Price is in logs.
Model (1d)/(2d): Y = (Industrial Production, Employment), all are in logs.
The yearly data go from 1909 to 1970, which amounts to 62 observations, while the quarterly data go from 1969Q1 to 1988Q4, which amounts to 80 observations.

For the intra-daily HSI data, we consider the following models:

Model (3a): Y = (HSI, HSIF), 9/24/99–11/11/99, before the listing of TRaHK.
Model (3b): Y = (HSI, HSIF), 11/12/99–12/29/99, after the listing of TRaHK.
Model (3c): Y = (HSI, HSIF, TraHK), 11/12/99–12/29/99, after the listing of TRaHK.
There are altogether 263 30-min data points before the listing and 261 30-min data points after the listing.

Tables 9–11 depict the portmanteau test statistics for different models. For each model, we report the statistic from full-rank estimation and the

Table 9. Test Statistics (Nelson and Plosser Yearly Data).

	r	Order	L					
			1	2	3	4	5	6
Model (1a)	3	6	12.8599	18.1148	24.0637	24.6242	24.9089	25.1327
	3	6	12.8599	18.1148	24.0637	24.6242	24.9089	25.1327
Model (1b)	4	6	8.1409	9.7717	9.9800	10.5655	12.4148	18.1311
	4	2	10.6578	14.1278	15.0724	16.0449	21.7625	27.9040
Model (1c)	2	6	2.9099	3.2514	3.2929	3.3235	3.3635	4.0028
	2	1	5.3201	6.4464	6.4745	6.7619	6.8520	6.8754
Model (1d)	2	6	12.8884	18.3267	18.3604	19.0747	20.9822	26.0623
	2	1	12.4447	18.2687	18.3154	18.3177	19.7787	21.0559

Number of observations = 62.

Table 10. Test Statistics (Extended Nelson and Plosser Quarterly Data).

	r	Order	L					
			1	2	3	4	5	6
Model (2a)	3	6	22.8602	38.6512	61.7940	73.9797	79.7754	84.2333
	3	2	29.8081	51.6621	86.0356	108.2400	126.5856	141.4628
Model (2b)	4	6	9.1575	12.5316	13.4587	14.0846	16.2893	22.9415
	1	2	10.2212	12.7757	15.9917	15.9941	17.5293	20.6749
Model (2c)	2	6	9.9412	10.0168	11.6834	11.6905	11.6917	12.1481
	2	1	8.9845	9.2218	9.9830	9.9831	11.2145	14.5033
Model (2d)	2	6	7.9039	10.7708	10.8277	12.6325	13.9533	13.9535
	2	1	10.2604	13.9844	13.9846	14.3856	14.4552	14.4996

Number of observations = 80.

Table 11. Test statistics (HSI and Its Derivatives, Intradaily Data).

	r	Order	L					
			1	2	3	4	5	6
Model (3a)	2	6	28.4225	43.7126	60.1957	70.6273	78.2277	89.7487
	2	3	27.3509	42.1721	58.7319	67.8527	73.8960	84.3190
Model (3b)	2	6	30.5071	66.6060	80.3720	95.5618	106.4083	115.6762
	2	3	24.8093	61.2617	72.5741	84.4421	90.5341	98.3391
Model (3c)	3	6	18.7755	43.8952	51.5977	58.9116	62.4958	64.9495
	3	6	18.7755	43.8952	51.5977	58.9116	62.4958	64.9495

maximum VAR-order of 6, as well as the combination of cointegrating rand and VAR-order that renders the lowest Akaike Information Criterion (AIC) value. The former minimizes the possibility of under-parameterization that will affect the size under the null of homoskedasticity, while the later minimizes the possibility of over-parameterization that will affect the power under the alternative of heteroskedasticity. All in all, our empirical results throw doubt on the efficiency of usual estimation of the ECM parameters, and more importantly, on the validity of the significance tests of an ECM.

From Tables 9–11, first we note that AIC often chooses $r = m$, that is the data are stationary. We should not take this seriously as, on the one hand this is not a formal test for stationarity or for cointegration. On the other, the AICs for some of reduced-rank models are close to minimum. Finally, one may conclude that there is multivariate GARCH in all the data, except somewhat surprisingly, possibly the quarterly Stock Price (in logs) and Interest Rate.

6. CONCLUSIONS

Macroeconomic or financial data are often modelled with cointegration and GARCH. Noticeable examples include those studies of price discovery, in which stock prices of the same underlying asset are cointegrated and they exhibit multivariate GARCH. It was not until recently that Li, et al.'s (2001) *Biometrika* paper formally derived the asymptotic distribution of the estimators for the ECM parameters, in the presence of conditional heteroskedasticity. As far as ECM parameters are concerned, the efficiency gain may be huge even when the deflated error is symmetrically distributed. Taking into consideration the different rates of convergence, this paper first shows that the standard distribution applies to a portmanteau test, even when the conditional mean is an ECM. Assuming the usual null of no multivariate GARCH, the performance of this test in finite samples is examined through Monte Carlo experiments. We then apply the test for GARCH to the *yearly* or *quarterly* (extended) Nelson–Plosser data, embedded with some prototype multivariate models. We also apply the test to the *intra-daily* HSI and its derivatives, with the spread as the ECT. The empirical results throw doubt on the efficiency of usual estimation of the ECM parameters, and more importantly, on the validity of the significance tests of an ECM.

ACKNOWLEDGEMENTS

This is a revision of the paper entitled, "Testing for Multivariate ARCH when the Conditional Mean is an ECM: Theory and Empirical Applications". We thank helpful comments from the editor (Dek Terrell), the anonymous referees, Shiqing LING and Heung WONG. This research is partially supported by the Hong Kong Research Grant Council competitive earmarked grant HKBU2014/02 H. The usual disclaimers apply.

REFERENCES

Ahn, S. K., & Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate models. *Journal of the American Statistical Association*, *85*, 813–823.

Bollerslev, T. (1990). Modelling the coherence in the short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics*, *72*, 498–505.

Chan, N. H., & Wei, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, *16*, 367–401.

Engle, R. F., & Granger, C. W. J. (1987). Cointegration and error correction: Representation, estimation and testing. *Econometrica*, *55*, 251–276.

Engle, R. F., & Patton, A. J. (2004). Impacts of trades in an error-correction model of quote prices. *Journal of Financial Markets*, *7*, 1–25.

Granger, C.W.J. (1983). *Cointegrated variables and error Correction models*. Discussion paper, USA: University of California at San Diego.

Hasbrouck, J. (2003). Intraday price formation in U.S. equity index markets. *Journal of Finance*, *58*, 2375–2399.

Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models*. New York: Oxford University Press.

Li, W. K., Ling, S., & Wong, H. (2001). Estimation for partially nonstationary multivariate autoregressive models with conditional heteroskedasticity. *Biometrika*, *88*, 1135–1152.

Li, W. K., & McLeod, A. I. (1981). Distribution of the Residual Autocorrelations in Multivariate ARMA Time Series Models. *Journal of the Royal Statistical Society Series B*, *43*, 231–239.

Ling, S., & Li, W. K. (1997). Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis*, *18*, 447–464.

Ling, S., & Li, W. K. (1998). Limiting distributions of maximum likelihood estimators for unstable ARMA models with GARCH errors. *Annals of Statistics*, *26*, 84–125.

Magnus, J. R. (1988). *Linear structures*. New York: Oxford University Press.

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. Chichester: Wiley.

Nelson, C. R., & Plosser, C. I. (1982). Trends and random walks in macroeconomic time series. *Journal of Monetary Economics*, *10*, 139–162.

Phillips, P. C. B., & Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *Review of Economic Studies*, *53*, 473–495.

Schotman, P. C., & van Dijk, H. K. (1991). On Bayesian routes to unit roots. *Journal of Applied Econometrics*, *6*, 387–401.

Sin, C.-y., & Ling, S. (2004). *Estimation and testing for partially nonstationary vector autoregressive models with GARCH. Revised and re-submitted to Journal of Econometrics.*

Stock, J. H., & Watson, M. W. (1993). A Simple estimator of cointegrating vectors in high order integrated system. *Econometrica*, *61*, 783–820.

Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal*, *5*, 358–373.

APPENDIX: LEMMAS

Given Assumptions 2.1–2.5, by Theorem 3.1 of Sin and Ling (2004), $\sqrt{n}(\hat{\delta} - \delta) = O_p(1)$. Denote $\hat{\varepsilon}_t = \varepsilon_t(\hat{\varphi})$ and $\varepsilon_t = \varepsilon_t(\varphi)$. As $\max_{1 \leq t \leq n} ||n^{-1/2} Z_{1t-1}|| = O_p(1)$,

$$\begin{aligned}\hat{\varepsilon}_t &= \varepsilon_t - [n(\hat{C} - C)P_1]\left[\frac{1}{n}Z_{1t-1}\right] - \left[\sqrt{n}(\hat{C} - C)P_2\right]\left[\frac{1}{\sqrt{n}}Z_{2t-1}\right] \\ &\quad - \sum_{j=1}^{s-1}\left[\sqrt{n}(\hat{\Phi}_j^* - \Phi_j^*)\right]\left[\frac{1}{\sqrt{n}}W_{t-j}\right] \\ &= \varepsilon_t - O_p\left(\frac{1}{\sqrt{n}}\right) - O_p\left(\frac{1}{\sqrt{n}}\right)U_{t-1} \end{aligned} \tag{A.1}$$

where we recall that $U_{t-1} = [(BY_{t-1})', W'_{t-1}, \ldots, W'_{t-s+1}]'$. All the $O_p(\cdot)$ and $o_p(\cdot)$, here and henceforth, hold uniformly in t and uniformly in $(\hat{\varphi}, \hat{\delta})$. Let ς stand for the series $\{\varepsilon_s = \varepsilon_s(\varphi)\}_{s=-\infty}^{n}$ and $\hat{\varsigma}$ stand for the series $\{\hat{\varepsilon}_s = \varepsilon_s(\hat{\varphi})\}_{s=1}^{n}$. Implicitly define $\hat{\gamma}_{ik}$ such that:

$$(1 - \sum_{j=1}^{p} \hat{b}_{ij}L^j)^{-1} = \sum_{k=0}^{\infty} \hat{\gamma}_{ik}L^k \tag{A.2}$$

γ_{ik} is defined similarly. Rewrite:

$$\begin{aligned} h_{it-1}(\hat{\varsigma}, \hat{\delta}_1) &= \sum_{k=0}^{t-1} \hat{\gamma}_{ik}\left[\hat{a}_{i0} + \left(\sum_{j=1}^{q} \hat{a}_{ij}\hat{\varepsilon}_{it-k-j}^2\right)\right] \\ h_{it-1}(\varsigma, \hat{\delta}_1) &= \sum_{k=0}^{t-1} \hat{\gamma}_{ik}\left[\hat{a}_{i0} + \left(\sum_{j=1}^{q} \hat{a}_{ij}\varepsilon_{it-k-j}^2\right)\right] \\ h_{it-1}^{\infty}(\varsigma, \hat{\delta}_1) &= \sum_{k=0}^{\infty} \hat{\gamma}_{ik}\left[\hat{a}_{i0} + \left(\sum_{j=1}^{q} \hat{a}_{ij}\varepsilon_{it-k-j}^2\right)\right] \\ h_{it-1}^{\infty}(\varsigma, \delta_1) &= \sum_{k=0}^{\infty} \gamma_{ik}\left[a_{i0} + \left(\sum_{j=1}^{q} a_{ij}\varepsilon_{it-k-j}^2\right)\right] \end{aligned} \tag{A.3}$$

where $\hat{\varepsilon}_{it}$ is the ith element of $\hat{\varepsilon}_t$, and ε_{it} is the ith element of ε_t.

In the rest of this appendix, when no ambiguity arises, denote $\hat{h}_{it-1} = h_{it-1}(\hat{\varsigma}, \hat{\delta}_1)$ and $h_{it-1} = h_{it-1}^{\infty}(\varsigma, \delta_1)$. Similarly, denote $\hat{D}_{t-1} = D_{t-1}(\hat{\varsigma}, \hat{\delta}_1)$, $D_{t-1} = D_{t-1}^{\infty}(\varsigma, \delta_1)$; $\hat{V}_{t-1} = V_{t-1}(\hat{\varsigma}, \hat{\delta}_1)$ and $V_{t-1} = V_{t-1}^{\infty}(\varsigma, \delta_1)$.

Lemma A.1. $\hat{h}_{t-1} - h_{t-1} = \nabla'_{\delta_1} h_{t-1}(\hat{\delta}_1 - \delta_1) + O_p(\rho^t) + A_{1t-1} + o_p(n^{-1/2}) = O_p(n^{-1/2}) + O_p(\rho^t)$, where $E[(\varepsilon'_{t-l}V_{t-1}^{-1}\varepsilon_{t-l} - m)H'_{t-1}A_{1t-1}] = 0$, $l = 0, 1, \ldots, L$. □

Proof. For an arbitrary i, define

$$\begin{aligned} h_{it-1}(\hat{\varsigma},\hat{\delta}_1) - h^{\infty}_{it-1}(\varsigma,\delta_1) &= (h_{it-1}(\hat{\varsigma},\hat{\delta}_1) - h_{it-1}(\varsigma,\hat{\delta}_1)) \\ &\quad + (h^{\infty}_{it-1}(\varsigma,\delta_1) - h_{it-1}(\varsigma,\hat{\delta}_1)) \\ &\quad + (h^{\infty}_{it-1}(\varsigma,\delta_1) - h^{\infty}_{it-1}(\varsigma,\delta_1)) \\ &\equiv A_{i1t-1} + A_{i2t-1} + A_{i3t-1} \end{aligned}$$

$$A_{i1t-1} = h_{it-1}(\hat{\varsigma},\hat{\delta}_1) - h_{it-1}(\varsigma,\hat{\delta}_1) = \sum_{k=0}^{t-1}\hat{\gamma}_{ik}\left[\hat{a}_{i0} + \sum_{j=1}^{q}\hat{a}_{ij}\left(\hat{\varepsilon}^2_{it-k-j} - \varepsilon^2_{it-k-j}\right)\right] \tag{A.4}$$

From (A.1) above,

$$\begin{aligned} \hat{\varepsilon}^2_{it} - \varepsilon^2_{it} &= \varepsilon_{it} O_p\left(\frac{1}{\sqrt{n}}\right) + \varepsilon_{it} O_p\left(\frac{1}{\sqrt{n}}\right) U_{t-1} \\ &\quad + O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{n}\right) U_{t-1} O_p(1) + O_p\left(\frac{1}{n}\right) U_{t-1} U'_{t-1} O_p(1) \end{aligned} \tag{A.5}$$

Put (A.5) into (A.4), correspondingly the first term in A_{i1t-1} is equal to

$$O_p\left(\frac{1}{\sqrt{n}}\right)\sum_{k=0}^{t-1}\hat{\gamma}_{ik}[\hat{a}_{i0} + (\sum_{j=1}^{q}\hat{a}_{ij}\varepsilon_{it-k-j}] = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{A.6}$$

as $\Sigma_{k=0}^{t-1}\hat{\gamma}_{ik}[\hat{a}_{i0} + (\Sigma_{j=1}^{q}\hat{a}_{ij}E|\varepsilon_{it-k-j}|)] < \infty$. Similarly, we can show that all other terms in A_{i1t-1} are $O_p(\frac{1}{\sqrt{n}})$ and thus so is A_{i1t-1}. Moreover, due to the symmetric distribution of η_t and the functional form of h_{t-1}, for $l = 0,1, \ldots ,L$,

$$E[(\varepsilon'_{t-l}V^{-1}_{t-1}\varepsilon_{t-l} - m)H'_{t-1}A_{i1t-1}] = 0_{1\times m} \tag{A.7}$$

Next we turn to A_{i2t-1}.

$$\begin{aligned} A_{i2t-1} &= h^{\infty}_{it-1}(\varsigma,\delta_1) - h_{it-1}(\varsigma,\hat{\delta}_1) \\ &= \sum_{l=1}^{\infty}\hat{\gamma}_{il+(t-1)}[\hat{a}_{i0} + (\sum_{j=1}^{q}\hat{a}_{ij}\varepsilon^2_{i1-l-j})] \\ &= O_p(\rho^t) \end{aligned} \tag{A.8}$$

Finally, we turn to A_{i3t-1}. Denote $\hat{h}_{it-1}^{\infty} = h_{it-1}^{\infty}(\varsigma, \hat{\delta}_1)$.

$$\begin{aligned}
A_{i3t-1} &= \hat{h}_{it-1}^{\infty} - h_{it-1} \\
&= (\hat{a}_{i0} - a_{i0}) + \sum_{j=1}^{q}(\hat{a}_{ij} - a_{ij})\varepsilon_{it-j}^2 \\
&\quad + \sum_{k=1}^{p}(\hat{b}_{ik} - b_{ik})h_{it-1-k} + \sum_{k=1}^{p}\hat{b}_{ik}(\hat{h}_{it-1-k}^{\infty} - h_{it-1-k}) \\
&= O_p(\tfrac{1}{\sqrt{n}})r_{t-1} + \sum_{k=1}^{p}\hat{b}_{ik}(\hat{h}_{it-1-k}^{\infty} - h_{it-1-k})
\end{aligned}$$

where $r_{t-1} = 1 + \Sigma_{j=1}^{q} \varepsilon_{it-j}^2 + \Sigma_{k=1}^{p} h_{it-1-k}$. Therefore, as $\Sigma_{k=1}^{\infty} \hat{\gamma}_{ik}E(r_{t-1-k}) < \infty$,

$$A_{i3t-1} = O_p(\frac{1}{\sqrt{n}})\sum_{k=1}^{\infty} \hat{\gamma}_{ik}r_{t-1-k} = O_p(\frac{1}{\sqrt{n}}) \tag{A.9}$$

On the other hand, as $A_{i3t-1-k} = (\hat{h}_{it-1-k}^{\infty} - h_{it-1-k}) = O_p(n^{-1/2})$,

$$\begin{aligned}
\sum_{k=1}^{p} \hat{b}_{ik}(\hat{h}_{it-1-k}^{\infty} - h_{it-1-k}) &= \sum_{k=1}^{p} b_{ik}(\hat{h}_{it-1-k}^{\infty} - h_{it-1-k}) \\
&\quad + \sum_{k=1}^{p} (\hat{b}_{ik} - b_{ik})(\hat{h}_{it-1-k}^{\infty} - h_{it-1-k}) \\
&= \sum_{k=1}^{p} b_{ik}(\hat{h}_{it-1-k}^{\infty} - h_{it-1-k}) + O_p(1/n)
\end{aligned}$$

As a result,

$$\begin{aligned}
A_{i3t-1} &= (\hat{a}_{i0} - a_{i0}) + \sum_{j=1}^{q} (\hat{a}_{ij} - a_{ij})\varepsilon_{it-j}^2 + \sum_{k=1}^{p}(\hat{b}_{ik} - b_{ik})h_{it-1-k} \\
&\quad + \sum_{k=1}^{p} b_{ik}a_{i3t-1-k} + o_p(\frac{1}{\sqrt{n}})
\end{aligned} \tag{A.10}$$

Let $\vec{\varepsilon}_t = (\varepsilon_{1t}^2, \ldots, \varepsilon_{mt}^2)'$ and $\gamma_l = (\gamma_{1l}, \ldots, \gamma_{ml})'$, where γ_{il} is as defined in (A.2). Tedious algebra shows that:

$$\nabla_{\delta_1} h_{t-1} \equiv (\nabla'_{a_0}h_{t-1}; \nabla'_{a_1}h_{t-1}, \ldots, \nabla'_{a_q}h_{t-1}; \nabla'_{b_1}h_{t-1}, \ldots, \nabla'_{b_p}h_{t-1})'$$

where

$$\nabla_{a_0} h_{t-1} = I_m + \sum_{l=1}^{p} (\nabla_{a_0} h_{t-1-l}) \mathrm{diag}(b_l) = \sum_{l=1}^{t-1} \mathrm{diag}(\gamma_l);$$

$$\begin{aligned}
\nabla_{a_j} h_{t-1} &= \mathrm{diag}(\vec{\varepsilon}_{t-j}) + \sum_{l=1}^{p} (\nabla_{a_j} h_{t-1-l}) \mathrm{diag}(b_l) \\
&= \sum_{l=1}^{t-j-1} \mathrm{diag}(\gamma_l \odot \vec{\varepsilon}_{t-l-j}), j = 1, \ldots, q; \text{ and} \\
\nabla_{b_k} h_{t-1} &= \mathrm{diag}(h_{t-1-k}) + \sum_{l=1}^{p} (\nabla_{b_k} h_{t-1-l}) \mathrm{diag}(b_l) \\
&= \sum_{l=1}^{t-k-1} \mathrm{diag}(\gamma_l \odot h_{t-1-l-k}), k = 1, \ldots, p.
\end{aligned} \tag{A.11}$$

That is, by (A.10) and (A.11),

$$A_{i3t-1} = \nabla'_{\delta 1} h_{t-1} (\hat{\delta}_1 - \delta_1) + O_p(n^{-1/2}) \tag{A.12}$$

All in all, define $A_{1t-1} \equiv (A_{11t-1}, A_{21t-1}, \ldots, A_{m1t-1})'$. By (A.4), (A.6), (A.7), (A.8), and (A.12), we complete the proof.□

Lemma A.2. $\hat{D}_{t-1} - D_{t-1} = O_p(n^{-1/2}) + O_p(\rho^t)$.□

Proof. It suffices to show that for an arbitrary i, $\sqrt{\hat{h}_{it-1}} - \sqrt{h_{it-1}} = O_p(n^{-1/2}) + O_p(\rho^t)$. First note that $1/\sqrt{\hat{h}_{it-1}} = O(1)$.

$$\begin{aligned}
\sqrt{\hat{h}_{it-1}} - \sqrt{h_{it-1}} &= \frac{\hat{h}_{it-1} - h_{it-1}}{\sqrt{\hat{h}_{it-1}} + \sqrt{h_{it-1}}} \\
&= \left(\frac{1}{\sqrt{\hat{h}_{it-1}}} \right) \left(\frac{1}{1 + \sqrt{h_{it-1} \big/ \hat{h}_{it-1}}} \right) \left(\hat{h}_{it-1} - h_{it-1} \right) \\
&= O_p\left(n^{1/2}\right) + O_p\left(\rho^t\right)
\end{aligned}$$

where the last equality follows by Lemma A.1. □

Lemma A.3. $\hat{V}_{t-1} - V_{t-1} = O_p(n^{-1/2}) + O_p(\rho^t)$. □

Proof.

$$\begin{aligned}
\hat{V}_{t-1} - V_{t-1} &= \hat{D}_{t-1}\hat{\Gamma}\hat{D}_{t-1} - D_{t-1}\Gamma D_{t-1} \\
&= (\hat{D}_{t-1}\Gamma\hat{D}_{t-1} - D_{t-1}\Gamma D_{t-1}) + (\hat{D}_{t-1}\hat{\Gamma}\hat{D}_{t-1} - \hat{D}_{t-1}\Gamma\hat{D}_{t-1}) \\
&= D_{t-1}\Gamma(\hat{D}_{t-1} - D_{t-1}) + (\hat{D}_{t-1} - D_{t-1})\Gamma D_{t-1} \\
&\quad + (\hat{D}_{t-1} - D_{t-1})\Gamma(\hat{D}_{t-1} - D_{t-1}) \\
&\quad + D_{t-1}(\hat{\Gamma} - \Gamma)D_{t-1} + (\hat{D}_{t-1} - D_{t-1})(\hat{\Gamma} - \Gamma)D_{t-1} \\
&\quad + D_{t-1}(\hat{\Gamma} - \Gamma)(\hat{D}_{t-1} - D_{t-1}) \\
&\quad + (\hat{D}_{t-1} - D_{t-1})(\hat{\Gamma} - \Gamma)(\hat{D}_{t-1} - D_{t-1}) \\
&= O_p(n^{-1/2}) + O_p(\rho^t)
\end{aligned}$$

by Lemma A.2. □

This page intentionally left blank

PART II: HIGH FREQUENCY VOLATILITY MODELS

This page intentionally left blank

SAMPLING FREQUENCY AND WINDOW LENGTH TRADE-OFFS IN DATA-DRIVEN VOLATILITY ESTIMATION: APPRAISING THE ACCURACY OF ASYMPTOTIC APPROXIMATIONS

Elena Andreou and Eric Ghysels

ABSTRACT

Despite the difference in information sets, we are able to compare the asymptotic distribution of volatility estimators involving data sampled at different frequencies. To do so, we propose extensions of the continuous record asymptotic analysis for rolling sample variance estimators developed by Foster and Nelson (1996, Econometrica, 64, 139–174). We focus on traditional historical volatility filters involving monthly, daily and intradaily observations. Theoretical results are complemented with Monte Carlo simulations in order to assess the validity of the asymptotics for sample sizes and filters encountered in empirical studies.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 155–181

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20006-3

INTRODUCTION

There are many strategies for estimating time-varying conditional variances and covariances of financial market data. Some involve parametric models, the most popular belong either to the ARCH or the SV class of models, while others are purely data-driven. We focus exclusively on the latter type of estimation strategies which typically involve a rolling or block sample approach.

Data-driven volatility estimates can be distinguished by (1) the sampling frequency, (2) the data window length, and (3) the weighting scheme. There are indeed various schemes, some consist of slicing the returns data into non-overlapping blocks (of equal size) while others rely on a moving window. The block-sampling approach estimates a time series process of conditional variances or covariances using exclusively data from a single block. This approach has its roots in various academic papers, such as Merton (1980); Poterba and Summers (1986); French, Schwert, and Stambaugh (1987); Schwert (1989, 1990a,b) and Schwert and Seguin (1990). Another commonly used strategy involves sliding spans of data resulting in rolling regressions. Early contributions include Fama and MacBeth (1973); Officer (1973) and Merton (1980). Asset-pricing applications often involve sampling at a monthly frequency with rolling windows between 5 and 10 years. Recent examples include Campbell, Lettau, Malkiel, and Xu (2001); Chan, Karceski, and Lakonishok (1999) and Fleming, Kirby, and Ostdiek (2001) among others. Practitioners compute daily volatilities using the same schemes applied to a daily and monthly sampling frequency, a prominent example is the *RiskMetrics* volatility measures.

There are no clear rules regarding the choice of data sampling frequency nor the number of lags to include. Some authors use monthly data and take a 60-month filter, following Fama and MacBeth (1973). Others have used daily data and take monthly sums and cross products of (squared) returns, following French et al. (1987). Some authors report estimators involving different data frequencies to check for robustness or emphasize different features. For instance, Schwert (1990b) considers volatility estimates based on yearly averages of monthly squared returns, daily returns squared and averaged across a month and finally 15-min squared returns across a trading day. Likewise, Campbell et al. (2001) consider a 12-month span rolling sample estimator of volatility, daily averages across a month as well as a quarter.

How do various data-driven volatility estimation schemes involving different sampling frequencies and window lengths compare? The main purpose of this paper is to provide some answers to this question. The theoretical work of Foster and Nelson (1996) formally establishes that with sufficiently high-frequency data, one can estimate instantaneous volatility, denoted by σ_t and sometimes also called spot volatility, using rolling and block sampling filters. They provide some powerful results about the estimation of spot volatility and establish the efficiency of different weighting schemes. They also show that for a large class of continuous path stochastic volatility models, the optimal weighting scheme is exponentially declining and provide formulas for the optimal lag (lead) length of various estimation procedures.

We propose extensions of the continuous record asymptotic analysis of Foster and Nelson for data-driven variance estimators. In particular, despite the difference in information sets we are able to compare the asymptotic distribution of estimators involving data sampled at different frequencies. A priori it may seem impossible to compare rolling sample estimators involving data sampled at different frequencies. However, Foster and Nelson impose certain regularity conditions, which enable us to maintain the same asymptotic distributional properties of the mean squared errors of volatility filters involving data sampled at different frequencies. This insight is the key to the results presented in this paper.

It obviously takes Monte Carlo simulations to back up the accuracy of any asymptotic distribution theory in sample sizes encountered in practice. For our comparison of sampling frequencies, this wisdom certainly applies even more so than is usually the case. The continuous record asymptotic arguments provide only an approximation to what we encounter in practice. To validate our asymptotic reasoning, we report an extensive Monte Carlo study.

The paper is organized as follows: In the first section we briefly review the relevant theoretical results from Foster and Nelson, which support the comparison of rolling sample estimators with various sampling frequencies. The second section provides examples of asymptotically equivalent filters. The simulation results are described in the third section. The final section concludes the paper.

1. THEORETICAL COMPARISON OF ESTIMATORS WITH DIFFERENT SAMPLING FREQUENCIES

We consider data-driven estimators for volatility which figure prominently in many asset-pricing applications. For convenience we assume a continuous time process and model the instantaneous returns $r_t \equiv dp_t$, where p_t is the log price, as a stochastic volatility process.[1] In particular, let

$$dp_t = \mu(p_t, \sigma_t)dt + \sigma_t dW_{1t}$$

$$d\sigma_t^2 = \zeta(p_t, \sigma_t)dt + \delta(p_t, \sigma_t)dW_{2t} \tag{1.1}$$

where W_{1t} and W_{2t} are standard Brownian motions (possibly correlated), and the functions $\mu(\cdot,\cdot)$, $\zeta(\cdot,\cdot)$ are continuous, and $\delta(\cdot,\cdot)$ strictly positive.

We propose extensions of the continuous record asymptotic analysis for rolling sample variance estimators of Foster and Nelson (1996) which enable us to compare the asymptotic distribution of estimators for σ_t involving data sampled at different frequencies. The powerful results in Foster and Nelson are driven by a continuous record asymptotic theory, which assumes that a fixed span of data is sampled at ever finer intervals. The basic intuition driving the results is that normalized returns, r_t/σ_t, over short intervals appear like approximately independent and identically distributed (i.i.d.) with zero conditional mean and finite conditional variance and have regular tail behavior which make the application of Central Limit Theorems possible.

Foster and Nelson impose several regularity conditions for the diffusion process appearing in (1.1). These can be considered standard assumptions, and in order to avoid introducing new notation as well as repeating Foster and Nelson, we outline the conditions and the implications which involve the following: (i) The higher order conditional moments are bounded with small changes over small invervals as $h \to 0$, which essentially allows us to apply the central limit theorem locally. This corresponds to Assumption A in Foster and Nelson. (ii) These conditional moments change slowly over time, which implies these hyper-parameters are regular and with uniform convergence, so that they can be estimated and that nondegenerate asymptotic distributions at the natural rate of convergence can be obtained. These correspond to Assumptions B, C, E and F of Foster and Nelson. (iii) The total number of lags and leads used in the rolling regression go to infinity at rate $h^{(-1/2)}$, though the time interval over which the weights are nonzero is shrinking to 0 at rate $h^{(-1/2)}$. Also it is required that the number of residuals assigned nonzero weights be bounded for each h. This accommodates the ARCH(p) process with p growing at rate $h^{(-1/2)}$ as $h \to 0$ and corresponds to Assumption D in Foster and Nelson. It is important to note, however, that the regularity conditions of the Foster and Nelson framework exclude processes with jumps. The significance and role of certain of these assumptions which are critical in our analysis are further discussed below.

We will adopt a notation slightly different from Foster and Nelson, but similar to that used by Drost and Nijman (1993); Drost and Werker (1996) and Andersen and Bollerslev (1998). Namely, let $r_{(m),\,t} \equiv p_t - p_{t-1/m}$ be the discretely observed time series of continuously compounded returns with m observations per day, per month or whichever benchmark applies. Henceforth, we will call $m = 1$ the benchmark frequency, which in the context of our paper will be either daily or monthly. Hence, the unit interval $r_{(1),\,t}$ is assumed to yield the daily or monthly return.[2]

The $r_{(m),\,t}$ process is a discrete step function with the $1/m$ horizon returns determining the step size. Therefore, when calculating monthly volatility estimates, one can either rely on daily data, i.e. use $r_{(22),\,t}$ with approximately 22 trading days per month, such as Merton (1980), French et al. (1987), Schwert (1990a), among others, or else use sliding spans of squared monthly returns $r_{(1),\,t}$ such as in Officer (1973), Merton (1980), Schwert (1989), among others. Similarly, to obtain daily volatility estimates, one can rely on high-frequency financial data, e.g. 5 min data (with $m = 288$ for FX data, as for instance in Andersen and Bollerslev (1998) and Andersen, Bollerslev, and Lange (1999) or $m = 78$ for equity markets as in Chin, Chan, and Karolyi (1991)).

In general, with ever finer sampling intervals, i.e. $m \to \infty$, we approach the continuously compounded returns, or equivalently $\rho_{(\infty),\,t} \equiv r_t$. The process $\{r_{(m),\,t}\}$ is adapted to the filtration $\{r_{(m),\,t}\}$, and conditional expectations and variances will be denoted as $E_{(m),\,t}(\cdot)$ and $Var_{(m),\,t}(\cdot)$, respectively, whereas unconditional moments follow a similar notation, $E_{(m)}(\cdot)$ and $Var_{(m)}(\cdot)$. From (1.1) we obtain the discrete time dynamics:

$$r_{(m),\,t} = \mu_{(m),\,t} m^{-1} + M_{(m),\,t} - M_{(m),\,t-1/m} \equiv \mu_{(m),\,t} m^{-1} + \Delta_{(m)} M_{(m),\,t}$$

which is the so-called Doob–Meyer decomposition of the $1/m$ horizon returns into a predictable component $\mu_{(m),\,t}$ and a local martingale difference sequence. Consequently,

$$Var_{(m),\,t}\left(r_{(m),\,t}\right) \equiv E\left[\left(\Delta_{(m)} M_{(m),\,t} - \mu_{(m),\,t}\right)^2 | F_{(m),\,t}\right] = \sigma^2_{(m),\,t} m^{-1}$$

where $\sigma^2_{(m),\,t}$ measures the conditional variance per unit of time. Various data-driven estimators for $\sigma^2_{(m),\,t}$ can generically be written as:

$$\hat{\sigma}^2_{(m),\,t} = \sum_{\tau} w_{(\tau-t)} \left(r_{(m),\,t} - \hat{\mu}_{(m),\,t}\right)^2 \tag{1.2}$$

where $w_{(\tau-t)}$ is a weighting scheme and $\hat{\mu}_{(m),\,t}$ a (rolling sample) estimate of the drift. To facilitate the discussion, we restrict our analysis to flat weighting schemes involving $n_L m^{-1/2}$ lags and $n_R m^{-1/2}$ leads. When $n_R = 0$, the filter is one-sided and backward-looking, a case of most practical interest. Note also that $m = 1$ implies that n_L is simply the number of days, or months, of squared returns used to compute the conditional volatility (again assuming $n_R = 0$).[3]

These schemes include the most commonly used volatility estimators involving equally weighting observations throughout the day, across all days of a month, or a sliding span of daily or monthly returns. The asymptotic efficiency of $\hat{\sigma}^2_{(m),\,t}$ only depends on the process characteristics once the filter weights are fixed, in this particular case once n_L and n_R are determined.[4]

Theorem 2 of Foster and Nelson establishes that $m^{1/4}\left(\hat{\sigma}^2_{(m),\,t} - \sigma^2_t\right) \rightarrow N\left(0, C_{(m),\,t}\right)$ as $m \rightarrow \infty$, where the continuous record asymptotic variance for a flat weighting scheme equals

$$C^F_{(m),\,t} \equiv \frac{\theta_{(m),\,t}}{n_R + n_L} + \sqrt{\theta_{(m),\,t}\Lambda_{(m),t}}\;\rho_{(m),\,t}\frac{n_R - n_L}{n_R + n_L} + \Lambda_{(m),\,t}\frac{n_R^3 + n_L^3}{3(n_R + n_L)^2} \quad (1.3)$$

The superscript F in (1.3) refers to the flat weighting scheme. Besides the window length parameters n_L and n_R, three other elements determine the asymptotic efficiency of the filter $\hat{\sigma}^2_{(m),\,t}$ They are $\Lambda_{(m),\,t}$, $\theta_{(m),\,t}$ and $\rho_{(m),\,t}$, each depending on the sampling frequency m and the characteristics of the underlying process (1.1). The process $\Lambda_{(m),\,t}$ represents the "variance of the variance," and therefore any increase of its value will increase $C^F_{(m),\,t}$ and deteriorate the asymptotic efficiency of filtering. The process $\theta_{(m),\,t}$ represents the conditional fourth moment. When the data span increases, namely when $n_R + n_L$ increases, then the first term on the right-hand side of Eq. (1.3) decreases, as the usual asymptotics would predict. Note, however, that the third term in the same equation increases with wider data spans, a result driven by the fact that only local cuts of the data exhibit a relatively stable variance. Finally, the process $\rho_{(m),\,t}$ measures the correlation between the empirical second moment and the conditional variance. As Foster and Nelson observe, the correlation is unity for ARCH-type processes and zero for continuous path diffusions. To streamline the discussion, we do not provide explicit characterizations of the three processes since details appear in Foster and Nelson.

We are interested in comparing $C^F_{(1),\,t}$, which is based on the benchmark sampling frequency, with $C^F_{(m),\,t}$ for $m \neq 1$, or equivalently compare the asymptotic efficiency of volatility estimators involving data sampled at different frequencies. The comparison can be demonstrated by the following illustrative example: Suppose one starts with a 30-day historical volatility estimate and instead of sampling daily one considers half-daily returns and therefore has twice as many returns. Now we ask the following hypothetical question: How many lags of half-daily returns does it take to attain the *same* efficiency as the historical 30-day filter? Please note that we do not change the weighting scheme. We only sample twice more often and try to find out how many lags of half-daily returns attain the same efficiency as a 30-day filter using daily returns. The answer is not 15 days worth of lagged returns sampled twice daily, i.e. the same number (i.e. 30) of observations. Indeed, to maintain the same (asymptotic) efficiency we obviously do not have a simple linear trade-off between sampling frequency and number of observations. It takes in fact more than 15 days of data to maintain the same efficiency. Hence, we need more and more lags as the sampling becomes finer to maintain a particular level of efficiency. It is important to note here that we only try to maintain a certain level of efficiency, and therefore sidestep the issue whether the efficiency one attains is adequate. The result in (1.3) clearly shows, one may have a very large or small asymptotic Mean Square Error (MSE) depending on the magnitude of the conditional higher moment terms.

The asymptotic efficiency of $\hat{\sigma}^2_{(m),\,t}$ only depends on the process characteristics once the filter weights are fixed, in this particular case once n_L and n_R are determined. Therefore, to compare $C^F_{(1),\,t}$ with $C^F_{(m),\,t}$ we must be able to appraise the difference between $\theta_{(m),\,t}$ and $\theta_{(1),\,t}$, and also compare $\Lambda_{(1),\,t}$, $\rho_{(1),\,t}$ with $\Lambda_{(m),\,t}$ and $\rho_{(m),\,t}$.

We expect $\Lambda_{(m),\,t}$ and $\theta_{(m),\,t}$ not only to vary through time, but also to exhibit different (dynamic) properties across sampling frequencies m.[5] In general, it is difficult to obtain an explicit formula for $\Lambda_{(m),\,t}$ and $\theta_{(m),\,t}$, for a given m. Therefore, since it is difficult to characterize the dynamics of the conditional kurtosis and the variance of variance, it is certainly also difficult to compare the processes across various sampling frequencies m. There are some results available, but they do not really help us. For example, Drost and Nijman (1993), Drost and Werker (1996) and Meddahi and Renault (2004) provide aggregation formulas for the *unconditional* kurtosis of GARCH(1,1) and GARCH diffusion processes. Unfortunately, we need to know the *conditional* kurtosis and variance of variance. Meddahi (2002) derives formulas for the first four conditional moments using general class of eigenfunction stochastic volatility models (which includes the GARCH diffusion). The conditional moments involve information sets of past continuous time observations. Computing conditional moments involving the continuous record of past observations is not so appealing, as it would amount to comparing conditional moments with the same past information, defying the purpose of computing $\Lambda_{(m),\,t}$ and $\theta_{(m),\,t}$ across various m.

Since computing $\Lambda_{(m),\,t}$ and $\theta_{(m),\,t}$ across m is impractical, we will rely on a critical assumption in Foster and Nelson, namely their Assumption B, which states that

$$Sup_{t\leqslant\tau\leqslant t+1/\sqrt{m}}\left|\theta_{(m),\tau}-\theta_{(m),t}\right|=o_p(1)$$
$$Sup_{t\leqslant\tau\leqslant t+1/\sqrt{m}}\left|\Lambda_{(m),\tau}-\Lambda_{(m),t}\right|=o_p(1)$$
$$Sup_{t\leqslant\tau\leqslant t+1/\sqrt{m}}\left|\rho_{(m),\tau}-\rho_{(m),t}\right|=o_p(1)$$

This assumption implies that the conditional fourth moments, variance of variance and correlation roughly stay constant over small intervals, where small is interpreted as an interval of size $1/\sqrt{m}$. This assumption guarantees that the process (1.1) is regular enough with higher moments changing slowly over time. We will replace Assumption B of Foster and Nelson with a slightly different condition, namely:

$$Sup_{t\leqslant\tau\leqslant t+1/\sqrt{m}}\left|\theta_{(m),\tau}-\theta_{(1),t}\right|=o_p(1)$$
$$Sup_{t\leqslant\tau\leqslant t+1/\sqrt{m}}\left|\Lambda_{(m),\tau}-\Lambda_{(1),t}\right|=o_p(1)$$
$$Sup_{t\leqslant\tau\leqslant t+1/\sqrt{m}}\left|\rho_{(m),\tau}-\rho_{(1),t}\right|=o_p(1)$$

This assumption is in the same vein as Assumption B of Foster and Nelson, the main difference is that the filtration is no longer kept constant at $\{F_{(m),\,t}\}$, We require the stronger condition that relative to the benchmark filtration $\{F_{(1),\,t}\}$, we have local stability of the conditional higher moments at other sampling frequencies m as well. In particular, we can write

$$\begin{aligned}&Sup_{t\leq\tau\leq t+1/\sqrt{m}}|\theta_{(m),\,\tau}-\theta_{(1),\,t}|\\&\leq|\theta_{(m),\,t}-\theta_{(1),\,t}|+Sup_{t\leq\tau\leq t+1/\sqrt{m}}|\theta_{(m),\,t}-\theta_{(m),\,t}|\end{aligned}$$

Hence, Assumption B of Foster and Nelson implies our condition, provided $|\theta_{(m),\,t}-\theta_{(1),\,t}|$ is $o_p(1)$. One can interpret this condition as saying that at the daily (monthly) level (and beyond) we have reached stability of all relevant conditional higher moments. To a certain degree this is what underlies the empirical application in Foster and Nelson who consider optimal filtering of daily volatility for the Standard and Poors 500 (S&P 500) index.

Though the regularity condition is fairly mild, it is of course not innocuous, and warrants some further discussion. A first observation worth noting is that the unconditional kurtosis can vary dramatically as one changes the sampling frequency (see Bai, Russell, and Tiao (1999) for recent empirical evidence). However, as noted before, our assumption pertains to *conditional* higher moments after the volatility dynamics are taken into account. Furthermore, for $\rho_{(m),\, t}$ the above assumption is often trivially satisfied when ARCH or SV-type processes are considered, since $\rho_{(m),\, t}$ is constant across m and t.

Foster and Nelson propose estimators for $\theta_{(m),\, t}$ and $\Lambda_{(m),\, t}$ (see formulae (10) and (11) in their paper). Hence, one can inspect the time series of point estimates of conditional higher moments, say $\theta_{(1),\, t}$ and $\theta_{(m),\, t}$ for $m \geq 1$ where t is for instance at the daily frequency. Moreover, there are several papers which document the behavior of conditional higher moments, particularly the kurtosis, as sampling frequencies increase. We consider empirical evidence from both the FX and equity markets. For instance, Baillie and Bollerslev (1989) find that GARCH parameter estimates and tail characteristics for FX daily data carry over to weekly, fortnightly, whereas the degree of leptokurtosis and time-dependent heteroskedasticity is reduced as the length of the sampling interval increases to monthly data. Engle, Ito, and Lin (1990) also examine 4-hourly FX data series and find that although all have excess kurtosis, these do not, however, deviate dramatically from the daily kurtosis levels encountered in the literature. More recently, Bollerslev and Domowitz (1993) and Andersen and Bollerslev (1998) also report FX intradaily results, which confirm the stability of the conditional kurtosis at high sampling frequencies.

Hsieh (1991) presents results pertaining to equity markets, namely he examines 15-minute and daily S&P 500 returns and finds comparable results. Chin, Chan, and Karolyi (1991) also examine S&P 500 returns at 5-minute sampling frequency and report sample kurtosis which do not substantially deviate from the daily sample kurtosis levels. This evidence suggests that at least for daily benchmark frequencies our assumption appears to be reasonable. For data sampled at the monthly frequency, it is well known that the conditional kurtosis increases when one moves from monthly to weekly and daily sampling frequencies. Therefore, when the monthly frequency is the benchmark frequency and we compare filters involving, say daily data, with monthly data filters our comparison may not be very accurate, even on asymptotic grounds, since we ignore the variation of higher moments. It should also be noted that for some models, such as the GARCH diffusion which we will use in the Monte Carlo simulations, we can compute for each m the entries to (1.3).

The advantage of our assumption (and the empirical results that lend support to it) is that we can simply drop all the subscripts in (1.3) and write the expression as:

$$C^F_{(m),\,t} \approx \frac{\theta}{n_R + n_L} + \sqrt{\theta\Lambda\rho}\,\frac{n_R - n_L}{n_R + n_L} + \Lambda\,\frac{n_R^3 + n_L^3}{3(n_R + n_L)^2} \tag{1.4}$$

for $m \neq 1$. It is worth noting that to facilitate their discussion, Foster and Nelson also simply drop all the subscripts in (1.3), see in particular their Eq. (9). This representation allows us to make relatively simple comparisons of asymptotically equivalent sampling schemes involving sampling at different frequencies m.

We are specifically interested in one-sided filters. Consider n_R, n_L, θ, Λ and ρ, fixed, and a one-sided window of length n_L. Furthermore, let there be two sampling schemes m_1 and m_2, with $m_2 < m_1$ (we will set $m_2 = 1$ later as the benchmark frequency). According to the results in Foster and Nelson (see in particular Theorem 4 in Foster and Nelson) the same asymptotic efficiency is achieved with one-sided window of length respectively $n_L m_1^{-1/2}$ and $n_L m_2^{-1/2}$. Since $n_L\sqrt{1/m_2} = n_L\sqrt{1/m_1}\left[\sqrt{1/m_2}/\sqrt{1/m_1}\right]$ we need to rescale the interval by $\left[\sqrt{1/m_2}/\sqrt{1/m_1}\right]$ to obtain the same asymptotic efficiency. For the benchmark frequency we assumed $m_2 = 1$, the rescaling has to be $\left[1/\sqrt{1/m_1}\right]$ or $\sqrt{m_1}$. Hence, a relatively simple formula emerges from the Foster and Nelson asymptotics. In the next section, we provide some examples of numerical calculations while in the remainder of the paper we examine how well this asymptotic approximation holds up in simulated experiments.

2. SOME EXAMPLES OF ASYMPTOTICALLY EQUIVALENT FILTERS

We present asymptotically MSE-equivalent one-sided volatility filters for different sampling frequencies. First, we consider the case where the benchmark frequency is daily data. Panels A–C of Table 1 report numerical comparisons for both the 24-h foreign exchange and 6.5 hour trading equity markets. The Panel A pertains to 22-day historical volatility filters, Panel B covers the 26-day filter, and Panel C pertains to 52-day filters. These three cases of n_L correspond roughly to the range of lags often encountered in practice. We examine filters which yield the same asymptotic $C^F_{(m),\,t}$ for intraday sample frequencies such as $m = 2$, 24, 13, 78 and 288. These sampling frequencies correspond, respectively, to half-daily sampling, hourly FX, half-hourly equity, 5-min equity and finally 5-min FX markets. We report the number of lags $n_L m^{1/2}$ for $n_L = 22$, 26, and 52 days and all values of m and translate these lags in number of days of FX and equity market

Table 1. Asymptotically Equivalent One-Sided Equal Weighting Schemes.

Frequency	m	Lags	Days FX	Days Eq.	Frequency	m	Lags	Months
Panel A: Equivalence to 22-day filter					*Panel D*: Equivalence to 60-month filter			
Half-daily	2	32	16	16	Daily	22	282	13
Hourly FX	24	108	4.5	—	Half-daily	44	398	9
Half-hourly equity	13	80	—	6.1	Hourly FX	528	1379	2.6
Five-min FX	288	374	1.3	—	Half-hourly equity	286	1015	3.5
Five-min equity	78	195	—	2.5				
Panel B: Equivalence to 26-day filter					*Panel E*: Equivalence to 12-month filter			
Half-daily	2	37	18.5	18.5	Daily	22	57	2.6
Hourly FX	24	128	5.3	—	Half-daily	44	80	1.8
Half-hourly equity	13	94	—	7.2	Hourly FX	528	276	0.5
Five-min FX	288	442	1.5	—	Half-hourly equity	286	203	0.7
Five-min equity	78	230	—	2.9				
Panel C: Equivalence to 52-day filter					*Panel F*: Equivalence to 24-month filter			
Half-daily	2	74	37	37	Daily	22	34	1.5
Hourly FX	24	255	11	—	Half-daily	44	118	2.7
Half-hourly equity	13	187	—	14	Hourly FX	528	87	0.2
Five-min FX	288	882	3	—	Half-hourly equity	286	407	1.4
Five-min equity	78	459	—	6				

Note: The entries to the table report numerical calculations based on Eq. (1.4) using $C^F_{(1),t}$ evaluated at $n_R = 0$ and $n_L = 22$, $n_L = 26$ and $n_L = 52$ days for the daily filters, and $n_L = 12$, 24 and 60 months for the monthly , as fixed. All asymptotically equivalent $C^F_{(m),t}$ filters require $n_L m^{1/2}$ lags, e.g. 22 $m^{1/2}$.

high-frequency observations. We note from Table 1 that for a 22-day filter we need 374 lags of 5-min observations in FX markets and 195 lags in equity markets to maintain the same asymptotic MSE. This amounts to 1.3 days FX data and 2.5 days stock data. When four daily observations are added, i.e. n_L equals 26, the trade-off gets obviously worse, with, respectively, 442 (FX) and 230 (equity) 5-min lags (or 1.5 and 2.9 days, respectively).

The remaining two panels in Table 1 cover two monthly benchmark frequency cases. These are the 12-, 24- and 60-month filters, again three cases commonly found in the literature. From Panel D we note that a 60-month filter and a 282-lag filter of daily returns are asymptotically equivalent, i.e. it takes 13 months of daily data to attain the same efficiency as a filter with 60 monthly observations. Half-daily sampling yields a 4 month gain, namely 398 lags or 9 months of data are necessary to maintain the same efficiency.

Taking hourly data reduces this further to slightly less than a quarter of observations (i.e. 2.6 months). The final panel shows a similar trade-off. The commonly used annual lag length of monthly returns is equivalent to about the same quarter length of daily data. This is in fact an interesting comparison. For instance, Campbell et al. (2001) use the annual filter with monthly data to extract volatility and decompose it into a market, industry- and firm-specific component and also use a quarterly block sampling scheme to compare the three volatility components with GNP growth rates. The results in Panel E of Table 1 show that these two schemes are roughly asymptotically equivalent (at the monthly and hence also quarterly frequency).

The arguments presented so far can be reversed as well. In Table 2 we report asymptotically equivalent historical window of 1-day and 1-month lengths for spot volatility filters. The entries to Table 2 report numerical calculations based on Eq. (1.4) using $n_L = \sqrt{m}$ as the number of daily or monthly observations in a one-sided historical volatility filter which is asymptotically equivalent to a one-day or one-month filter with sampling frequency $1/m$. Panel A covers the daily benchmark frequency whereas Panel B covers the monthly benchmark frequency. From Table 2 we learn that an hourly FX filter (i.e. $m = 24$) is equivalent to a 5-day historical volatility filter, while half-hourly equity market data filter with $m = 13$ is

Table 2. Equivalent Historical and Benchmark Frequency Volatility Filters.

Panel A: Daily Benchmark Frequency			*Panel B*: Monthly Benchmark Frequency		
Frequency	m	Equivalent number of days	Frequency	m	Equivalent number of months
Hourly FX	24	5	Daily	22	5
Half-hourly equity	13	4	Hourly FX	528	23
Five-min FX	288	17	Half-hourly equity	268	17
Five-min equity	78	9	Five-min equity	1716	41
One-min FX	1440	38	Five-min FX	6336	80
One-min equity	390	20			

Note: The entries to the table report numerical calculation based on Eq. (1.4) using $n_L = \sqrt{m}$ as the number of daily observations in a one-sided historical volatility filter which is asymptotically equivalent to a 1-day (Panel A) or 1-month (Panel B) volatility filter with sampling frequency $1/m$. All results are rounded off to the next integer.

worth one day less. The more interesting case of a 5-min FX market filter (i.e. $m = 288$) is asymptotically equivalent to a 17-day historical volatility filter. The equity market filter with the same frequency, is as efficient as a 9-day filter. We know from the figures reported in Table 1 that we only gain the efficiency of the usual historical volatility once we sample at 1-min interval. Indeed, as the results in Table 2 indicate a 1-min FX filter with $m = 1440$ is as efficient as a 38-day historical volatility estimate and a 1-min stock market filter is equal to a 20-day one.

Panel B of Table 2 deals with one month worth of data and reports comparable monthly frequency historical filters. For example, the first entry in Panel B shows that a 22-day historical volatility filter is equivalent to a 5-month filter. One month of hourly FX data corresponds to a filter of monthly data almost two years long. The most extreme case reported is that of 5-min FX data when sampled for an entire month correspond to an 80-month historical volatility filter.

Obviously, at this point we do not know whether the theoretical asymptotic trade-offs described in Tables 1 and 2 are a good approximation of what we encounter in practice, hence the need to conduct a thorough Monte Carlo investigation to which we return in Section 3. Before turning our attention to simulations, it is worth noting that the arguments presented in this section easily extend to weighting schemes other than the flat scheme discussed so far. For instance, Theorem 6 of Foster and Nelson covers the so-called dominating flat weights, which have the same sliding span of data as the flat scheme but where the actual weights are reshaped (see Formula (17) in Foster and Nelson). The resulting $C^{D}_{(m),\,t}$, where D stands for the dominating flat scheme, is again a function of $\rho_{(m),\,t}$, $\theta_{(m),\,t}$ and $\Lambda_{(m),\,t}$. Hence, under the same regularity conditions we can compare dominating flat weighting schemes for different m on the basis of n_L and n_R. The optimal exponentially declining weighting scheme considered in Theorem 5 of Foster and Nelson is slightly more complicated as it involves, at least in principle infinite weighting schemes. It is noted though that in practice such weights need to be truncated (otherwise they would also violate Assumption D of Foster and Nelson).

3. MONTE CARLO STUDY

The objective of the Monte Carlo study is to examine whether the predictions of the continuous record asymptotic theory describe adequately the sampling behavior of filters when applied to actual data. Therefore, we aim for a design tailored to (1) applications routinely found, and (2) predictions derived from continuous record asymptotics.

3.1. Monte Carlo Design

In this section, we provide the details about filters and sample sizes used in the Monte Carlo design. The models used for the simulation study are representative of the FX and equity financial markets, popular candidates of which are taken to be returns on the YN/US$ exchange rate and S&P 500 stock index. We consider the following continuous time stochastic volatility model which is based on the results of Drost and Nijman (1993) and Drost and Werker (1996):

$$\begin{aligned} d\ln Y_t &= \sigma_t dW_{pt} \\ d\sigma_t^2 &= \theta(w - \sigma_t)dt + (2\lambda\theta)^{1/2} dW_{\sigma t}. \end{aligned} \tag{3.1}$$

The so-called GARCH diffusion yields exact GARCH(1,1) discretizations which are represented by the following equations:

$$\begin{aligned} \ln Y_t - \ln Y_{t-1/m} &= r_{(m),t} = \sigma_{(m)} z_{(m),t} \\ \sigma^2_{(m),t} &= \phi_{(m)} + \alpha_{(m)} r^2_{(m),t-1/m} + \beta_{(m)} \sigma^2_{(m),t-1/m} \end{aligned} \tag{3.2}$$

where $r_{(m),\,t}$ is the returns process sampled at frequency $1/m$. The diffusion parameters of (3.1) and the GARCH parameters of (3.2) are related via formulas appearing in Drost and Werker (1996, Corollary 3.2). Likewise, Drost and Nijman (1993) derive the mappings between GARCH parameters corresponding to processes with $r_{(m),\,t}$ sampled with different values of m. This allows us to estimate a GARCH process using, say daily data with $m = 1$, and computing the GARCH parameters $\alpha_{(m)}$, $\beta_{(m)}$, $\phi_{(m)}$, for any other frequency m as well as the diffusion parameters θ, ω and λ. We consider GARCH processes driven by an error term $z_{(m),\,t}$ that is Normally distributed or exhibits leptokurtosis specified by a Student's t distribution.

The GARCH model parameters are, $\phi_{(m)}$, $\alpha_{(m)}$ and $\beta_{(m)}$, as defined in (3.2). The kurtosis parameter is $\kappa(m)$. The unconditional variance is $v(m) = \phi_{(m)} \big/ (1 - \alpha_{(m)} - \beta_{(m)})$. The parameters used in the simulation experiments appear in Table 3. The daily FX parameter estimates are taken from Andersen and Bollerslev (1998). For the period 01/10/87–30/09/92, the YN/US$ estimates are $\theta = 0.054$, $\omega = 0.476$ and $\lambda = 0.480$. Based on these, the implied GARCH(1,1) parameters $\alpha_{(m)}$, $\beta_{(m)}$ and $\phi_{(m)}$ are disaggregated for the

5-min frequency and reported in Table 3. Following the same paradigm we consider an analogous example for the equity market with only 6.5 h of trading as opposed to the 24-h trading in FX markets. The S&P 500 GARCH(1,1) parameters estimates cover the daily samples 04/01/86–29/08/97 ($T = 2884$ observations). We consider the equivalent intraday frequencies for $m = 78$ to denote the 5-min frequency. The results in the top panel

Table 3. GARCH(1,1) Models Used in Simulation Design.

	Daily frequency	
	YN/US\$ $m = 1$	S&P 500 $m = 1$
Parameters		
$\phi_{(m)}$	0.026	0.033
$\alpha_{(m)}$	0.104	0.029
$\beta_{(m)}$	0.844	0.967
$\kappa(m)$	3	3
$v(m)$	0.500	0.750
5-min frequency		
	$m = 288$	$m = 78$
Parameters		
$\phi_{(m)}$	3×10^{-7}	6×10^{-6}
$\alpha_{(m)}$	0.023	0.004
$\beta_{(m)}$	0.973	0.996
$\kappa(m)$	2.494	2.857
$v(m)$	7×10^{-5}	0.006
Monthly frequency		
	$n = 22$ days	$n = 22$ days
Parameters		
$\phi_{(m)}$	7.602	15.358
$\alpha_{(m)}$	0.059	0.077
$\beta_{(m)}$	0.293	0.836
$\kappa(m)$	3.773	3.416
$v(m)$	11.73	176.6

Note: Entries to the table provide the parameters of the models used in the Monte Carlo simulations. The GARCH model parameters are, $\phi_{(m)}$, $\alpha_{(m)}$ and $\beta_{(m)}$, as defined in (3.2). The kurtosis parameter is $\kappa_{(m)}$. The unconditional variance is $v(m) = \varphi_{(m)}/(1 - \alpha_{(m)} - \beta_{(m)})$. The daily parameters (on the top panel) for the YN/US\$ were obtained from Andersen and Bollerslev (1998) and cover the period 01/10/87–30/09/92. The S&P 500 estimated parameters cover the daily samples 04/01/86–29/08/97 ($T = 2884$ obs.). The disaggregated daily GARCH parameters based on Drost and Nijman (1993) for the 5-minute frequency and aggregated parameters for the monthly frequency are reported.

of Table 3 refer to the daily frequency Normal-GARCH(1,1) estimated parameters for the S&P 500 and the disaggregated models are also reported in the panel below that. In light of the most widely early as well as recent empirical applications of data-driven volatility filters (outlined in the Introduction), we carry this analysis to the monthly frequency. We now aggregate the daily GARCH parameters for the monthly frequency using the approximation of 22 trading days per month (see for instance, French et al., 1987; Schwert, 1989) to obtain the monthly GARCH parameters.

Certain adjustments are required in order to translate some of the continuous record asymptotic results into a meaningful simulation design. There are two issues we need to highlight.

First, it should be noted that when Foster and Nelson discuss spot volatility, they consider a measure of volatility normalized by the sampling interval, i.e. a measure of volatility per unit of time. Second, we need to incorporate this into our simulation design in order to make, for instance, comparisons across sampling frequencies. We have tailored our discussion around a benchmark, or reference, frequency, i.e. $m = 1$. which refers to the daily and monthly frequencies. We will use the reference frequency as a benchmark to measure volatility. This implies that if we sample at, say 5-min intervals and the benchmark frequency is daily we will actually rescale the 5-min volatility estimates by the sampling frequency m, so that they have a daily volatility interpretation. Schwert (1990b) is a practical example of a comparison of daily volatility computed from rescaled intradaily returns (see Schwert, 1990b, Figure D).

3.2. Data-Driven Volatility Filters

In light of the asymptotic properties of various types of data-driven SV measures discussed in Section 1, we will consider the following two families of spot volatility filters. The first family refers to the RV estimators with a given window of 22, 26 and 52 days. To examine the asymptotic MSE approximation of Foster and Nelson adapted here to higher sampling frequencies, we compare the above windows of daily observations with the asymptotically equivalent window of intra-day information. Hence we examine a second family of spot volatility filters starting with the daily benchmark frequency,

(I) *One-day Spot Volatility* $\hat{\sigma}_t^{SV1}$ as the intradaily (rescaled) mean of the log of squared returns $r_{(m),t}$ for different values of m, to produce the end-day spot volatility measure:

$$\hat{\sigma}_t^{SV1} = \sum\nolimits_{j=1}^{m} r^2_{(m),\ t+1-j/m} \qquad t = 1, \ldots, n_{\text{days}}. \tag{3.3}$$

where for the 5-min sampling frequency the lag length take values, $m = 288$ for financial markets open 24-h per day (e.g. FX markets) and $m = 78$ for a stock market open 6.5 hours per day. The filter will be denoted $SV1$ and it is rescaled to yield daily volatility estimates and therefore resembles an integrated volatility measure.

(II) *k-day Spot Volatility*, $\hat{\sigma}_t^{SVk}$, is a one-sided moving k-day average of $\hat{\sigma}_t^{SV1}$. The purpose of selecting k is to make window length comparisons on MSE ground with the daily rolling volatility filters. In our analysis we will set $k = 2$ and 3 days. These windows are specified according to most empirical volatility estimators (see the Introduction) and to examine the theoretical equivalence results in Table 1.

Finally we also consider:

(III) *One-sided Rolling daily window Volatility*, $\hat{\sigma}_t^{RV}$, defined as:

$$\sigma_t^{RV} = \sum_{j=1}^{n_L} w_j \left(r_{(1),t+1-j}\right)^2 \qquad t = 1, \ldots, n_{\text{days}}. \tag{3.4}$$

where n_L is the lag length of the rolling window in days. When the weights w_j are equal to n_L^{-1} then one considers flat weights. Since the asymptotic approximations we propose depend on flat weights we focus on these and consider the three windows of 22, 26 and 52 days.[6]

The above filters are also defined for the monthly benchmark frequency: In the one-month spot volatility, we have 22 trading days, which we extend to 2 and 3 months (or approximately 44 and 66 days). In the rolling monthly volatilities we define windows, n_L, of 12, 24 and 60 months (e.g. Officer, 1973; Merton, 1980; Campbell et al., 2001; Chan et al., 1999).

3.3. Measures of Appraisal

In the Monte Carlo design, we consider the sample sizes n that are representative of the empirical results of 5-year samples for the YN/US$ and S&P 500 (Section 3.1). For the daily volatility simulation design we consider a sample size of six years.[7] We assume that one year has 250 trading days. The intradaily sample of 5minutes for the GARCH processes defined in Table 3 yield sample sizes of n equal to 432,000 and 117,000 observations for the FX and equity examples, respectively. The monthly simulation analysis is based on a sample size of 50 years, often encountered in practice.[8] Each experiment is performed with 1,000 replications.

For the daily benchmark frequency case we simulate $r_{(m),\,t}$ for the 5-min frequency, $m = 288$ and $m = 78$, based on the FX and equity GARCH(1,1) models in Table 3, respectively. Hence, we obtain the highest intraday sampling frequency which we consider to be the true generating process. Next, we apply the GARCH dynamics to obtain the spot volatility, $\sigma_{(m),\,t}$, which for $m = 1$ refers to the daily spot volatility. The extraction error is the difference between the simulated volatility, which is model based, and the modelfree data-driven spot volatility filters:

$$\varepsilon_t^i = \sigma_{(m),t} - \hat{\sigma}_t^i \tag{3.5}$$

for $i = SV1,\ SV2,\ SV3,\ RV22,\ RV26$ and $RV52$ for daily frequency. Obviously, for the rolling window schemes we have different lag lengths. Moreover, for the SV schemes we will consider the 24-h market cases as well as the shorter equity trading opening hours. To avoid further complicating the notation we will denote the monthly spot volatility filters by $SV1,\ SV2,\ SV3,\ RV12,\ RV24$ and $RV60$. It will always be clear from the context that we refer to the monthly filters when SVk will be discussed.

The behavior of the extraction error, ε_t^i, is examined according to the following two dimensions:

1. We examine the efficiency of filters using the MSE which we compare across different filters and sample sizes. Note that we also obtain the Mean Absolute Error (MAE) since Andersen et al.(1999) argue that this criterion is more robust than the RMSE which is susceptible to outliers.[9] The relative efficiency of one filter vis-à-vis another is studied by computing ratios of MSEs. To facilitate comparison, the MSE of all spot volatility filters will be benchmarked against the MSE of the 1-day spot filter. We therefore obtain the following ratios:

$$\mathrm{MSE}^i/\mathrm{MSE}^{SV1} \qquad or \qquad \mathrm{MAE}^i/\mathrm{MAE}^{SV1} \tag{3.6}$$

   where i refers to the MSEs obtained from $\hat{\sigma}_t^i$ for different windows and weights, i.e. for the daily case $i = SVk,\ RV22,\ RV26$ and $RV52$. This analysis extends to the monthly frequency.
2. We study the out-of-sample forecast performance analogous to Andersen, Bollerslev, and Lange (1999). Following Baillie and Bollerslev (1992) the h-period linear projection from the weak GARCH(1,1) model with

returns that span $1/m$ day(s) is expressed as:

$$P_{(m),t}\left(r^2_{(1/h),t+h}\right) = m \cdot h \cdot \sigma^2_{(m)} + \left(\alpha_{(m)} + \beta_{(m)}\right) \cdot \left[1 - \left(\alpha_{(m)} + \beta_{(m)}\right)^{m \cdot h}\right] \cdot \left[1 - \alpha_{(m)} - \beta_{(m)}\right]^{-1} \cdot \left(\sigma^2_{(m),t} - \sigma^2_{(m)}\right) \tag{3.7}$$

where $\sigma^2_{(m)} \equiv \phi_{(m)} \cdot (1 - \alpha_{(m)} - \beta_{(m)})^{-1}$ and $\sigma^2_{(m),\,t}$ would be the alternative volatility filters analyzed above. We shall consider $h = 20$ days as in Andersen, et al. (1999) and obtain the MSE and MAE for each 20-day (or long-run) out-of-sample volatility filter forecast. For the monthly experiment we consider $h = 12$ months. We also consider short horizon forecasts with $h = 1$ day or month for the two benchmark frequencies. It is interesting to note that if we ignore parameter uncertainty, which is the case for our Monte Carlo simulations, we can view (3.7) as a functional transformation of $\sigma^2_{(m),\,t}$, and therefore the asymptotic distribution of the forecast MSE is easily obtained from the asymptotic distribution of the volatility estimator using the usual delta method. This is quite useful as we can easily compare MSEs of forecasts in empirical applications, whereas MSEs of filters can only be computed in a simulation context where the true data generating process is observed. We therefore consider the forecast of MSEs as a bridge between the simulation-based results and empirical studies.

3.4. Simulation Results

We examine whether the simulation results provide supportive evidence for the asymptotics with respect to the MSE equivalence of volatility filters with windows or lag lengths based on different sampling frequencies.

Table 4 reports the Monte Carlo simulation results of the contemporaneous MSE (and MAE) ratios defined in (3.6). For the daily and monthly volatilities the MSE (and MAE) ratios are based on the benchmark of the 1-day and 1-month Spot Volatility ($SV1$), respectively. The theoretical results are based on MSE efficiency and hence we focus our discussion on this criterion, though it should be noted that the MAE results are expected to provide a more robust measure of efficiency comparisons for volatility. We focus on the three high-frequency volatility filters used in practice, $SV1$, $SV2$ and $SV3$ which refer to intraday information for daily volatilities and daily information for monthly volatilities. The window length of these

Table 4. Monte Carlo Simulated MSE and MAE Ratios.

	Daily Frequency, Benchmark: $SV1$				
	$SV2$	$SV3$	$RV22$	$RV26$	$RV52$
S&P 500					
MSE	0.817	0.747	0.627	0.467	0.177
MAE	0.907	0.869	0.782	0.678	0.422
YN/US$					
MSE	0.893	0.812	0.439	0.386	0.735
MAE	0.949	0.908	0.667	0.698	0.853
	Monthly Frequency, Benchmark: $SV1$				
	$SV2$	$SV3$	$RV12$	$RV24$	$RV60$
S&P 500					
MSE	0.752	0.644	0.502	0.370	0.367
MAE	0.870	0.807	0.708	0.617	0.621
YN/US$					
MSE	0.645	0.483	0.556	0.294	0.134
MAE	0.816	0.718	0.818	0.615	0.423

Note: The MSE and MAE are the mean square error and mean absolute error ratios, respectively, defined in (3.6), and obtained from the extraction error (3.5). The daily spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-day spot volatilities, respectively. $RV22$, $RV26$ and $RV52$ are the 22-, 26- and 52-day one-sided rolling volatilities, respectively. The monthly spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-month spot volatilities, respectively. $RV12$, $RV24$ and $RV60$ are the 12-, 24- and 60-month rolling volatilities, respectively. The daily simulation results refer to the 5-year sample of 5-min intradaily GARCH models and the monthly results refer to a 50-year sample size of daily GARCH processes as reported in Section 3.4.

filters is compared on MSE grounds with the rolling volatility estimators RVk where for daily data we focus on $k = 22$, 26 and 52 days and for monthly data we consider $k = 22$, 44 and 66. The choice of these windows is driven from empirical applications and the objective is to report the simulation comparisons in a concise manner.

The theoretical results summarized in Table 1 suggest that the following volatility estimates are MSE asymptotically equivalent:

(i) Consider first the daily frequency. For the equity market, the daily rolling volatilities $RV26$ and $RV22$ are MSE asymptotically equivalent to approximately $SV3$ and $SV2.5$ based on three and two and half days of 5-minute window, respectively (shown in Panels A and B, Table 1). In contrast for the FX market $SV3$ has the same MSE asymptotic efficiency as $RV52$ whereas $SV1.5$ is MSE equivalent to $RV26$ and approximately equivalent to $RV22$ (as shown in Panels A–C, Table 1). For the equity market it is expected that $RV52$ would have the lowest MSE compared with any of the daily SVk filters where k ranges from 1 to 3 days of intraday data (shown in Panel C, Table 1).

(ii) For the monthly benchmark frequency we find that the monthly spot volatility based on a two and a half day window $SV2.5$ is approximately MSE asymptotically equivalent to $RV12$, whereas $RV24$ is MSE equivalent to $SV1.5$ (shown in Panels E and F, Table 1). It is expected that the monthly rolling volatility with longer window length $RV60$ would be less efficient than any of the monthly SVk filters where k ranges from 1 to 3 months of daily data (shown in Panel D, Table 1).

Following the simulation design discussed in Sections 3.1–3.3 for a Normal GARCH process, we obtain three types of MSE and MAE ratios where the benchmark filter is $SV1$: (a) contemporaneous, presented in Table 4 (b) 1-step ahead forecast, presented in Table 5, and (c) h-step ahead forecasts based on longer horizons, $h = 20$ days and 12 months for daily and monthly filters, respectively, shown in Table 6. Moreover, we extend the above results to the case of Student's t GARCH model for generating returns, and the MSEs and MAEs ratios are reported in Table 7. For each case we evaluate the results for both the equity and foreign exchange markets reported by the S&P 500 and YN/US$ simulated series.

A synthesis of the simulation results reported in all of the above Tables yields the following three broad conclusions:

(1) For the equity market represented by the S&P 500 index simulated process given in Table 3 with 6.5 h of trading, the 5-min intraday volatility filter $SV3$ that involves a window length of three days is MSE asymptotically equivalent to a rolling volatility that involves interdaily data with a window of 22 days instead of 26 days. Hence, although the theoretical evaluations in Table 1 (Panels A and B) suggest that $RV26$ and $RV22$ are approximately equivalent to a window of 3 and 2.5 days of 5-min data for the equity market, respectively, the simulation evidence suggests that only $RV22$ attains this MSE equivalence with $SV3$. This conclusion holds whether we compare MSEs from the contemporaneous extraction error or whether that is based in 1-day ahead forecast error (compare for instance the results in Tables 4 and 5). This result

Table 5. Monte Carlo Simulated of 1-day and 1-month ahead MSE and MAE Ratios.

	$h = 1$-day ahead, Daily Frequency, Benchmark: $SV1$				
	$SV2$	$SV3$	$RV22$	$RV26$	$RV52$
S&P 500					
MSE	0.816	0.746	0.627	0.474	0.180
MAE	0.908	0.870	0.782	0.680	0.420
YN/US$					
MSE	0.901	0.829	0.507	0.573	0.905
MAE	0.953	0.916	0.708	0.756	0.938
	$h = 1$-month ahead, Monthly Frequency, Benchmark: $SV1$				
	$SV2$	$SV3$	$RV12$	$RV24$	$RV60$
S&P 500					
MSE	0.749	0.639	0.499	0.353	0.389
MAE	0.868	0.804	0.669	0.602	0.637
YN/US$					
MSE	0.602	0.466	0.509	0.376	0.293
MAE	0.793	0.702	0.777	0.659	0.548

Note: The h-period ahead forecast MSE and MAE are the mean square error and mean absolute error ratios, respectively, defined in (3.6), and obtained from the extraction error in (3.5), using the h-period linear projection GARCH(1,1) equation in (3.7). The daily spot volatilities $SV1$, $SV2$ and $SV3$ are the 1, 2- and 3-day spot volatilities, respectively. $RV22$, $RV26$ and $RV52$ are the 22-, 26- and 52-days one-sided rolling volatilities, respectively. The monthly spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-month spot volatilities, respectively. $RV12$, $RV24$ and $RV60$ are the 12-, 24- and 60-months rolling volatilities, respectively. These daily simulation results refer to the 5-year sample and the 5-minute intraday frequency. Similar results apply to the 10-year sample. The monthly simulation results refer to a 50-year sample size.

also extends to situations whether the driving error process has a Student's t distribution with four degrees of freedom as shown in Table 6. However, for longer horizon forecasts of $h = 20$ we find that the MSE efficiency equivalence does not hold since for such horizons the window does not seem to play any role. All filters for the S&P 500 appear to have the same MSE and MAE. Moreover, as expected from the theory, the simulation results show that for the equity market the $RV52$ yields the lowest MSE compared to the two- and three-day window SVk, $k = 2, 3$.

Table 6. Monte Carlo Simulated 20-day and 12-month ahead MSE and MAE Ratios.

$h = 20$-day ahead, Daily Frequency, Benchmark: $SV1$					
	$SV2$	$SV3$	$RV22$	$RV26$	$RV52$
S&P 500					
MSE	0.987	0.982	1.066	1.040	0.999
MAE	0.999	0.999	1.003	1.002	1.002
YN/US$					
MSE	0.977	0.972	1.472	1.500	1.744
MAE	0.990	0.986	1.216	1.232	1.349
$h = 12$-month ahead, Monthly Frequency, Benchmark: $SV1$					
	$SV2$	$SV3$	$RV12$	$RV24$	$RV60$
S&P 500					
MSE	0.978	0.969	1.011	0.957	0.910
MAE	0.999	0.999	1.000	1.000	1.000
YN/US$					
MSE	0.999	0.999	0.999	0.999	0.999
MAE	1.000	1.000	1.000	1.000	1.000

Note: The h-period ahead forecast MSE and MAE are the mean square error and mean absolute error ratios, respectively, defined in (3.6), and obtained from the extraction error in (3.5), using the h-period linear projection GARCH (1,1) equation in (3.7). The daily spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-day spot volatilities, respectively. $RV26$ and $RV52$ are the 26- and 52-days one-sided rolling volatilities, respectively. The monthly spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-months spot volatilities, respectively. $RV12$, $RV24$ and $RV60$ are the 12-, 24- and 60-months Rolling Volatilities, respectively. These daily simulation results refer to a 5-year sample of 5-min intraday frequency. Similar results apply to the 10-year sample. The monthly simulation results refer to a 50-year sample size.

As a final note we mention that the MAEs are often higher than the MSEs and they provide additional support to the above results.

(2) For the FX market as generated by the YN/US$ defined in Table 3 we find that the 5-min 24-h market yields a daily spot volatility filter $SV3$, which is MSE equivalent to $RV52$ as the theoretical results suggest in Table 1 (Panel C). This result is valid whether the extraction error is defined contemporaneously or from a short-horizon prediction of 1-day ahead. Moreover, it holds whether we generate a GARCH process with

Table 7. Monte Carlo Simulated MSE and MAE Ratios under Student's $t(v)$.

Daily Frequency, Intradaily GARCH with $t(0,1;\ v=4)$ Benchmark: $SV1$										
	SV2	SV3	RV22	RV26	RV52	SV2	SV3	RV22	RV26	RV52
	S&P 500					YN/US$				
	$h=0$									
MSE	0.817	0.747	0.636	0.476	0.179	0.894	0.813	0.815	0.420	0.691
MAE	0.908	0.869	0.785	0.683	0.424	0.949	0.909	0.908	0.656	0.841
	$h=1$									
MSE	0.819	0.749	0.636	0.465	0.175	0.902	0.830	0.829	0.505	0.844
MAE	0.908	0.871	0.784	0.677	0.417	0.952	0.916	0.915	0.706	0.936
	$h=20$									
MSE	0.987	0.982	1.062	1.044	0.999	0.977	0.972	0.829	0.496	0.851
MAE	0.999	1.000	1.000	1.002	1.002	0.991	0.989	0.836	0.702	0.924
Monthly Frequency, Daily GARCH with $t(0,1;\ v=6)$ Benchmark: $SV1$										
	SV2	SV3	RV12	RV24	RV60	SV2	SV3	RV12	RV24	RV60
	S&P 500					YN/US$				
	$h=0$									
MSE	0.752	0.644	0.502	0.370	0.367	0.645	0.483	0.553	0.292	0.133
MAE	0.870	0.807	0.709	0.617	0.622	0.816	0.718	0.817	0.614	0.423
	$h=1$									
MSE	0.749	0.639	0.448	0.353	0.391	0.603	0.467	0.507	0.374	0.292
MAE	0.869	0.804	0.669	0.602	0.638	0.793	0.702	0.778	0.659	0.548
	$h=12$									
MSE	0.978	0.969	1.010	0.957	0.910	0.999	0.999	1.000	0.999	0.990
MAE	0.999	1.000	0.999	0.999	0.999	1.000	1.000	1.000	1.000	1.000

Note: The MSE and MAE are the mean square error and mean absolute error ratios, respectively, defined in (3.6), and obtained from the extraction error (3.5). The simulated process refers to a GARCH(1,1) with conditional Student's t distribution and degrees of freedom v, where v is equal to 4 for the intradaily GARCH and 6 for the daily GARCH. The parameters of the GARCH(1,1) model for the five-minute S&P 500 are defined in Table 3. The daily spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-day spot volatilities, respectively, as well as $RV26$ and $RV52$ which are the 26 and 52-days one-sided rolling volatilities, respectively. The monthly spot volatilities $SV1$, $SV2$ and $SV3$ are the 1-, 2- and 3-months spot volatilities, respectively. $RV12$, $RV24$ and $RV60$ are the 12-, 24- and 60-months rolling volatilities, respectively. The intradaily and daily samples are equal to 5 and 30 years, respectively.

Student's t innovations and for longer samples of 10 years of intradaily data. It is also important to note that although the theoretical results show that the rolling volatilities with shorter windows of 22- and 26- days are MSE equivalent to SV with 1.3 and 1.5 days, we find no supportive simulation evidence if we compare them with $SV2$.

(3) The monthly volatility simulation results show support for the MSE equivalence of the rolling volatility with 12 months, $RV12$, and the spot volatility with 3 days, $SV3$ (Panel E). However, there is no supportive simulation evidence that $RV24$ is MSE asymptotically efficient to either $SV1$ or $SV2$. This result is valid for contemporaneous extraction errors (Table 4) and short horizon forecasts of $h = 1$ month (Table 5) as well as Student's t error driven GARCH processes. However, for longer horizon forecasts of $h = 12$, there seems to be no gain in efficiency or MSE equivalence between volatility filters with different windows.

Overall the simple prediction in the paper that extends the Foster and Nelson (1996) MSE efficiency results to the comparison of volatility filters with flat weights for alternative window lengths of different sampling frequencies gains simulation support within the boundaries of the design described above for certain cases: For the contemporaneous and 1 step ahead forecast error comparisons of the MSE equivalence of (a) the 3-day spot volatility $SV3$ in the equity and FX markets with daily rolling volatilities, $RV22$ and $RV26$, respectively (b) the 3-month spot volatility $SV3$ with the 12-month rolling volatility, $RV12$.

4. CONCLUSIONS

The paper extends one of the theoretical results in Foster and Nelson (1996), namely equally efficient spot volatility estimators are derived for alternative sampling frequencies based on the continuous record asymptotics. The theoretical results are examined by a Monte Carlo study that provides support for comparing volatility filters with alternative window lengths and sampling frequencies on MSE grounds. Our analysis can also be extended to study optimal weighting schemes.

NOTES

1. As noted by Foster and Nelson, their analysis applies not only to SV diffusions but also, with appropriate modifications, to discrete time SV, to ARCH models and to certain types of random coefficient models. While we start with a continuous time

SV framework, we will focus later on a particular case which yields a GARCH(1,1) model using exact discretization methods (see also Drost & Werker, 1996, Meddahi & Renault, 2004 and Andersen & Bollerslev, 1998).

2. The notion of a benchmark frequency will be used extensively, particularly when simulating models. We use the daily and monthly examples as they are most commonly encountered in applications. The benchmark frequency will also serve as a reference frequency to normalize simulation results and make them comparable across frequencies as will be discussed later.

3. This includes the zero lag, since we assume end-of-day, or end-of-month, volatilities.

4. Foster and Nelson assume that the $\mathrm{Sup}_\tau\{\tau : w_{\tau-t}>0\} - \mathrm{Inf}_\tau\{\tau : w_{\tau-t}>0\} = O(m^{-1/2})$ and hence shrinks as m increases and is bounded in probability by $m^{-1/2}$ (see Foster and Nelson (1996, Assumption D)). For flat scheme involving $n_{\mathrm{L}}m^{-1/2}$ lags and $n_{\mathrm{R}}m^{-1/2}$ leads, the weights can be characterized as $w_{(m),t} = m^{-1/2}(n_L + n_R)^{-1}I\{\tau\varepsilon[-n_L m^{-1/2}, n_R m^{-1/2}]\}$.

5. We focus here on $\Lambda_{(m),t}$ and $\theta_{(m),t}$, since $\rho_{(m),t}$ is unity for ARCH-type processes and zero for continuous path diffusions. Therefore, $\rho_{(m),t}$ is in many cases constant across time and sampling frequencies.

6. Note that geometrically declining weights can also be adopted where $w_j = \exp(-\alpha j)$ with $\alpha = 0.0665$. However, the comparisons we have in Table 3 focus on flat weights.

7. Note that for the one-sided rolling estimates the window length presupposes trimming a percentage of the initial sample. The effective sample is normalized for all filters for evaluation and comparison purposes.

8. With $n_{\text{years}} = 50$ we have $n_{\text{months}} = 600$ and $n_{\text{days}} = 7,500$

9. Andersen et al. also consider analogous statistics which are adjusted for heteroskedasticity and which are found to have significant improvements in the volatility forecasting analysis.

ACKNOWLEDGMENTS

The authors would like to thank the hospitality and financial support of CIRANO, Montréal and the University of Manchester. Material in this paper is an expanded version of Sections 1 and 3, taken from an earlier paper (Andreou and Ghysels, 2000).

REFERENCES

Andersen, T., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*, 885–905.

Andersen, T., Bollerslev, T., & Lange, S. (1999). Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance*, *6*, 457–477.

Andreou, E., & Ghysels, E. (2000). Rolling-sample volatility estimators: Some new theoretical, simulation, and empirical results. CIRANO Discussion Paper 2000s-19.

Bai, X., Russell, J., & Tiao, G. (1999). Beyond merton's utopia: Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data, Discussion Paper, University of Chicago.

Baillie, R. T., & Bollerslev, T. (1989). The message in daily exchange rates: A conditional variance tale. *Journal of Business and Economic Statistics*, *7*, 297–305.

Baillie, R. T., & Bollerslev, T. (1992). Prediction in dynamic models with time dependent conditional variances. *Journal of Econometrics*, *74*, 3–30.

Bollerslev, T., & Domowitz, I. (1993). Trading patterns and prices in the interbank foreign exchange market. *Journal of Finance*, *XLVIII*(4), 1421–1443.

Campbell, J. Y., Lettau, M., Malkiel, B. G., & Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance*, *56*, 1–43.

Chan, L. K. C., Karceski, J., & Lakonishok, J. (1999). On portfolio optimization: Forecasting covariances and choosing the risk model. *Review of Financial Studies*, *12*, 937–974.

Chin, K., Chan, K. C., & Karolyi, G. A. (1991). Intraday volatility in the stock index futures markets. *Review of Financial Studies*, *4*, 637–684.

Drost, F. C., & Nijman, T. (1993). Temporal aggregation of GARCH processes. *Econometrica*, *61*, 909–927.

Drost, F. C., & Werker, B. M. J. (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics*, *74*, 31–57.

Engle, R. F., Ito, T., & Lin, W.-L. (1990). Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica*, *58*, 525–542.

Fama, E., & MacBeth, J. D. (1973). Risk, return, and equilibrium: empirical tests. *Journal of Political Economy*, *81*, 607–636.

Fleming, J., Kirby, C., & Ostdiek, B. (2001). The economic value of volatility timing. *Journal of Finance*, *56*(1), 329–352.

Foster, D., & Nelson, D. (1996). Continuous record asymptotics for rolling sample estimators. *Econometrica*, *64*, 139–174.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, *19*, 3–29.

Hsieh, D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance*, *46*, 1839–1877.

Meddahi, N. (2002). Moments of continuous time stochastic volatility models. CIRANO Discussion Paper.

Meddahi, N., & Renault, E. (2004). Temporal aggregation of stochastic volatility models. *Journal of Econometrics*, *119*, 355–379.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*, 323–361.

Officer, R. R. (1973). The variability of the market factor of the New York stock exchange. *Journal of Business*, *46*, 434–453.

Poterba, J. M., & Summers, L. H. (1986). The persistence of volatility and stock market fluctuations. *American Economic Review*, *76*, 1142–1151.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance*, *44*, 1115–1154.

Schwert, G. W. (1990a). Stock volatility and the crash of '87. *Review of Financial Studies*, *3*, 77–102.

Schwert, G. W. (1990b). Stock market volatility. *Financial Analysts Journal*, *46*, 23–34.

Schwert, G. W., & Seguin, P. J. (1990). Heteroskedasticity in stock returns. *Journal of Finance*, *45*, 1129–1155.

This page intentionally left blank

MODEL-BASED MEASUREMENT OF ACTUAL VOLATILITY IN HIGH-FREQUENCY DATA

Borus Jungbacker and Siem Jan Koopman

ABSTRACT

In this chapter, we aim to measure the actual volatility within a model-based framework using high-frequency data. In the empirical finance literature, it is widely discussed that tick-by-tick prices are subject to market micro-structure effects such as bid-ask bounces and trade information. These market micro-structure effects become more and more apparent as prices or returns are sampled at smaller and smaller time intervals. An increasingly popular measure for the variability of spot prices on a particular day is realised volatility that is typically defined as the sum of squared intra-daily log-returns. Recent theoretical results have shown that realised volatility is a consistent estimator of actual volatility, but when it is subject to micro-structure noise and the sampling frequency increases, the estimator diverges. Parametric and nonparametric methods can be adopted to account for the micro-structure bias. Here, we measure actual volatility using a model that takes account of micro-structure noise together with intra-daily volatility patterns and stochastic volatility. The coefficients of this model are estimated by maximum likelihood methods that are based on importance sampling techniques. It is shown that such Monte Carlo techniques can be employed successfully for our purposes in

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 183–210

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20007-5

a feasible way. As far as we know, this is a first attempt to model the basic components of the mean and variance of high-frequency prices simultaneously. An illustration is given for three months of tick-by-tick transaction prices of the IBM stock traded at the New York Stock Exchange.

1. INTRODUCTION

1.1. Some Background

The filtering of efficient prices and volatilities in financial markets using high-frequency intra-day spot prices has gained much interest from both the professional and academic communities. The Black–Scholes (BS) model is still the dominating framework for the pricing of contingencies such as options and financial derivatives while the generalised autoregressive conditional heteroskedasticity (GARCH) models are widely used for the empirical modelling of volatility in financial markets. Although the BS and GARCH models are popular, they are somewhat limited and do not provide a satisfactory description of all the dynamics in financial markets. In this chapter, we focus on the measurement of daily volatility in financial markets using high-frequency data. A model-based approach is taken that considers both prices and volatilities.

Measuring the volatility in prices of financial assets is essentially not much different than measuring any other unobserved variable in economics and finance. For example, many contributions in the economic literature have appeared on the measurement of the business cycle that can be defined as the unobserved component for medium-term deviations from a long-term trend in economic activity. Nonparametric methods (e.g. the Hodrick–Prescott filter) as well as model-based methods (e.g. the Beveridge–Nelson decomposition) have been proposed and developed for the measurement of business cycles. In the case of measuring volatility using high-frequency data, most, if not all, of the emphasis so far is on nonparametric methods. The properties of nonparametric estimates of volatility are investigated in detail and rely on advanced and novel asymptotic theory in stochastics and econometrics (see Ait-Sahalia, Mykland, & Zhang, 2004; Barndorff-Nielsen, Hansen, Lunde, & Shephard, 2004). In this chapter, we explore model-based approaches for the measurement of volatility. By allowing for intra-day effects and stochastic volatility (SV), efficient estimates of volatility can be obtained. However, the modelling framework required for this purpose is non-standard and does easily move away from linearity and Gaussianity.

1.2. Measuring Actual Volatility

The price of a financial asset is denoted by P_t. A common assumption in the finance literature is that the log of P_t can be represented by a stochastic differential equation (SDE) of the form

$$\mathrm{d}\log P_t = \mu_t(\psi)\,\mathrm{d}t + \sigma_t(\psi)\,\mathrm{d}B_t, \quad t > 0 \tag{1}$$

where $\mu_t(\psi)$ is a drift function representing expected return, $\sigma_t(\psi)$ a stochastic process representing the spot volatility, B_t a standard Brownian motion and ψ is a vector of unknown parameters, see Campbell, Lo, and MacKinlay (1997) for more background. For different purposes the financial economist is interested in measuring and predicting the variability of the asset price. This variability is mainly determined by what is called integrated volatility

$$\sigma^{*2}(0,t) = \int_0^t \sigma_t^2(\psi)\mathrm{d}t \tag{2}$$

where the dependence of ψ is implied. The related concept of actual volatility for the interval $[t_1, t_2]$ is defined as $\sigma^{*2}(t_1, t_2)$ where

$$\sigma^{*2}(t_1,t_2) = \sigma^{*2}(0,t_2) - \sigma^{*2}(0,t_1) \tag{3}$$

It should be noted that integrated and actual variance would be the more precise names for integrated and actual volatility, respectively. However, we choose to follow the convention in much of the financial econometrics literature and refer to these quantities as volatilities.

1.3. Realised Volatility

The realised price of an asset can be observed when a trade takes place. Heavy trading takes place in international financial markets, on a continuously basis. The Trades and Quotes (TAQ) database of the New York Stock Exchange (NYSE) contains all equity transactions reported on the so-called consolidated tape and it includes transactions from the well-known NYSE, AMEX and NASDAQ markets but also from various other important exchange markets. By collecting all prices in a certain period, a so-called high-frequency dataset is obtained. We refer to high-frequency data when observations are sampled at very small time intervals. In the finance literature, this usually means that observations are taken at the intra-daily interval of 5 or 1 minutes (calendar time sampling) or that observations are recorded trade-by-trade (business time sampling). The trade-by-trade data are regarded as the ultimate high-frequency collection of prices. In a time scale of seconds, we may even have multiple trades within the same time interval, although, this is unlikely. It is however more likely that many prices will be missing since trades do not take place every second in most financial markets with the possible exception of foreign exchange markets.

The observed log price at the discrete time point t_n (in seconds) is denoted by $Y_n = \log P_{t_n}$ for observation index n. The number of time points (seconds) in one trading day is denoted by N_d. We therefore have potentially N_d observations $Y_1, Y_2, \ldots, Y_{N_d}$ of the log price of a trade on a particular day d. The index t_0 refers to the start of the period while in this chapter the distance $t_n - t_{n-1}$ is assumed constant for $n = 1, \ldots, N_d$. The value Y_n will not be available when no trade has taken place at time t_n. Such values will be treated as missing. The number of trades is denoted by $N \leq N_d$ so that we have $N_d - N$ missing values in day d.

A natural estimator of actual volatility is given by the so-called realised volatility and denoted by $\tilde{\sigma}^{*2}(t_0, t_{N_d})$. Realised volatility can be computed by

$$\tilde{\sigma}^{*2}(t_0, t_{N_d}) = \sum_{j=2}^{N_d/m} (Y_{mj} - Y_{mj-m})^2 \tag{4}$$

where m is the sampling frequency (see Andersen, Bollerslev, Diebold, & Labys, 2001). For example, when the sampling frequency is 5 min, m equals 300 assuming that the index of Y_n refers to the nth second. In the case a transaction has not taken place at time n, so that Y_n is missing in (4), it can be approximated via an interpolation method using observed values in the neighbourhood of Y_n (see Malliavin & Mancino, 2002 and Hansen & Lunde, 2004 for discussions of different filtering methods). Novel asymptotic theory is developed for the realised volatility estimator (4) as the number of observations in the fixed interval $[t_0, t_n]$ increases (or as m decreases, see Barndorff-Nielsen & Shephard, 2001). Specifically, it is shown that $\tilde{\sigma}^{*2}(t_0, t_{N_d})$ is a consistent estimator of actual volatility. This result suggests that if we sample the log price process $\log P_t$ more frequently within a fixed time interval, by taking m small, the efficiency of the estimator increases. Empirical work on this subject however indicates the complete opposite (see, in particular, Andreou & Ghysels, 2001 and Bai, Russell, &

Tiao, 2000). If the realised volatility is computed using more observations, the estimate seems to diverge. A possible cause of this phenomenon is the fact that the efficient price is not observed directly. Observed trading prices Y_n are contaminated by the so-called micro-structure noise that has various causes such as bid-ask bounces, discrete price observations and irregular trading (see Campbell et al., 1997 for a further discussion with references). Micro-structure noise can generate significant dependence in first and higher-order moments of spot prices. This dependence typically vanishes with aggregation.

It is therefore argued that the micro-structure noise ruins the reliability of realised volatily as an estimator. Recently non-parametric methods have been proposed by Barndorff-Nielsen et al. (2004) and Ait-Sahalia et al. (2004) that produce consistent estimates of volatility even in the presence of micro-structure noise.

1.4. Plan of this Chapter

In this chapter, we take a model-based approach to measure volatility using high-frequency prices that are observed with micro-structure noise. Standard formulations for price and volatility from the finance literature will be considered. Further, the model allows for an intra-daily volatility pattern and SV. The details of the model are described in Section 2. In this way, the salient features of high-frequency prices are described and efficient estimates of actual volatility can be produced. However, the estimation of parameters in this class of models is nonstandard and simulation-based methods need to be employed. This also applies to methods for the measurement of volatility. We propose importance sampling techniques in Section 3 and it is shown that such methods can work effectively and efficiently for our purposes. This is illustrated in Section 4 in which daily volatilities are measured from high-frequency IBM prices recorded at the NYSE for a period of three months. The detailed results show that the implemented methods work satisfactory for estimation and measurement. A short discussion in Section 5 concludes this chapter.

2. MODELS FOR HIGH-FREQUENCY PRICES

2.1. Model for Price with Micro-Structure Noise

Different specifications for the drift and diffusion components of model (1) have been proposed in the finance literature. Throughout this chapter we assume that the drift term equals zero so that var$(P_{t+\tau}|P_t)$, for $\tau > 0$, only depends on the diffusion term $\sigma_t(\psi)$ in (1) (see Andersen, Bollerslev, & Diebold, 2002). For the volatility process $\sigma_t(\psi)$ we consider three different specifications in this section. The first and most basic specification is where the volatility is kept constant over time, that is $\sigma_t(\psi)\sigma_t(\psi)$. These assumptions lead us to the following model for the efficient price process

$$\mathrm{dlog}P_t = \sigma(\psi)\mathrm{d}B_t \tag{5}$$

where B_t is standard Brownian motion. Other specifications for $\sigma_t(\psi)$ are discussed in Sections 2.2 and 2.3.

It is assumed that the observed trade price Y_n is a noisy observation of the efficient price. In other words, the price is possibly contaminated by micro-structure noise. We therefore have $Y_n = \log P_t + U_n$ for $t = t_n$ where U_n represents microstructure noise that is assumed to have zero mean and variance σ_U^2. The noise process U_n can be subject to serial correlation, although initially we assume an independent sequence for U_n, see also the discussion in Section 3.1. The discrete time model then becomes

$$Y_n = p_n + \sigma_U U_n, \quad U_n \sim \mathrm{IID}(0,1) \tag{6}$$

$$p_{n+1} = p_n + \sigma_\varepsilon \varepsilon_n, \quad \varepsilon_n \sim \mathrm{NID}(0,1) \tag{7}$$

where $p_n = \log P_t$ is the unobserved price (in logs) at time $t = t_n$ for $n = 1, \ldots, N_d$. Here n refers to an index of seconds leading to equidistances $t_n - t_{n-1}$

In this framework we have a simple expression for actual volatility

$$\sigma^{*2}(t_n, t_{n+1}) = (t_{n+1} - t_n)\sigma_\varepsilon^2$$

The model implies that the observed return

$$R_n = \Delta Y_{n+1} = \Delta p_{n+1} + \sigma_U \Delta U_{n+1} = \sigma_\varepsilon \varepsilon_n + \sigma_U U_{n+1} - \sigma_U U_n$$

follows a moving average (MA) process of order one, that is $R_n \sim \mathrm{MA}(1)$ (see Harvey, 1989 for a further discussion of the local level model). The direct consequence of this model is that the returns are not white noise. This is not an indication that prices are realised at inefficient markets. The serial correlation is caused by the high-frequency of the realisations and is due to micro-structure bounces and related effects.

The initial assumption of constant volatility is too strong for a relatively long period, even for, say, one day. However, this simple framework allows us to obtain a preliminary estimate of daily volatility using high-frequency data. The estimation of the local level model (7) is explored in detail by Durbin and Koopman (2001, Chapter 1) and is based on the standard

Kalman filter equations. The possibly many missing values in the series Y_n can be accounted for within the Kalman filter straightforwardly. When it is assumed that micro-structure noise U_n and price innovation ε_n are Gaussian distributed error terms, exact maximum likelihood (ML) estimates of σ^2_U and σ^2_ε are obtained by numerically maximising the Gaussian likelihood function that can be evaluated via the Kalman filter. When the Gaussian assumptions do not apply, these estimates can be referred to as quasi-maximum likelihood (QML) estimates.

Ait-Sahalia et al. (2004) also consider the local level model framework to describe the true process of the observed log prices and also observe that the returns therefore follow an MA (1) process. In their theoretical analysis, it is argued that distributional properties of U_n do not matter asymptotically. The main conclusions of their analysis are that (i) "*modelling the noise explicitly restores the first order statistical effect that sampling as often as possible is optimal*" and (ii) "*this remains the case if one misspecifies the assumed distribution of the noise term*". We take these findings as an endorsement of our modelling approach. They further discuss possible extensions of the local level model by modelling U_n as a stationary autoregressive process and by allowing for contemporaneous correlation between U_n and ε_n. In our modelling framework, the former extension can be incorporated straightforwardly although the estimation of elaborate autoregressive moving average (ARMA) processes for financial data may be hard in practice. The latter proposed extension is more difficult from an inference point of view since the correlation coefficient between U_n and ε_n is not identified when both variances are unrestricted (see the discussions in Harvey & Koopman, 2000).

2.2. Intra-Daily Seasonal Patterns in Volatility

In empirical work, it is often found that estimates of actual volatility for different intervals within the day show a strong seasonal pattern. At the opening and closure of financial markets, price changes are more volatile than at other times during the trading session. In 24-hour markets, such different volatile periods within the day can be found too. Discussions of this phenomenon and empirical evidence are given by, among many others, Dacarogna, Miiller, Nagler, Olsen, and Pictet (1993) and Andersen and Bollerslev (1997). To account for the intra-daily variation of integrated volatility we replace the constant spot volatility σ^2 in (5) by an intra-daily seasonal specification in the volatility, that is

$$\sigma_t^2 = \sigma^2 \exp g(t), \quad \text{or} \quad \log \sigma_t^2 = \log \sigma^2 + g(t)$$

where $g(t)$ is a deterministic function that can represent a diurnal pattern and starts at zero, that is $g(0) = 0$. The function $g(t)$ is typically very smooth so that deviations from a diurnal pattern are not captured by $g(t)$. An example of an appropriate specification for $g(t)$ is given in Appendix A. The integrated volatility becomes

$$\sigma^{*2}(0,t) = \int_0^t \sigma_s^2 \, \mathrm{d}s = \sigma^2 \int_0^t \exp g(s) \, \mathrm{d}s \tag{8}$$

The actual volatility can be analytically derived from (8) or it can be approximated by

$$\sigma^{*2}(t_n, t_{n+1}) \approx \sigma^2 \sum_{s=t_n}^{t_{n+1}} \exp \; g(s)$$

with σ^2 representing the constant variance part and where the index step length can be chosen to be very small. As a result, σ_ε^2 in (7) is replaced by $\sigma_{\varepsilon,n}^2 = \sigma^{*2}(t_n, t_{n+1})$. The function $g(t) = g(t;\psi)$ depends on parameters that are collected in vector ψ, together with the variances σ_ε^2 and σ_U^2. This parameter vector can be estimated by ML methods. As a result, model (7) is unchanged except that the state variance has become dependent of a deterministic function of time. The Kalman filter can incorporate time-varying coefficients and therefore the estimation methodology remains straightforward.

2.3. Stochastic Volatility Model

Various specifications for SV models have been proposed. To keep the analysis and estimation simple, we will assume one of the most basic, non-trivial specifications. The efficient price process (5) is extended as follows. The constant volatility σ is replaced by a stochastic time-varying process. The price process can then be described by the system of SDE's given by

$$\begin{aligned} \mathrm{d} \log P_t &= \sigma_t \mathrm{d}B_t^{(1)} \\ \log \sigma_t^2 &= \log \sigma_t^{\prime 2} + \xi \\ \mathrm{d} \log \sigma_t^{\prime 2} &= -\lambda \log \sigma_t^{\prime 2} \mathrm{d}t + \sigma_\eta \mathrm{d}B_t^{(2)} \end{aligned} \tag{9}$$

where $B_t^{(1)}$ and $B_t^{(2)}$ are independent Brownian motions and where $\log \sigma_t^{'2}$ is modelled by an Ornstein–Uhlenbeck process. The fixed mean of log volatility is given by the constant ξ. The vector of unknown parameters is $\psi = (\lambda, \xi, \sigma_\eta^2)'$ Using the Euler–Maruyama method (see Kloeden and Platen, 1999 for details) we obtain an approximation to the solution of the system of SDEs (9) as given by the discrete model representation

$$\begin{array}{lcl} \log P_{t_{n+1}} & = & \log P_{t_n} + \sigma_n \varepsilon_n, \quad \varepsilon_n \sim \text{NID}(0,1) \\ \log \sigma_n^2 & = & \log \sigma_n^{'2} + \xi, \\ \log \sigma_{n+1}^{'2} & = & (1-\lambda)\log \sigma_n^{'2} + \sigma_\eta \eta_n, \quad \eta_n \sim \text{NID}(0,1) \end{array} \tag{10}$$

for $n = 1, \ldots, N_d$. Note that $\lambda = \sigma_\eta = 0$ implies constant volatility with $\log \sigma_n^2 = \xi$. The set of equations (10) represents the standard discrete SV model (see Ghysels, Harvey, & Renault, 1996 for an overview). It follows that the actual volatility is approximated by

$$\sigma^{*2}(t_n, t_{n+1}) = \int_{t_n}^{t_{n+1}} \sigma_s^2 \, \mathrm{d}s \approx (t_{n+1} - t_n)\sigma_n^2$$

Finally, assuming that a particular day d consists of Nd intra-day intervals, the actual volatility of day d is approximated by

$$\sigma^{*2}(t_0, t_{N_d}) \approx \sum_{n=0}^{N_d-1} (t_{n+1} - t_n)\sigma_n^2$$

To analyse the stochastic log prices (mean) and the SV (variance) simultaneously, it is more convenient to represent the model in terms of returns $\log(P_{t_{n+1}}/P_{t_n})$. It follows from the discussion in Section 2.1 that when the model for log prices accounts for micro-structure noise, the observed returns R_n follows an MA (1) process. By further allowing for SV, we obtain

$$R_n = \sigma_n \varepsilon_n + \sigma_U W_n \tag{11}$$

where $\log \sigma_n^2$ is modelled as in (10) and $W_n = U_{n+1} - U_n$ such that $W_n \sim$ MA (1). From an estimation point of view, it will be argued in the next section that ML estimation of the model (11) with an MA (1) noise term or a general ARMA term is intricate. We therefore leave this problem as for future research and consider a white noise process for W_n in the empirical part of this chapter.

The final model that we consider is the price model with SV that also accounts for the intra-daily seasonal pattern. In the previous section, we have introduced the flexible deterministic function $g(t)$ for this purpose. The final model is therefore based on the system of SDEs

$$\begin{aligned} \mathrm{dlog}\, P_t &= \sigma_t \mathrm{d}B_t^{(1)} \\ \log \sigma_t^2 &= \log \sigma_t^{'2} + g(t) + \xi \\ \mathrm{dlog}\, \sigma_t^{'2} &= -\lambda \log \sigma_t^{'2} \mathrm{d}t + \sigma_\eta \mathrm{d}B_t^{(2)} \end{aligned}$$

The flexible function $g(t)$ is incorporated in the SV specification (10) in the same way as described in Section 2.2. In particular, $\log \sigma_n^2$ in (10) is replaced by

$$\log \sigma_n^2 = \log \sigma_n^{'2} + g(n) + \xi$$

where $t_{n+1} - t_n$ is assumed constant for all n.

3. ESTIMATION METHODS

3.1. The Problem of Estimation

It is already argued in Sections 2.1 and 2.2 that the model for prices with constant or deterministic time-varying volatilities is relatively straightforward to estimate by using Kalman filter methods. However, estimating the model with SV is known to be much more intricate. Various methods have been developed for the estimation of the SV model without micro-structure noise. Such methods have been based on QML, Bayesian Markov chain Monte Carlo procedures, importance sampling techniques, numerical integration, method of moments, etc. Presenting an overview of all these methods is beyond the scope of this chapter but the interested reader is referred to the collection of articles in Shephard (2005). None of these methods have considered the existence of micro-structure noise in the returns since most empirical applications have only been concerned with returns data measured at lower frequencies such as months, weeks and days. The issue of micro-structure noise is less or not relevant in such cases. This section discusses feasible methods for the estimation of parameters in models for returns with SV plus noise since this is relevant for high-frequency data. We limit ourselves to approximate and ML methods. Bayesian and (efficient and/or simulated) method of moments can be considered as well and in fact we believe that such methods are applicable too. However, given our favourable experiences with ML estimation using importance sampling techniques for standard SV models, we have been encouraged to generalise these methods for the models described in the previous section.

To focus the discussion on estimation, the model for returns with SV, intra-daily seasonality and micro-structure noise is represented as the nonlinear state space model

$$R_n = \sigma_\varepsilon \varepsilon_n + \sigma_U W_n \tag{12}$$

$$\sigma_n^2 = \exp\{h_n + g(n) + \xi\}$$

$$h_{n+1} = \phi h_n + \sigma_\eta \eta_n \tag{13}$$

where $h_n = \log \sigma_n^{'2}$ and$\phi = 1 - \lambda$ The log-volatily h_n follows an autoregressive process of order one and the micro-structure noise W_n follows an MA process of order one. These processes can be generalised to other stationary time-series processes. The disturbances driving the time-series processes for h_n and W_n together with ε_n are assumed Gaussian and independent of each other, contemporaneously and at all time lags. These assumptions can be relaxed, see the discussion in Section 3.3. The returns model (12) is nonlinear and depends on a state vector with log variance log $\sigma_n^{'2}$ modelled as a linear autoregressive process together with constant ξ and with intra-daily volatility pattern $g(t)$. The nonlinearity is caused by the term $\exp(1/2h_n)\varepsilon_n$ in (12) since both h_n and ε_n are stochastic. Conditional on the unobservable h_n, model (12) can be viewed as a linear Gaussian ARMA model (for the micro-structure noise $\sigma_U W_n$) with additive heteroscedastic noise (for the returns $\log P_{t_{n+1}} - \log P_{t_n}$).

Different approximation methods for the estimation of the unknown parameters in model (12) and (13) can be considered. For example, the multiplicative term $\exp(h_n/2)\varepsilon_n$ can be linearised by a first-order Taylor expansion in h_n. The resulting linearised model can be considered by the Kalman filter. This approach is referred to as the Extended Kalman filter. The details will not be discussed here since we believe that this approach will provide a poor approximation especially when the volatility is relatively large or small, that is, when $|h_n|$ is large. Some improvements may be obtained when the resulting estimate of h_n is inserted in the model so that a linear model is obtained which can be treated using standard methods. Such a mix of approximate methods does not lead to a satisfactory estimation strategy and therefore we aim to provide a ML estimation method in the next section.

3.2. Estimation Using Importance Sampling Techniques

The estimation of parameters in discretised SV models, that is models (12) and (13) with $\sigma_U = 0$, is not standard since a closed expression for the likelihood function does not exist. Estimation can be based on approximations such as QML (see Harvey, Ruiz, & Shephard, 1994), numerical integration methods for evaluating the likelihood (see Fridman & Harris, 1998), and Markov chain Monte Carlo (MCMC) methods (see Jacquier, Polson, & Rossi, 1994 and Kim, Shephard, & Chib, 1998). In this chapter, we focus on Monte Carlo methods of evaluating the likelihood function of the SV model (see Danielsson, 1994 and Sandmann & Koopman, 1998 for some earlier contributions in this respect). The evaluation of the likelihood function using Monte Carlo importance sampling techniques has been considered for the models (12) and (13) with $\sigma_U = 0$ by Shephard and Pitt (1997) and Durbin and Koopman (1997). Further details of this approach have been explored in Part II of the monograph of Durbin and Koopman (2001). The basic ingredients of this approach are as follows:

- The approximate linear Gaussian model
$$y = \theta + u, \quad u \sim \mathrm{NID}(c, V) \tag{14}$$
is considered with its conditional density denoted by $g(y|\theta)$ where y is the vector of observations and θ the associated unobserved signal. In the SV model without noise, we have $y = (R_1, \ldots, R_{N_d})'$ and $\theta = (h_1, \ldots, h_{N_d})'$ The approximate conditional Gaussian density $g(y|\theta)$ depends on mean vector c and diagonal variance matrix V which are chosen such that
$$\dot{g}(y|\theta) = \dot{p}(y|\theta), \quad \ddot{g}(y|\theta) = \ddot{p}(y|\theta)$$
where $\dot{q}(\cdot)$ and $\ddot{q}(\cdot)$ are the first and second derivatives, respectively, of the density $q(\cdot)$ with respect to θ. Further, $p(\cdot)$ refers to the density of models (12) and (13), here with $\sigma_U = 0$. To obtain the mean and variance of $g(y|\theta)$, we require to estimate θ from the approximate linear Gaussian model (14) that also depends on θ. Therefore an iterative method involving Kalman filtering and smoothing needs to be carried out.

- Given the importance density associated with the approximate model (14), simulations from density $g(\theta|y)$ can be obtained using simulation smoothing algorithms such as the recent ones of de Jong and Shephard (1995) and Durbin and Koopman (2002). The resulting simulated $\theta' s$ are denoted by $\theta^{(i)} \sim g(\theta|y)$.

- The importance sampling estimator of the likelihood is based on

$$\begin{aligned} L(\psi) = p(y;\psi) &= \int p(y,\theta)\,\mathrm{d}\theta = \int \frac{p(y,\theta)}{g(\theta|y)} g(\theta|y)\,\mathrm{d}\theta \\ &= g(y;\psi) \int \frac{p(y,\theta)}{g(y,\theta)} g(\theta|y)\,\mathrm{d}\theta \end{aligned}$$

and since $p(\theta) = g(\theta)$, we obtain the convenient expression

$$L(\psi) = L_g(\psi) \int \frac{p(y|\theta)}{g(y|\theta)} g(\theta|y)\,\mathrm{d}\theta$$

where $L_g(\psi) = g(y;\psi)$ is the likelihood function of the approximating model. All densities $p(\cdot)$ and $g(\cdot)$ depend on parameter vector ψ even when this is not made explicit. The importance sampling estimator of the likelihood function $L(\psi)$ is therefore given by

$$\widehat{L}(\psi) = L_g(\psi) \sum_{i=1}^{M} \frac{p(y|\theta^{(i)})}{g(y|\theta^{(i)})}$$

where $\theta^{(i)} \sim g(\theta|y)$ for $i = 1, ..., M$. It is noted that the densities p(y|θ) and g(y|θ) are relatively easy to evaluate. The likelihood function evaluated by importance sampling is exact but subject to Monte Carlo error.

The last items are general and do not depend on the particular model specification. Finding an approximate linear Gaussian model from which we can generate simulation samples from $g(y|\theta)$, does obviously depend on the model in question. The details of obtaining an approximate model for the standard SV model for importance sampling can be found in Shephard and Pitt (1997) and Durbin and Koopman (2001, p. 195). The values for c_n and V_n, the *n-th* element of c and the *n-th* diagonal element of V, respectively, in this case are obtained by

$$V_n = 2\frac{\exp(h_n)}{R_n^2}, \quad c_n = \frac{1}{2}V_n + R_n - h_n - 1 \tag{15}$$

For the case with micro-structure noise the values for c and V need to be derived as hinted in the first item. The details of the derivations are given in

Appendix B. For the case of IID noise, that is $W_n \sim \mathrm{IID}(0,1)$, the actual values are given by

$$V_n^{-1} = \frac{1}{2}(b_n - b_n^2) + \left(b_n - \frac{1}{2}\right)\frac{b_n}{a_n}R_n^2, \quad c_n = R_n - h_n - \frac{1}{2}V_n b_n\left(\frac{R_n^2}{a_n} - 1\right) \tag{16}$$

where $a_n = \exp(h_n) + \sigma_U^2$ and $b_n = \exp(h_n)/a_n$. We note that $a_n > 0$ and $0 < b_n \leq 1$. A strictly positive variance $V_n > 0$ for all n can not be guaranteed in this case except when $\sigma_U^2 < \exp(h_n)$ since this implies that $b_n > \frac{1}{2}$. However, in a recent development reported by Jungbacker and Koopman (2005), it is shown how the "negative variance" problem can be resolved in a satisfactory way.

3.3. Discussion of Estimation Methods

Details of importance sampling methods for estimating the general model are presented in the previous section. It is assumed that W_n is IID while the basic modelling framework for the prices in Section 2.1 insists that W_n should be modelled by an MA (1) process or possibly an ARMA process. The consideration of an ARMA disturbance term in the measurement equation requires multivariate sampling devices which are intricate and need to be developed in future work.

For estimation purposes the price model with SV is reformulated in terms of returns. The ultimate aim, however, is to estimate models as specified in (10). The estimation of parameters in such models is not an easy task and various methods can be considered. In this chapter, we have considered Kalman filter and importance sampling techniques. This leads to feasible methods but it is not yet clear how they can be utilised more effectively to treat models such as (10) directly.

Other estimation techniques can also be adopted with numerical integration, simulated method of moments and Bayesian methods as obvious examples. It should be noted that the number of transactions in one trading day can be as big as 23,400 but is usually between 1,000 and 5,000 for a liquid stock. As a consequence, the integral of the likelihood function is of a very high dimension and therefore numerical integration is not feasible.

As far as we know, effective methods of moments and Bayesian methods are not developed as yet for models such as (10). For example, the MCMC method of Kim et al. (1998), in which candidate samples are generated by approximate densities based on mixture of normals, can not be used straightforwardly for this class of models.

4. EMPIRICAL RESULTS FOR THREE MONTHS OF IBM PRICES

4.1. Data

A small subset of the the TAQ database for the NYSE is analysed in the empirical study below. We only consider the IBM equity transactions reported on Consolidated Tape. The IBM stock is a heavily traded and liquid stock. The NYSE market opens at 9:30 AM and closes at 4 PM. Prices of transactions made outside these official trading hours have been deleted. The resulting database consists of prices (measured in cents of US dollars) and times (measured in seconds) of transactions realised in the three months November 2002, December 2002 and January 2003. No further manipulations have been carried out on this dataset. The prices for each trading day are considered as a time series with the time index in seconds. This time series have possibly many missing observations. For example, when no trade has taken place in the last 2 minutes, we have at least 120 consecutive missing values in the series. The treatment of missing values is handled by the Kalman filter and does not lead to computational or numerical inefficiencies.

4.2. Measuring Actual Volatility for One Day

As a first illustration, we consider tick-by-tick prices and returns of IBM realised on the NYSE trading day of November 1, 2002. In Fig. 1 the prices and returns are presented for the hourly intervals of the trading day. The number of trades that has taken place on this day is 3,321. Given that a trading day consists of 23,400 seconds, that is 6.5 trading hours times 3,600 seconds in 1 hour, the average duration between trades is 7.05 seconds. In other words, on average, 511 trades in 1 hour and 8.5 trades in 1 minute has been realised. However, approximately, the first 300 trades took place before 10 am and the last 600 trades took place after 3 pm. The time series of prices and returns presented in Fig. 1 are against an index of seconds. This means that 23,400 observations can be displayed but

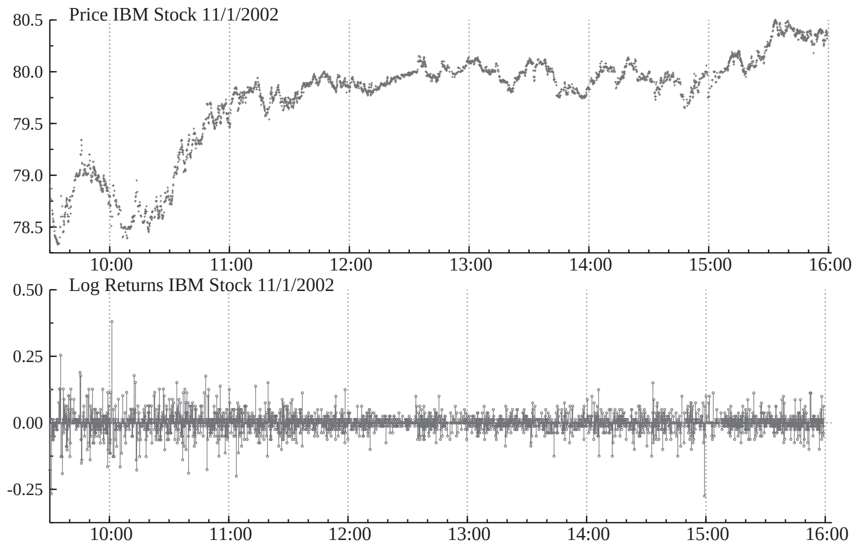


Fig. 1. IBM Stock Prices and Returns for All Trades on November 1, 2002. The tick-by-tick data is Presented Against Time in Seconds.

only 3,321 transactions have been realised, resulting in 20,079 missing values on this day. We note that no multiple trades occurred in the same second. These facts aim to put the plots of Fig. 1 into some perspective. Due to the lack of resolution in our graphs, the majority of missing values go almost unnoticed.

In Fig. 2 we present prices, returns and squared log returns of the IBM stock for November 1, 2002. Here the index is trade by trade. Nevertheless, the series of prices in Figs. 1 and 2 appear to be very similar. This is again due to the limited resolution that can be provided in these graphs. In any case, both plots of returns show that volatility is substantially higher at the beginning of the trading day and somewhat higher at the end of the trade session. The small price variation in the middle of the trading day is probably due to the fact that no relevant information has arrived in these hours.

To analyse the trade prices on this day, we first consider the model for prices (5) with constant volatility σ and intra-daily pattern $g(t)$ for spot volatility. For the function of $g(t)$ we adopt the cubic spline as described in Appendix A with three knots $\{\gamma_1, \gamma_2, \gamma_3\}$ of which two are at either ends of the trading day and one is placed in the middle of the day.

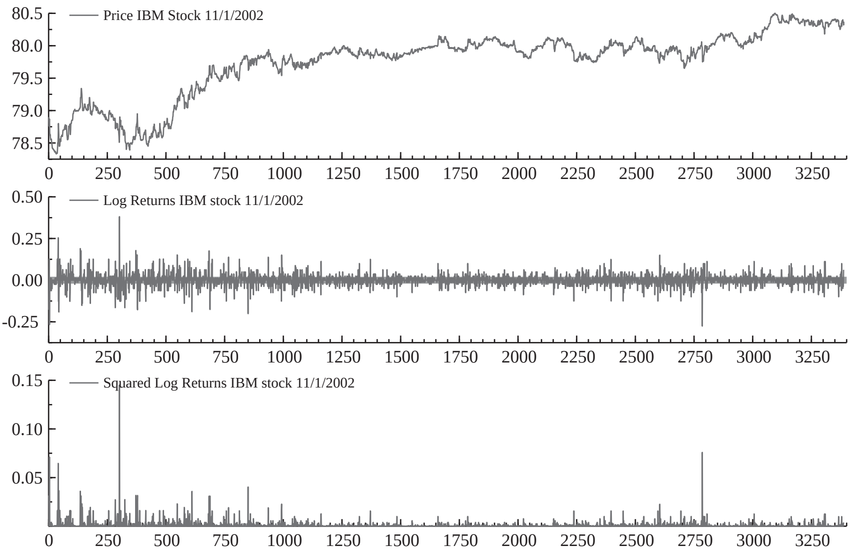


Fig. 2. IBM Stock Prices, Returns and Squared Log Returns for All Trades on November 1, 2002. The Data are Presented Against the Trade Index So that Every Tick is One Trade.

The first knot coefficient γ_1 is restricted to be zero so that $g(t_0) = 0$. It is argued in Section 2.1 that the standard Kalman filter can be used in the estimation of coefficients for this model. The Kalman filter as implemented in the SsfPack package of Koopman, Shephard, and Doornik (1999) allows for missing values and deterministic time-varying variances. We have implemented the calculations in the Ox package of Doornik (2001) using the state space library of SsfPack, version 3. The estimation results are given by

$$\log \widehat{\sigma} = -5.112, \quad \widehat{\gamma}_2 = -1.747, \quad \widehat{\gamma}_3 = -1.135$$

These reported values provide some initial indication of results that can be obtained from a high-frequency analysis.

More interestingly from theoretical and empirical perspectives are the results for the returns model with SV and intra-daily seasonality. In particular, we focus on the differences in estimation results for models with or without micro-structure noise. The estimation method for the model with

SV and noise requires importance sampling techniques as discussed in Section 3.2. The necessary calculations are implemented in Ox with intensive use of the SsfPack to obtain an approximating model and to simulate random samples of log volatility conditional on returns.

The parameter estimates are as follows. For models (12) and (13) without micro-structure noise $\sigma_U = 0$, we have

$$\widehat{\phi} = 0.961, \quad \widehat{\sigma}_\eta^2 = 0.0619, \log \widehat{\sigma} = -7.977, \quad \widehat{\gamma}_2 = -1.654, \quad \widehat{\gamma}_3 = -1.135$$

For the SV model with micro-structure noise, we have

$$\begin{aligned} &\widehat{\sigma}_U^2 = 0.00003985, \quad \widehat{\sigma}_U = 0.00631, \quad \widehat{\phi} = 0.955, \\ &\widehat{\sigma}_\eta^2 = 0.0821, \quad \log\widehat{\sigma} = -8.033, \quad \widehat{\gamma}_2 = -1.629, \quad \widehat{\gamma}_3 = -1.065 \end{aligned}$$

In comparison with the earlier results for a model with a constant plus spline volatility, the estimates $\hat{\sigma}$ are smaller since the stochastic part of log volatility also accounts for part of the variance. The persistence of log volatility is in the same order when the model is estimated with noise or without noise. Although apparently the micro-structure noise seems low, it has a big impact on the estimate $\hat{\sigma}_\eta$. In fact, this estimate $\hat{\sigma}_\eta$ has increased after micro-structure noise is included in the model. It can be concluded that more variation is attributed to the stochastic part rather than the constant part of volatility, especially when micro-noise is excluded from the observed returns.

In Fig. 3, we present the estimated volatility components for this day. The time-series length is 23,400 seconds for which 20,079 seconds have recorded no price. During the model estimation process, these 20,079 non-available prices are treated as missing observations. This approach does not lead to computational or numerical inefficiencies. The estimated prices and returns are obtained using the importance sampling methods for filtering and smoothing, (see Durbin & Koopman, 2001, Chapter 11 and Appendix B for further details). As a result, we obtained 23,400 estimates for which the vast majority of values are the result of interpolations implied by the estimated model. To provide a somewhat more detailed insight, we also present estimates of log σ_t' for a smaller interval of 30 minutes and four intervals of 5 minutes in Figs. 4 and 5, respectively. It shows clearly that when returns are sampled every 30 minutes or every 5 minutes, much of the variation in the returns is unaccounted for.

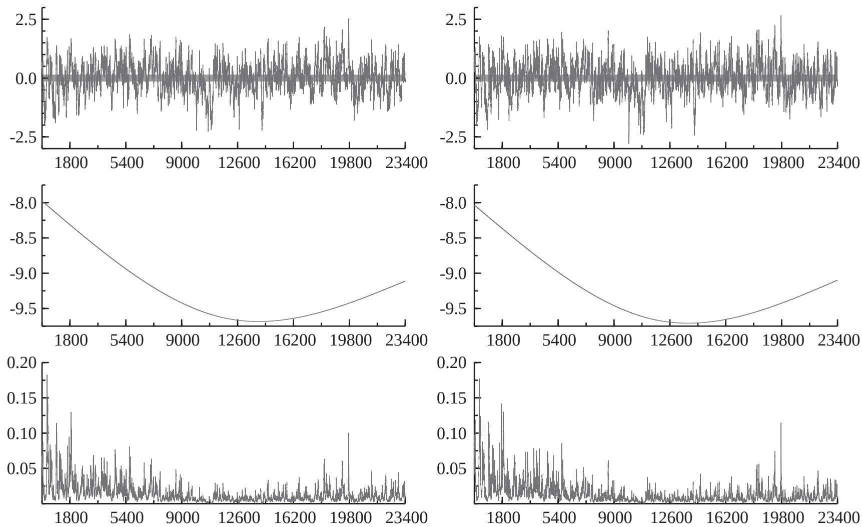


Fig. 3. Estimated Intra-Day Volatility: (i) Log-Volatility Component Log $\hat{\sigma}_n^{'}$; (ii) Intra-Daily Volatility Pattern $\hat{g}(n)$; (iii) Integrated Volatility $\sigma^{*2}(t_{n-1}, t_n)$.

4.3. Measuring Actual Volatility for Three Months

The model-based methods for estimating coefficients and for measuring volatility are implemented satisfactorily. Several limited simulation studies have been carried out and they confirm the reliability of the implemented procedures. Subsequently, we repeat the analysis for a large dataset of IBM stock returns for 61 consecutive trading days in the months November 2002, December 2002 and January 2003. We present in Fig. 6 the measures obtained for standard realised volatility calculations, for a model with constant volatility plus micro-structure noise, for a model with constant, spline and SV, and for a model with constant, spline and SV plus micro-structure noise.

The patterns of the volatility measures are similar although the variation among different days is different. The levels of realised volatility and of estimates from a constant volatility plus noise model are comparable with each other. However, the model-based measure is somewhat less noisy since the micro-structure noise is separated from the underlying price changes. The levels of volatility measures obtained from models with splines and SV are higher compared to the earlier two basic measures. The difference is due to the fact that variations between, say, 5 minutes are not considered in the

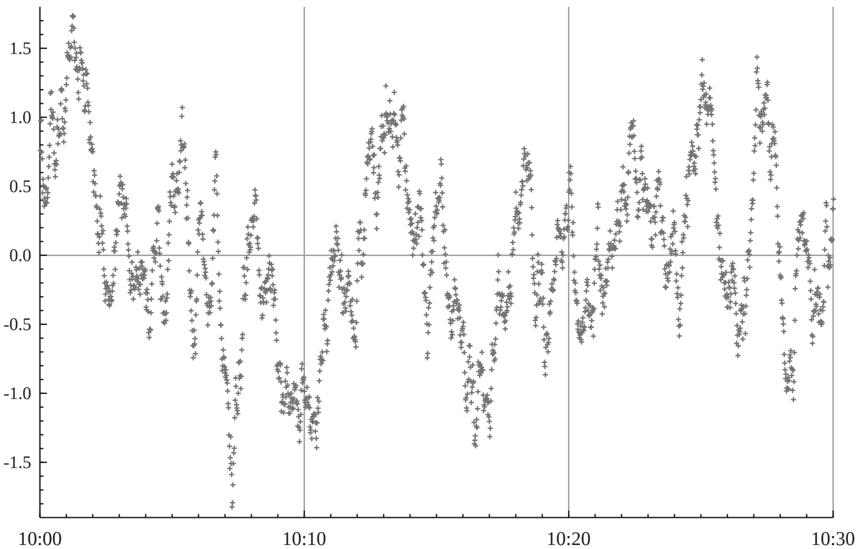


Fig. 4. Estimated Log-Volatility Component Log $\hat{\sigma}_n'$ for an interval of 30 min.

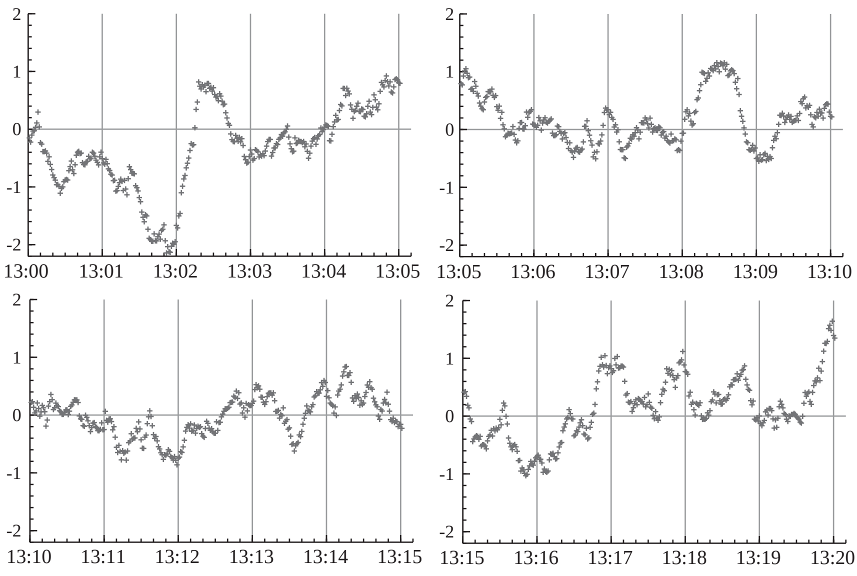


Fig. 5. Estimated Log-Volatility Component Log $\hat{\sigma}_n'$ for Four Intervals of 5 min.

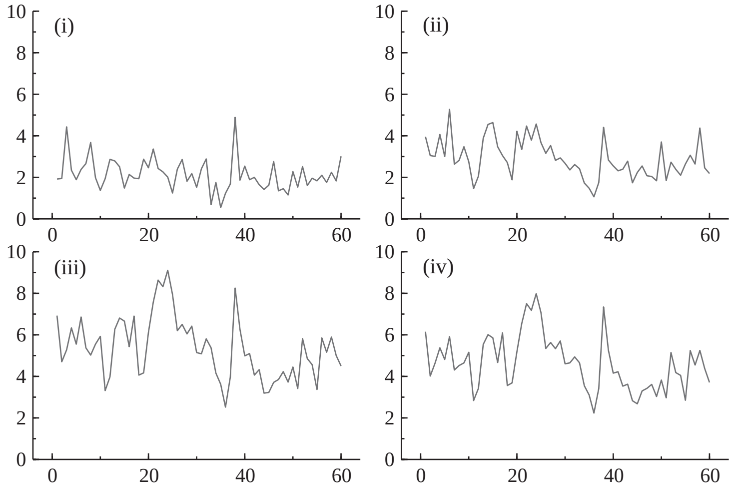


Fig. 6. Volatility measures: (i) Realised Volatility; (ii) Estimates from a Constant Volatility Model with Noise; (iii) Estimates from a Constant, Spline and SV Model; (iv) Estimates from a Constant, Spline and SV Model with Noise. The Volatility Estimates are for the 61 Trading Days of November 2002, December 2002 and January 2003.

constant volatility measures. Fig. 5 nicely illustrates the amount of variation that is missed when sampling takes place every 5 minutes or even every 1 minute. In the SV modelling framework, all variations within the day, at a frequency as high as seconds, are taken into account. This clearly leads to a higher scaling of volatility. In the case of the model with constant volatility, the estimates are lower since they are close to a mean of squared log returns which implies that excessive variations are averaged out. This does not apply to the other SV models, where the variance is decomposed into different effects such as intra-day diurnal effects and SV.

The difference between the models with micro-structure noise and without noise seems relatively small. However, we note that the number of trades in each day are between 2,000 and 5,000. It is clear that the volatility estimates for models with SV and noise are somewhat lower compared to SV models without noise, as the former model attributes part of the noise to micro-structure effects.

Finally, we display the sample autocorrelation functions for the daily volatility measures in Fig. 7. Although it is somewhat surprising that the

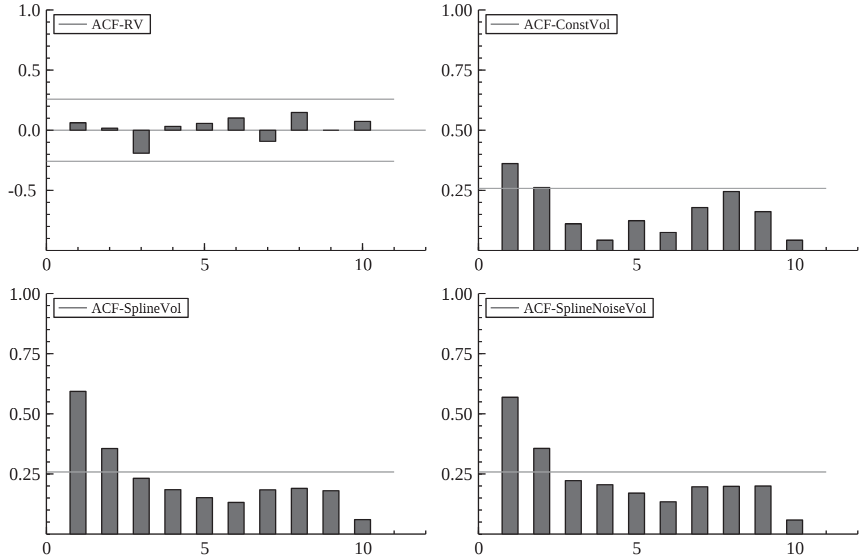


Fig. 7. Sample Autocorrelation Functions of the Volatility Measures from Fig. 6. The Autocorrelation Coefficients are Therefore Based on 61 Datapoints.

correlogram for realised volatility is not significant at any lag despite the widely accepted view that realised volatility is serially correlated and can be effectively modelled as an autoregressive fractional integrated moving average (ARFIMA) process (see, for example, Andersen, Bollerslev, Diebold, & Labys, 2003). However, for the realised volatility series analysed in Koopman, Jungbacker, and Hol (2005), many instances are encountered where the correlogram is also not significant when random subsamples of length 100 are considered. Note that for the full sample of approximately 1,500 daily realised volatilities, a significantly persistent correlogram is present. In the analysis of this section, it appears that model-based measures of volatility are persistent over days, especially when SV is modelled explicitly. The daily time series of model-based volatility measures are relatively smooth and clearly contain some level of persistency. These preliminary results may have shown that the supposed long memory property of realised volatility may not exist for a relatively small number of days whereas for high-frequency measures the persistence of daily volatility estimates remain to exist. In the latter case, daily volatilities can still be modelled as ARFIMA processes even for small samples.

5. DISCUSSION AND CONCLUSION

We have proposed to measure volatility from high-frequency prices using a model-based framework. A standard basic model is considered that captures the salient features of prices and volatilities in financial markets. In particular, it accounts for micro-structure noise, an intradaily volatility pattern and SV. Feasible estimation methods have been implemented for this class of models and the illustration shows that this approach can work effectively in determining the volatility in financial markets using tick-by-tick data. As a result, no information is lost as opposed to realised volatility for which prices are sampled at a low frequency, say 5 or 10 minutes. Therefore, a part of the variation in prices is lost in realised volatility. When more detailed comparisons are made between realised volatility and the high-frequency measures, it is shown that the supposed long memory property of realised volatility may not be identified from a relatively small number of days whereas for high-frequency measures the persistence of daily volatility estimates remain. However, more empirical investigation is needed to obtain further insights on this issue. Nonparametric methods have also been proposed recently to tackle the problem of micro-structure noise. However, as far as we know, this chapter presents a first attempt to analyse ultra high-frequency prices using a model that simultaneously accounts for micro-structure noise and SV.

REFERENCES

Ait-Sahalia, Y., Mykland, P. A. & Zhang, L. (2004). How often to sample a continuous-time proces in the presence of market microstructure noise. *Review of Financial Studies 17,* forthcoming.

Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, *4*, 115–158.

Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2002). Parametric and nonparametric volatility measurement. In: *Handbook of Financial Econometrics* Amsterdam: North-Holland. (Forthcoming)

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association*, *96*, 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*, 579–625.

Andreou, E., & Ghysels, E. (2001). Rolling sample volatility estimators: Some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics*, *20*, 363–376.

Bai, X., Russell, J. R. & Tiao, G. C. (2000). *Beyond mertons utopia: Effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data*. Manuscript, Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2004). *Regular and modified kernel-based estimators of integrated variance: The case with independent noise*. Working paper, Nuffield College.

Barndorff-Nielsen, O. E., & Shephard, N. (2001). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of Royal Statistical Society B*, *63*, 253–280.

Campbell, J. Y., Lo, A. W., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton, NJ: Princeton University Press.

Dacarogna, M. M., Miiller, U. A., Nagler, R. J., Olsen, R. B., & Pictet, O. V. (1993). A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Financial Economics*, *12*, 413–438.

Danielsson, J. (1994). Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics*, *61*, 375–400.

de Jong, P., & Shephard, N. (1995). The simulation smoother for time series models. *Biometrika*, *82*, 339–350.

Doornik, J. A. (2001). *Object-oriented matrix programming using Ox 3.0* (4th ed.). London: Timberlake Consultants Ltd. See http://www.doornik.com.

Durbin, J., & Koopman, S. J. (1997). Monte Carlo maximum likelihood estimation of non-Gaussian state space model. *Biometrika*, *84*, 669–684.

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.

Durbin, J., & Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, *89*, 603–616.

Fridman, M., & Harris, L. (1998). A maximum likelihood approach for non-gaussian stochastic volatility models. *Journal of Business and Economic Statistics*, *16*, 284–291.

Ghysels, E., Harvey, A. & Renault, E. (1996). Stochastic volatility. In *Handbook of Statistics, (Vol. 14),* (pp. 119–191). *Statistical Methods in Finance,* Amsterdam: North-Holland.

Hansen, P. R. & Lunde, A. (2004). A realized variance for the whole data based on intermittent high-frequency data. Working paper, Stanford University.

Harvey, A. C. (1989). *Forecasting, structural time series models and the kalman filter*. Cambridge: Cambridge University Press.

Harvey, A. C., & Koopman, S. J. (2000). Signal extraction and the formulation of unobserved components models. *Econometrics Journal*, *3*, 84–107.

Harvey, A. C., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *Review Economic Studies*, *61*, 247–264.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics*, *12*, 371–417.

Jungbacker, B., & Koopman, S. J. (2005). *On importance sampling for state space models*. Amsterdam: mimeo, Vrije Universiteit.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review Economic Studies*, *65*, 361–393.

Kloeden, P., & Platen, E. (1999). *Numerical solution of stochastic differential equations*. Berlin: Springer-Verlag.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, *12*, 445–475.

Koopman, S. J., Shephard, N. & Doornik, J. A. (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal, 2,* 113–66. http://www.ssfpack.com/.

Malliavin, P., & Mancino, M. (2002). Fourier series method for measurment of multivariate volatilities. *Finance and Stochastics*, *6*, 49–61.

Poirier, D. J. (1976). *The econometrics of structural change: with special emphasis on spline functions*. Amsterdam: North-Holland.

Sandmann, G., & Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics*, *87*, 271–301.

Shephard, N. (2005). *Stochastic volatility: Selected readings*. Oxford: Oxford University Press.

Shephard, N., & Pitt, M. K. (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, *84*, 653–667.

APPENDIX A. CUBIC SPLINE FOR INTRA-DAILY VOLATILITY PATTERN

The intra-daily pattern of volatility is captured by a flexible function $g(t)$. In this chapter, we take $g(t)$ as a cubic spline function. We follow Poirier (1976) in developing a cubic spline. Given a mesh of, say 3, x values ($\{x_0, x_1, x_2\}$) and a set of corresponding y values ($\{y_0, y_1, y_2\}$, respectively), the y values for $x_{j-1} \leq x \leq x_j$ can be interpolated by

$$\begin{aligned} y = g(x) = & \frac{(x_j - x)^3}{6(x_j - x_{j-1})} z_{1,j-1} + \frac{(x - x_{j-1})^3}{6(x_j - x_{j-1})} z_{1,j} + (x_j - x) z_{2,j} \\ & + (x - x_{j-1}) z_{3,j} + z_{4,j}, \quad j = 1, 2 \end{aligned}$$

where $z_{i,j}$ is an unknown coefficient for $i = 1, 2, 3, 4$ and $j = 1, 2$. The coefficients $z_{i,j}$ are determined by restricting smoothness conditions on $g(x)$ such as continuity at x_j $(j = 1, 2)$ of the spline itself and its first and second derivatives. The resulting set of equations can be solved in $z_{i,j}$ via standard matrix algebra. Given a solution for $z_{i,j}$, the spline function can be expressed as

$$g(x) = \sum_{j=0}^{2} w_j y_j, \quad \sum_{j=0}^{2} w_j = 1$$

where weights w_j depend on x and the mesh $\{x_0, x_1, x_2\}$.

APPENDIX B. APPROXIMATING MODEL FOR SV WITH NOISE

Consider a non-linear state-space model where the state equation is linear Gaussian and the distribution of the observations $Y=(Y_1, \ldots, Y_N)$ conditional on the states $h=(h_1, \ldots, h_N)$ is determined by the probability density $p(Y_n|h_n)$, $n=1, \ldots, N$. It is evident that the SV models considered in the main text are special cases of this class of models, the interested reader is referred to Durbin and Koopman (2001) for more examples. For the importance sampling procedure a linear Gaussian approximating model is chosen with the same state equation as the true model but with an observation equation given by

$$Y_n = c_n + h_n + u_n, \quad u_n \sim \mathrm{NID}(0, V_n) \tag{B.1}$$

where the constants c_n and V_n have to be chosen in a suitable manner. The approach advocated in Durbin and Koopman (2001) consists of choosing c_n and V_n for $n=1, \ldots, N$ such that the true smoothing density, $p(h|Y)$, and the smoothing density of the approximating model, $g(h|Y; V, c)$, have the same modes and equal curvatures around these modes. This means, denoting $V=(V_1, \ldots, V_N)'$ and $c=(c_1, \ldots, c_N)'$, that V and c are solutions to the system of equations defined by

$$\frac{\partial p(Y,h)}{\partial h_n} = \frac{\partial g(Y,h; c, V)}{\partial h_n} = 0$$

and

$$\frac{\partial^2 p(Y,h)}{\partial h_n^2} = \frac{\partial^2 g(Y,h; c, V)}{\partial h_n^2}$$

for $n=1, \ldots, N$. Solving these equations for c and V is as follows. First of all, the mean and the mode have the same location for a Gaussian density. This means that, conditional on c and V, the mode, $\hat{h}=(\hat{h}_1, \ldots, \hat{h}_N)$, can be obtained by computing the mean of $g(h|V, c)$, a problem that is routinely handled by the Kalman filter and smoother. On the other hand, the fact that the marginal distribution of h is equal for both the true and the approximating models, combined with the monotonicity of the log transformation, implies that the system of equations is equivalent to

$$\frac{\partial \log p(Y|h)}{\partial h_n} = \frac{\partial \log g(Y|h; c, V)}{\partial h_n} = 0$$

and

$$\frac{\partial^2 \log p(Y|h)}{\partial h_n^2} = \frac{\partial^2 \log g(Y|h; c, V)}{\partial h_n^2}$$

implying that conditional on the mode h a solution to this set of equations is given by the vectors V and c satisfying

$$\left.\frac{\partial \log p(Y_n|h_n)}{\partial h_n}\right|_{h_n=\hat{h}_n} = \left.\frac{\partial \log g(Y_n|h_n; c, V)}{\partial h_n}\right|_{h_n=\hat{h}_n}$$

and

$$\left.\frac{\partial^2 \log p(Y_n|h_n)}{\partial h_n^2}\right|_{h_n=\hat{h}_n} = \left.\frac{\partial^2 \log g(Y_n|h_n; c, V)}{\partial h_n^2}\right|_{h_n=\hat{h}_n}$$

for $n = 1, \ldots, N$. If we now use

$$\frac{\partial \log g(Y_n|h_n; c, V)}{\partial h_i} = \frac{Y_n - h_n - c_n}{V_n}$$

and

$$\frac{\partial^2 \log g(Y_n|h_n; c, V)}{\partial h_n^2} = \frac{1}{V_n}$$

then these expressions imply

$$V_n = \left(\frac{\partial^2 \log p(Y_n|h_n)}{\partial h_n^2}\right)^{-1} \tag{B.2}$$

and

$$c_n = Y_n - h_n - V_n \frac{\partial \log p(Y_n|h_n)}{\partial h_n} \tag{B.3}$$

These two observations suggest the following algorithm

1. Choose a starting value h^1 for $\hat{h}$.
2. Adopt h^i to obtain c^i and V^i using (B.2) and (B.3) for $i = 1, 2, \ldots$. Create a new proposal for $\hat{h}$ that is h^{i+1}, by applying the Kalman smoother to $Y_1, \ldots, Y_N$ for the model defined by (B.1), with $c = c^i$and $V = V^i$
3. Keep repeating 2 until $||h^{i+1} - h^i|| < \varepsilon_c$, where ε_c is some small threshold value.

To implement this algorithm for the SV models considered in this chapter, the only thing that remains is the calculation of the derivatives in (B.2) and

(B.3). For the SV model defined in (10) these derivatives are given by

$$\frac{\partial \log p(Y_n|h_n)}{\partial h_n} = \frac{1}{2}\left(\frac{Y_n^2}{\exp h_n} - 1\right), \quad \frac{\partial^2 \log p(Y_n|h_n)}{\partial h_n^2} = -\frac{Y_n^2}{2\exp h_n}$$

For the SV model with micro-structure noise defined in (11), we have

$$\frac{\partial \log p(Y_n|h_n)}{\partial h_n} = \frac{1}{2} b_n \left(\frac{Y_n^2}{a_n} - 1\right)$$

$$\frac{\partial^2 \log p(Y_n|h_n)}{\partial h_n^2} = \left(\frac{1}{2} - b_n\right) \frac{b_n Y_n^2}{a_n} - \frac{1}{2}\left(b_n - b_n^2\right)$$

where $a_n = \exp(h_n) + \sigma_U^2$ and $b_n = \exp(h_n)/a_n$.

NOISE REDUCED REALIZED VOLATILITY: A KALMAN FILTER APPROACH

John P. Owens and Douglas G. Steigerwald

ABSTRACT

Microstructure noise contaminates high-frequency estimates of asset price volatility. Recent work has determined a preferred sampling frequency under the assumption that the properties of noise are constant. Given the sampling frequency, the high-frequency observations are given equal weight. While convenient, constant weights are not necessarily efficient. We use the Kalman filter to derive more efficient weights, for any given sampling frequency. We demonstrate the efficacy of the procedure through an extensive simulation exercise, showing that our filter compares favorably to more traditional methods.

1. INTRODUCTION

Long-standing interest in asset price volatility, combined with recent developments in its estimation with high-frequency data, has provoked research on the correct use of such data. In this paper we offer a framework for high-frequency measurement of asset returns that provides a means of clarifying

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 211–227

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20008-7

the impact of microstructure noise. Additionally, we provide Kalman filter-based techniques for the efficient removal of such noise.

In a series of widely cited articles, Andersen, Bollerslev, Diebold, and Ebens (2001) and Barndorff-Nielsen and Shephard (2002a,b) lay out a theory of volatility estimation from high-frequency sample variances. According to the theory, realized volatility estimators can recover the volatility defined by the quadratic variation of the semimartingale for prices. Realized volatility estimators are constructed as the sums of squared returns, where each return is measured over a short interval of time.[1]

Realized volatility differs markedly from model-based estimation of volatility. The widely used class of volatility models derived from the ARCH specification of Engle (1982), place constraints on the parameters that correspond to the interval over which returns are measured. Empirical analyses of these models rarely support the constraints. In contrast, realized volatility estimators do not require a specified volatility model.

The asymptotic theory underpinning realized volatility estimators suggests that the estimators should be constructed from the highest frequency data available. One would then sum the squares of these high-frequency returns, giving each squared return equal weight. In practice, however, very high-frequency data is contaminated by noise arising from the microstructure of asset markets.

By now, it is widely accepted that market microstructure contamination obscures high-frequency returns through several channels. For example, transaction returns exhibit negative serial correlation due to what Roll (1984) terms the bid-ask bounce. When prices are observed at only regular intervals, or are treated as if this were the case, measured returns exhibit nonsynchronous trading biases as described in Cohen, Maier, Schwartz, and Whitcomb (1978, 1979), and Atchison, Butler, and Simonds (1987), and Lo and MacKinlay (1988, 1990). Because transaction prices are discrete and tend to cluster at certain fractional values, prices exhibit rounding distortions as described in Gottlieb and Kalay (1985), Ball (1988), and Cho and Frees (1988). Noise cannot be removed simply by working with the middle of specialist quotes; while mid-quotes are less impacted by asynchronous trade and the bid-ask bounce, mid-quotes are distorted by the inventory needs of specialists and by the regulatory requirements that they face.[2]

Simulations by Andersen and Bollerslev (1998) and Andreou and Ghysels (2002), among many others, illustrate the effects of finite sampling and microstructure noise on volatility estimates under a variety of specifications. Differences in model formulation and assumed frictions make drawing robust conclusions about the effects of specific microstructure features difficult. Nevertheless, from the cited work, it is clear that microstructure frictions, as a group, cannot be safely ignored.

Essentially, three strands of research exist that treat the problem of microstructure noise in realized volatility estimation. The first attempts to remove the noise with a simple moving-average filter as in Zhou (1996). Andersen, Bollerslev, Diebold, and Ebens (2001) and Corsi, Zumbach, Müller, and Dacorogna (2001) select a sample frequency of five minutes based on a volatility signature plots, and then apply a moving-average filter. In contrast, Russell and Bandi (2004) work with an explicit model of microstructure noise. Rather than filtering the data to reduce the noise, they determine an optimal sampling frequency in the presence of noise. To do so, they construct a mean-squared error criterion that trades off the increase in precision against the corresponding increase in noise that arises as the sampling frequency increases. Although squared returns are given equal weight for a given asset, the optimal sampling interval that arises can vary across assets. Ait-Sahalia, Mykland, and Zhang (2003) and Oomen (2004) offer similar treatments. Finally, Hansen and Lunde (2004) derive a Newey and West (1987) type correction to account for spurious correlations in observed returns.

Theory suggests that noise volatility remains relatively constant. However, it is known that return volatility varies markedly. Thus, the relative contributions of noise and actual returns toward observed returns vary. During periods of high-return volatility, return innovations tend to dominate the noise in size. In consequence, we propose a somewhat different estimator in which the weight given to each return varies. Observed returns during periods of high volatility are given larger weight.

Our argument has three parts. First, we frame a precise definition of noise in terms of market microstructure theory. Second, we show how the Kalman filter can be used to remove the microstructure noise. We pay particular attention to how the variability of the optimal return weights depends on high-frequency volatility. Third, we demonstrate the efficacy of the filter in removing the noise.

2. MODEL

To formalize, consider a sequence of fixed intervals (five-minute periods, for example) indexed by t. The log of the observed price at t is $\tilde{p}_t = p_t + \eta_t$, where p_t denotes true price and η_t denotes microstructure noise. The observed return is

$$\tilde{r}_t = r_t + \varepsilon_t \tag{1}$$

where $\tilde{r}_t = p_t - p_{t-1}$ is the latent (true) return and $\varepsilon_t = \eta_t - \eta_{t-1}$ is the return noise.

We employ assumptions typical of the realized volatility literature. Our first assumption concerns the latent (true) price process.

Assumption 1. (*Latent price process*). The log true price process is a continuous local martingale. Specifically,

$$p_\tau = \int_0^\tau v_s \mathrm{d}w_s$$

where w_s is standard Brownian motion and the spot volatility process, v_s is a strictly positive cadlag process such that the quadratic variation (or integrated volatility) process, V_τ, obeys

$$V_\tau = \int_0^\tau v_s \mathrm{d}s < \infty$$

with probability one for all τ.

Assumptions about noise dynamics must be selected with care. Close study of microstructure noise reveals strong positive correlation at high frequency.[3] The correlation declines sharply with the sampling frequency, due to intervening transactions. To understand these effects, we discuss three prominent sources of noise.

The bid–ask bounce, discussed by Roll (1984), arises because transactions cluster at quotes rather than the true price. Hasbrouck and Ho (1987) show that this source of noise may be positively correlated as a result of clustered trade at one quote (due to the break up of large block trades). However, the positive noise correlation due to trade clustering nearly vanishes between trades more than a few transactions apart. In a similar fashion, positive noise correlation arising from the common rounding of adjacent transactions, vanishes at lower sampling frequencies.

The nonsynchronous trading effect, discussed by Lo and MacKinlay (1990), arises when transactions are relatively infrequent. If transactions are infrequent relative to the measurement of prices at regular intervals, then multiple price measurements refer to the same transaction, inducing positive noise correlation. Again, the positive noise correlation vanishes as the sampling frequency declines.

To determine the sampling frequency at which noise is uncorrelated, Hasbrouck and Ho (1987) study a large sample of NYSE stocks. They find no significant correlation for observations sampled more than 10 transactions apart. Hansen and Lunde (2004) find supporting evidence in their recent study of Dow Jones Industrial Average stocks. In consequence, we assume that the sampling interval contains at least 10 transactions to justify treating microstructure noise as an i.i.d. sequence.

Assumption 2. (*Microstructure noise*). The microstructure noise forms an i.i.d. sequence of random variables each with mean zero and variance $\sigma_\eta^2 < \infty$ and independent of the latent return process.

We do not make any distributional assumptions about microstructure noise. However, as the noise is composed of a sum of several largely independent features, and because these features tend to be symmetric, the assumption of normally distributed noise may be a plausible approximation. Consequently, we consider normally distributed microstructure noise in Section 3.

Under Assumption 2, it is clear that return noise forms an MA(1) process with a unit root. To determine the covariance structure of observed returns, we assume that latent returns form a weakly stationary martingale.

Lemma 1. If in addition to Assumption 1 and Assumption 2, r_t forms a weakly stationary process with unconditional mean zero and unconditional variance σ_r^2, then the autocovariance function of observed returns obeys

$$Cov(\tilde{r}_t, \tilde{r}_{t-k}) = \begin{cases} \sigma_r^2 + 2\sigma_\eta^2 & \text{if} \quad k = 0 \\ -\sigma_\eta^2 & \text{if} \quad k = 1 \\ 0 & \text{if} \quad k > 1 \end{cases}$$

Moreover, the first-order autocorrelation is given by

$$\rho = -\frac{\sigma_\eta^2}{\sigma_r^2 + 2\sigma_\eta^2}$$

For each day, which contains n intervals, define the following three volatility measures.

1. The integrated volatility, $V = \Sigma_{t=1}^n \sigma_t^2$.
2. An infeasible estimator, constructed from latent returns, $\overline{V} = \Sigma_{t=1}^n r_t^2$.
3. A feasible estimator, $\hat{V} = \Sigma_{t=1}^n \hat{r}_t^2$, where $\hat{r}_t = \tilde{r}_t$ in the absence of noise.

To motivate the form of the feasible estimator, decompose the estimation error as

$$V - \hat{V} = V - \overline{V} + \overline{V} - \hat{V} \tag{2}$$

The behavior of $V - \overline{V}$ as a function of step length and the underlying volatility process has been studied by Barndorff-Nielsen and Shephard (2002a). If the step length is chosen (hence n is fixed), then this part of the error is beyond the control of the researcher. Therefore, we focus on minimizing the mean squared error $\mathrm{E}\left(\overline{V} - \hat{V}\right)^2$, where $\tilde{r}' = (\tilde{r}_t, \cdots, \tilde{r}_T)$ and $T = n \cdot J$ (J is the number of days in the sample). It is well known that the mean squared error is minimized by choosing

$$\hat{V} = \mathrm{E}\left(\overline{V}|\tilde{r}\right) = \mathrm{E}\left(\sum_{t=1}^{n} r_t^2 \middle| \tilde{r}\right) = \sum_{t=1}^{n} \mathrm{E}\left(r_t^2|\tilde{r}\right)$$

Thus, in order to minimize the effects of microstructure noise, we must extract expected squared latent returns from observed returns. The effectiveness with which the extraction can be achieved depends on the correct treatment of the microstructure noise.

2.1. Kalman Filter and Smoother

The Kalman filter provides a technique to separate (observed) contaminated returns into two components: the first corresponds to (latent) true returns and the second to microstructure noise. To construct the filter, we follow the notation in Hamilton (1994) (Harvey, 1989 also provides textbook treatment). The state vector consists of latent variables, $\xi_t' = \left(r_t, \eta_t, \eta_{t-1}\right)$. The observation equation relates the state vector to observed returns

$$\tilde{r}_t = \mathbf{H}'\xi_t \tag{3}$$

where $H' = (1,\ 1,\ -1)$. The state equation describes the dynamic evolution of the state vector

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{R}\mathbf{v}_{t+1} \tag{4}$$

where $v_t' = \left(r_t, \eta_t\right)$ with covariance matrix

$$Q_t = \begin{pmatrix} \sigma_t^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix}$$

and the coefficient matrices are

$$\mathbf{F} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{R} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

The Kalman filter delivers the linear projection of the state vector ξ_t, given the sequence of observations $\{\tilde{r}_1, \cdots, \tilde{r}_t\}$. The Kalman smoother delivers the corresponding linear projection onto the extended sequence $\tilde{r}$. While these linear projections make efficient use of information, they may have larger MSE's than nonlinear projections. As conditional expectations need not be linear projections, we distinguish between linear projections and conditional expectations. Let $\hat{\mathrm{E}}_t$ represent linear projection onto $F_t = \{\tilde{r}_t, \tilde{r}_{t-1}, \ldots, \tilde{r}_1, 1\}$. Let $\hat{\xi}_{\tau|t} = \hat{\mathrm{E}}_t(\xi_\tau)$ and let

$$\mathbf{P}_{\tau|t} = \mathrm{E}\left[\left(\xi_\tau - \hat{\xi}_{\tau|t}\right)\left(\xi_\tau - \hat{\xi}_{\tau|t}\right)'\right] \tag{5}$$

represent the mean squared error matrices of these projections. For example, the one-step-ahead mean squared error matrix $\mathbf{P}_{t|t-1}$ is a diagonal matrix with the first diagonal element equal to σ_t^2 and the second equal to σ_η^2. The third diagonal element we define as $c_t = Var\left(\hat{\eta}_{t-1|t-1}\right)$. The c_t are determined though a recursion described below. Let u_t denote the one-step-ahead prediction error for the observed returns. Then it follows that the variance of u_t, which we denote by M_t, is given by $M_t = H'P_{t|t-1}H = \sigma_t^2 + \sigma_\eta^2 + c_t$.

The projections from the Kalman filter are given by the recursion

$$\hat{r}_{t|t}\left(\sigma_t^2\right) = \frac{\sigma_t^2}{M_t}\left(\tilde{r}_t + \hat{\eta}_{t-1|t-1}\right) \tag{6}$$

$$\hat{\eta}_{t|t} = \frac{\sigma_\eta^2}{M_t}\left(\tilde{r}_t + \hat{\eta}_{t-1|t-1}\right) \tag{7}$$

$$c_{t+1} = \frac{\sigma_\eta^2\left(\sigma_t^2 + c_t\right)}{M_t} \tag{8}$$

The recursion and the boundary conditions $\hat{\eta}_{t|t} = 0$ and $c_1 = \sigma_\eta^2$ determine the sequence of filtered returns and filtered noise.

The projections from the Kalman smoother, which are the elements of $\hat{\xi}_{\tau|T} = \hat{\mathrm{E}}_T(\xi_\tau)$, are

$$\hat{r}_{t|T}\left(\sigma_t^2\right) = \hat{r}_{t|t}\left(\sigma_t^2\right) - \frac{\sigma_t^2}{\sigma_t^2 + c_t}\left(\hat{\eta}_{t|T} - \hat{\eta}_{t|t}\right) \tag{9}$$

$$\hat{\eta}_{t-1|T} = \hat{\eta}_{t-1|t} + \frac{c_t}{\sigma_t^2 + c_t}\left(\hat{\eta}_{t|T} - \hat{\eta}_{t|t}\right) \tag{10}$$

Smoothed quantities exhibit smaller variances than their filtered counterparts. For example, if we let $d_{t+1} = Var\left(\hat{\eta}_{t|T}\right)$, then it can be shown that

$$d_t = c_t\left(\frac{\sigma_t^2 + \sigma_\eta^2}{\sigma_t^2 + \sigma_\eta^2 + c_t}\right) - \left(\frac{c_t}{\sigma_t^2 + c_t}\right)^2 (c_{t+1} - d_{t+1}) \tag{11}$$

As is easily verified from their definitions, $d_{T+1} = c_{T+1}$. Consequently, $d_T < c_T$, and, by an induction argument, it follows that $d_t < c_t$ for $t = 1, 2, \ldots, T$, establishing that $Var\left(\hat{\eta}_{t|T}\right) < Var\left(\hat{\eta}_{t|t}\right)$ for $t = 1, 2, \ldots, T-1$.

The smoother estimates the latent returns as weighted averages of contemporaneous, lagged, and future observed returns. If variance is constant, so that $\sigma_t^2 = \sigma_r^2$ for all t, then the weights are nearly the same for all smoothed returns. To see the point clearly, consider a numerical example. If $\sigma_r^2 = 10$, $\sigma_\eta^2 = 1$, and $T = 7$, then, ignoring weights less than 0.001, we have that

$$\tilde{r}_{4|7} = 0.006\tilde{r}_2 + 0.0709\tilde{r}_3 + 0.8452\tilde{r}_4 + 0.0709\tilde{r}_5 + 0.006\tilde{r}_6$$

while

$$\tilde{r}_{3|7} = 0.0059\tilde{r}_1 + 0.0709\tilde{r}_2 + 0.8452\tilde{r}_3 + 0.0709\tilde{r}_4 + 0.006\tilde{r}_5.$$

Thus, an almost identical weighting scheme determines the third and fourth optimally estimated latent returns. For large samples, the weights are even more consistent. Except for a few returns at the beginning and end of the sample, the assumption of constant volatility leads to estimates of latent returns that are essentially a weighted average of the observed returns where the weights, for all practical purposes, are constants.

If, as is almost certainly the case in practice, latent returns do not exhibit constant volatility, then the optimal weights for estimating latent returns in (6)–(8) and (9)–(10) are not constant. Instead, during periods of high volatility, the optimal weights are larger for the currently observed return, and lower for the other returns.

2.2. Bias

With the estimated latent returns $\hat{r}_{t|T} = \hat{\mathrm{E}}(r_t|\tilde{r})$, it seems natural to estimate realized volatility by $\Sigma_{t=1}^{n} \hat{r}_{t|T}^2$ (note, $\hat{r}_{t|T}^2$ stands for $\left(\hat{r}_{t|T}\right)^2$). Yet, because filtering is a linear transformation, while squaring is not, $\hat{r}_t^2$ is a downward biased estimator of $\mathrm{E}\left(r_t^2|\tilde{r}\right)$. Fortunately, the size and direction of the bias is determined in the normal course of constructing the Kalman smoother. For the bias $\mathrm{E}\left(r_t^2 - \hat{r}_{t|T}^2\right)$, the properties of projection mappings imply[4]

$$\mathrm{E}\left[\left(r_t - \hat{r}_{t|T}^2\right)^2\right] = \mathrm{E}\left[\hat{\mathrm{E}}_T\left(r_t^2 - 2\hat{r}_{t|T} r_t + \hat{r}_{t|T}^2\right)\right] = \mathrm{E}\left[r_t^2 - \hat{r}_{t|T}^2\right] \tag{12}$$

Thus, the bias equals the (1, 1) element of the mean squared error matrix for $\hat{r}_{t|T}$.

We find that the bias is

$$b_t\left(\sigma_t^2\right) = b_t^f\left(\sigma_t^2\right) - \left(\frac{\sigma_t^2}{\sigma_t^2 + c_t}\right)^2 (c_{t+1} - d_{t+1}) \tag{13}$$

where $b_t^f\left(\sigma_t^2\right)$ (the bias of the filtered return $\hat{r}_{t|t}$) is

$$b_t^f\left(\sigma_t^2\right) = \sigma_t^2\left(\frac{\sigma_\eta^2 + c_t}{\sigma_t^2 + \sigma_\eta^2 + c_{t-1}}\right) \tag{14}$$

Recall that c_t is the element of the variance for the filtered prediction of η_{t-1} and d_t is the corresponding variance element for the smoothed prediction. As shown in (11), for $t < T$ the smoothed estimator has lower variance than the filtered estimator.[5] As a result, the bias of the smoothed estimator is smaller than the bias of the filtered estimator, and so we concentrate on the smoothed estimator in what follows.

To determine the magnitude of the bias, consider the simple case in which $\sigma_t^2 = \sigma_r^2$ (constant volatility). The bias, b_t, is well approximated by

$$\sigma_r^2\left(1 - \frac{1}{\sqrt{1 + 4\sigma_\eta^2/\sigma_r^2}}\right)$$

In accord with intuition, the bias is a decreasing function of the return variance and an increasing function of the noise variance. If $\sigma_\eta^2/\sigma_r^2 = .1$, then the bias is 15 percent of the return variance and 50 percent larger than the expected squared noise term.[6] If the noise variance dominates, so that $\sigma_r^2/\sigma_\eta^2 \simeq 0$, then the bias is approximately σ_r^2. If the return variance dominates, then the bias is near zero.

3. MULTIVARIATE NORMAL APPROACH

To analyze the multivariate normal case, it is convenient to work in vector form. Let $\mathbf{r} = (r_1, r_2, \cdots, r_T)'$ and $\eta = (\eta_0, \eta_1, \ldots, \eta_T)'$ so that

$$\tilde{\mathbf{r}} = \mathbf{r} + \mathbf{B}\eta$$

Here $\mathbf{B}$ is a selection matrix with first row $[-1, -1, 0, \cdots, 0]$. The covariance matrix of r is $\Lambda = diag(\sigma_t^2)$. The Kalman smoother equations (in vector form) are

$$\hat{\mathbf{r}} = \Lambda\left(\Lambda + \sigma_\eta^2\mathbf{B}\mathbf{B}'\right)^{-1}\tilde{\mathbf{r}} \text{ and } \Sigma = \sigma_\eta^2\mathbf{B}\left(\mathbf{I} + \sigma_\eta^2\mathbf{B}'\Lambda^{-1}\mathbf{B}\right)^{-1}\mathbf{B}' \tag{15}$$

From Assumption 1, it follows that $r_t|\sigma_t^2 \sim N(0, \sigma_t^2)$. If we extend the assumption to

$$\begin{pmatrix} \mathbf{r} \\ \eta \end{pmatrix} \sim N_T\left(0, \begin{pmatrix} \Lambda & 0 \\ 0 & \sigma_\eta^2\mathbf{I} \end{pmatrix}\right)$$

then it is simple to derive the conditional distribution of $\tilde{r}|r$. Specifically,

$$\mathbf{r}|\tilde{\mathbf{r}} \sim N_T(\hat{\mathbf{r}}, \Sigma) \tag{16}$$

where $\hat{\mathbf{r}}$ and Σ are identical to the quantites from the Kalman smoother (15).

Under the assumption of joint normality $\hat{r} = \mathrm{E}(r|\tilde{r})$, so, the smoothed estimator is the conditional expectation rather than simply the optimal linear projection. Similarly, $\Sigma = Var(r|\hat{r})$ rather than simply the MSE matrix of the linear projection. This is especially useful for understanding the source of the bias that arises from squaring filtered returns. Here

$$Var(r_t|\tilde{\mathbf{r}}) = \mathrm{E}\left(r_t^2|\tilde{\mathbf{r}}\right) - \mathrm{E}^2(r_t|\tilde{\mathbf{r}}) \tag{17}$$

The optimal estimator $\mathrm{E}\left(r_t^2|\tilde{\mathbf{r}}\right)$ exceeds the square of the optimal estimator for latent returns $\mathrm{E}^2(r_t|\tilde{\mathbf{r}})$. The correction term $Var(r_t|\tilde{\mathbf{r}})$ does more than simply correct for the bias. Because the correction term corresponds to the conditional covariance matrix of r given the observed returns, the correction delivers the conditional expectation of squared returns. In consequence, we are able to form an optimal nonlinear estimator from an optimal linear estimator as

$$\mathrm{E}\left(r_t^2|\tilde{\mathbf{r}}\right) = \mathrm{E}^2(r_t|\tilde{\mathbf{r}}) + Var(r_t|\tilde{\mathbf{r}})$$

4. IMPLEMENTATION

The above analysis, in which it is assumed that $\{\sigma_t^2\}$ and σ_η^2 are known, suggests the use of the bias-corrected estimator

$$\hat{V}(\sigma_t^2) = \sum_{t=1}^{n} \left[\hat{r}_{t|T}^2(\sigma_t^2) + b_t(\sigma_t^2)\right]$$

To implement the method, we need estimators of $\{\sigma_t^2\}$ and σ_η^2.

If the latent return variance is assumed constant, then $\sigma_t^2 = \sigma_r^2$ and the bias-corrected estimator $\hat{V}(\sigma_r^2)$ is a function only of σ_r^2 and σ_η^2. From Lemma 1, the first two autocovariances of the observed returns series are sufficient for determining the variance of the noise and the expected variance of the true returns. (If one wishes to make further distributional assumptions, then ML estimators may be used in place of the method of moments estimators.) Andersen et al. (2001) employ an MA(1) estimator that, while similar to $\hat{V}(\sigma_r^2)$, does not contain smoothed estimates and makes no bias correction.

For the case in which the return variances are not constant, we begin with $\hat{r}_{t|T}^2(\hat{\sigma}_r^2)$ and $b_t(\hat{\sigma}_r^2)$. We then estimate the time-varying variance with a rolling window.[7]

$$\hat{\sigma}_{t|T}^2 = \frac{1}{25} \sum_{k=t-12}^{t+12} \left(\hat{r}_{k|T}^2(\hat{\sigma}_r^2) + b_k(\hat{\sigma}_r^2)\right) \tag{18}$$

The estimated time-varying variances from (18) together with $\hat{\sigma}_\eta^2$, yield $\hat{r}_{t|T}^2(\hat{\sigma}_t^2)$ from (9) and $b_t(\hat{\sigma}_t^2)$ from (13).

For the case of constant variance, laws of large numbers ensure the consistency of $\hat{\sigma}_r^2$ and $\hat{\sigma}_\eta^2$. Similar results are derived for ML estimators in Ait-Sahalia et al. (2003). To establish consistency if the return variance is not constant, it seems natural to specify a dynamic structure for $\{\sigma_t^2\}$. Rather than focus on this problem, we seek to recover latent realized volatility with a general purpose filter that minimizes mean squared error.

5. PERFORMANCE

To test the performance of the suggested filter against realistic scenarios, we use a model for the simulated latent returns that is consistent with the return behavior of the S&P 500 stock index. A popular special case of Assumption 1 is

$$\begin{aligned} \mathrm{d}p_t &= \sigma_t \mathrm{d}w_t \\ \mathrm{d}\sigma_t^2 &= \theta(\omega - \sigma_t)\mathrm{d}t + (2\lambda\theta)^{1/2}\mathrm{d}w_{\sigma_t} \end{aligned} \tag{19}$$

where w_t and w_{σ_t} are independent Brownian motion. Drost and Werker (1996, Corollary 3.2) provide the map between (19) and the (discrete) GARCH(1,1),

$$\begin{aligned} p_t - p_{t-1/m} = r_{(m),t} &= \sigma_{(m),t}\, z_{(m),t} \\ \sigma_{(m),t}^2 &= \phi_{(m)} + \alpha_{(m)} r_{(m),t}^2 + \beta_{(m)} \sigma_{(m),t-1/m}^2 \end{aligned} \tag{20}$$

where $z_{(m),t}$ is (for the purposes of simulation) i.i.d. $N(0,1)$. Andreou and Ghysels (2002) find that 5-minute returns from the S&P 500 index are well approximated by the values

$$\phi_{(m)} = 0.0004, \ \ \alpha_{(m)} = 0.0037, \ \ \beta_{(m)} = 0.9963. \tag{21}$$

These parameters imply an unconditional return variance of $\sigma_r^2 = 7.9$ basis points over the 5-minute interval. While this unconditional variance is high (daily estimates of return variance are roughly eight basis points), an appropriate rescaling by multiplying by 1/78 results in such small parameter values that simulation is difficult. As the relative mean squared error measurements that we report are invariant to such scaling, we follow Andreou and Ghysels and use the values in (21).

From (20) and (21) we simulate latent returns, r_t. We construct observed returns as $\tilde{r}_t = r_t + \eta_t - \eta_{t-1}$, where η_t is generated as an i.i.d. $N\left(0, \sigma_\eta^2\right)$ random variable. To determine the noise variance, we invert the formula for ρ in Lemma 1 to obtain

$$\sigma_\eta^2 = \frac{-\rho}{1+2\rho}\sigma_r^2$$

Hasbrouck and Ho (1987) report estimates of ρ between $-.4$ and $-.1$, so we allow ρ to take the values $[-.4, -.3, -.2, -.1]$. As decreasing the value of ρ increases σ_η^2, the resultant values of noise variance vary from $\sigma_\eta^2 = 1$ (for $\rho = -.1$) to $\sigma_\eta^2 = 15.8$ (for $\rho = -.4$).

To mirror trading days on the NYSE, which are 6.5 hours long, each simulated day contains 78 5-minute returns. We generate 10,000 trading days, a span that roughly corresponds to 50 years. For each day, j, we construct the latent realized volatility

$$\overline{V} = \sum_{t=(j-1)78+1}^{78j} r_t^2,$$

the feasible bias-corrected realized volatility estimator

$$\hat{V}_j\left(\hat{\sigma}^2_{t|T}\right) = \sum_{t=(j-1)78+1}^{78j} \left[\hat{r}^2_{t|T}\left(\hat{\sigma}^2_{t|T}\right) + b_t\left(\hat{\sigma}^2_{t|T}\right)\right]$$

and the infeasible bias-corrected estimator $\hat{V}_j\left(\sigma^2_t\right)$.

To compare this filter to methods that assume constant return variance, such as the MA(1) filter mentioned above, we construct $\hat{V}_j(\hat{\sigma}^2_r)$. To determine the gains from smoothing, we also construct the estimator based on filtered (rather than smoothed) quantites

$$\hat{V}^f_j\left(\hat{\sigma}^2_{t|t}\right) = \sum_{t=(j-1)78+1}^{78j} \left[\hat{r}^2_{t|t}\left(\hat{\sigma}^2_{t|t}\right) + b^f_t\left(\hat{\sigma}^2_{t|t}\right)\right]$$

where $\hat{\sigma}^2_{t|t}$ is obtained from (18) with $\hat{r}^2_{k|t}\left(\hat{\sigma}^2_r\right)$ and $b^f_k\left(\hat{\sigma}^2_r\right)$ in place of $\hat{r}^2_{k|T}\left(\hat{\sigma}^2_r\right)$ and $b_k\left(\hat{\sigma}^2_r\right)$, respectively. Finally, for completeness, we construct $\hat{V}^f_j(\hat{\sigma}^2_r)$.

To judge the quality of the realized volatility estimators, we measure the mean squared error (MSE) of each estimator relative to the infeasible (optimal) estimator. For example, the relative MSE for $\hat{V}\left(\hat{\sigma}^2_{t|T}\right)$ is

$$\frac{MSE\left[\hat{V}\left(\hat{\sigma}^2_{t|T}\right)\right]}{MSE\left[\hat{V}\left(\sigma^2_t\right)\right]} = \frac{\sum_{j=1}^{10000}\left(\overline{V}_j - \hat{V}_j\left(\hat{\sigma}^2_{t|T}\right)\right)^2}{\sum_{j=1}^{10000}\left(\overline{V}_j - \hat{V}_j\left(\sigma^2_t\right)\right)^2}$$

In Table 1, we present the relative efficiency calculations. Regardless of the degree of noise variance, or indeed of the decision to smooth, the gain from estimating a time-varying return variance is substantial. For the case with the smallest noise variance, the relative MSE for the smoothed estimator is reduced by more than half (from 3.5 to 1.4). As one would expect,

Table 1. Relative Efficiency.

σ^2_η	Constant Variance		Time-Varying Variance	
	$\hat{V}^f(\hat{\sigma}^2_r)$	$\hat{V}(\hat{\sigma}^2_r)$	$\hat{V}^f(\hat{\sigma}^2_{t\|t})$	$\hat{V}(\hat{\sigma}^2_{t\|T})$
15.8	8.4	6.7	5.7	4.7
5.9	7.6	6.0	3.9	3.4
2.6	6.1	5.1	2.4	2.2
1.0	3.9	3.5	1.5	1.4

increasing the noise variance renders the estimation problem more difficult, yet even for the highest noise variance the relative MSE for the smoothed estimator is substantially reduced (from 6.7 to 4.7). Moreover, while smoothing always leads to an efficiency gain, the magnitude of the efficiency gain resulting from smoothing is dominated by efficiency gain from allowing for time-varying volatility.

To determine the impact of diurnal patterns, we generate time-varying volatility that mirrors the U-shape pattern often observed in empirical returns. To do so, we construct a new sequence of return variances $\left\{\sigma^{2*}_{(m),t}\right\}$:

$$\sigma^{2*}_{(m),t} = \sigma^{2}_{(m),t}\left(1 + 1/3\cos\left(\frac{2\pi}{78}t\right)\right)$$

where $\sigma^{2}_{(m),t}$ is obtained from (20). Note that with the cyclic component, the expected variance doubles between the diurnal peak and trough. This process mimics the U-shape pattern as the maximum of the cosine term to corresponds to the beginning and ending of each day.

In Table 2, we find that the relative MSE measurements are surprisingly robust to the presence of diurnal patterns. When noise variance is about an order of magnitude smaller than the expected innovation variance (when $\sigma^2_\eta = 1.0$), the MSE of the realized volatility estimator is about 4 percent larger when based on filtered returns. When noise variance is roughly twice as large as the expected innovation variance (when $\sigma^2_\eta = 15.8$), the filter-based mean squared error is about 10 percent larger. Larger gains are achieved by the estimator based on the rolling volatility proxy, especially when noise volatility is relatively small. The mean squared errors based on the naive estimators are between 40 and 160 percent larger than corresponding mean squared errors based on the volatility proxy. The improvements from smoothing, relative to filtering, are shown in the last column.

Table 2. Relative Efficiency with a Diurnal Pattern.

σ^2_η	Constant Variance		Time-Varying Variance	
	$\hat{V}^f(\hat{\sigma}^2_r)$	$\hat{V}(\hat{\sigma}^2_r)$	$\hat{V}^f(\hat{\sigma}^2_{t\|t})$	$\hat{V}(\hat{\sigma}^2_{t\|T})$
15.8	8.1	6.5	5.5	4.6
5.9	7.4	5.9	3.8	3.3
2.6	6.0	5.0	2.4	2.2
1.0	3.9	3.5	1.5	1.4

Although currently used filters vary widely, we are aware of none that exploit the gains available from either smoothing or from the used of a high-frequency volatility proxy. Most filtering methods in uses are similar to the filtered naïve estimator. Notice that the mean squared errors of the filtered naïve estimators are more than double those of the smoothed estimators based on our proposed smoothed estimator based on the volatility proxy.

6. CONCLUSIONS

This article applies market microstructure theory to the problem of removing noise from a popular volatility estimate. The theory suggests that a Kalman smoother can optimally extract the latent squared returns, which are required for determining realized volatility from their noisy observable counterparts. However, the correct specification of the filter requires knowledge of a latent stochastic volatility state variable, and is therefore infeasible. We show that a feasible Kalman smoothing algorithm based on a simple rolling regression proxy for high-frequency volatility can improve realized volatility estimates. In simulations, the algorithm substantially reduces the mean squared error of realized volatility estimators even in the presence of strong diurnal patterns. The broad conclusion is that realized volatility estimators can be improved in an obvious way, by smoothing instead of merely filtering the data, and in a less obvious way, by bias correcting and using a straightforward proxy of latent high-frequency volatility.

NOTES

1. Andersen, Bollerslev, and Diebold (2005) provides a survey of both theory and empirics for realized volatility.
2. Surveys by O'Hara (1995), Hasbrouck (1996), Campbell, Lo, & MacKinlay (1997), and Madhavan (2000) document these and other microstructure frictions.
3. Noise outcomes of adjacent price measurements are almost perfectly correlated when no transaction intervenes (they are not perfectly correlated because, although measured price remains constant in the absence of new transactions, the latent true price changes through time).
4. Brockwell and Davis (1987, Proposition 2.3.2).
5. If $t = T$, then the smoothed estimator is identical to the filtered estimator.
6. The bias of the filtered estimator is approximately $-2\rho\sigma_r^2$ (recall $\rho < 0$).
7. The rolling window width of 24 corresponds to two hours, which balances bias and variance in the presence of diurnal features.

REFERENCES

Ait-Sahalia, Y., Mykland, P. & Zhang, L. (2003). *How often to sample a continuous-time process in the presence of market microstructure noise*. Working Paper w9611, NBER.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*, 885–905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2005). Parametric and nonparametric volatility measurement. In: Y. Aït-Sahalia & L. P. Hansen (Eds), *Handbook of financial econometrics*. Amsterdam: Elsevier Science B.V.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*, 43–76.

Andreou, E., & Ghysels, E. (2002). Rolling sample volatility estimators: Some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics*, *20*, 363–376.

Atchison, M., Butler, K., & Simonds, R. (1987). Nonsynchronous security trading and market index autocorrelation. *Journal of Finance*, *42*, 111–118.

Ball, C. (1988). Estimation bias induced by discrete security prices. *Journal of Finance*, *43*, 841–865.

Barndorff-Nielsen, O. E. & Shephard, N. (2002a). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, Series B 64, 253–280.

Barndorff-Nielsen, O. E., & Shephard, N. (2002b). Estimating quadratic variation using realised variance. *Journal of Applied Econometrics*, *17*, 457–477.

Brockwell, P., & Davis, R. (1987). *Time series: Theory and methods*. New York: Springer.

Campbell, A., Lo, W., & MacKinlay, A. (1997). *The econometrics of financial markets*. Princeton: Princeton University Press.

Cho, D., & Frees, E. (1988). Estimating the volatility of discrete stock prices. *Journal of Finance*, *43*, 451–466.

Cohen, K., Maier, S., Schwartz, R., & Whitcomb, D. (1978). The returns generation process, returns variance, and the effect of thinness in securities markets. *Journal of Finance*, *33*, 149–167.

Cohen, K., Maier, S., Schwartz, R., & Whitcomb, D. (1979). On the existence of serial correlation in an efficient securities market. *TIMS Studies in the Management Sciences*, *11*, 151–168.

Corsi, F., Zumbach, G., Müller, U., & Dacorogna, M. M. (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes*, *30*, 183–204.

Drost, F., & Werker, B. (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics*, *74*, 31–57.

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1007.

Gottlieb, G., & Kalay, A. (1985). Implications of the discreteness of observed stock prices. *Journal of Finance*, *40*, 135–154.

Hamilton, J. (1994). *Time series analysis*. Princeton: Princeton University Press.

Hansen, P., & Lunde, A. (2004). *An unbiased measure of realized variance*. Manuscript. Department of Economics, Brown University.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge: Cambridge University Press.

Hasbrouck, J. (1996). Modeling market microstructure time series. In: G. S. Maddala & C. R. Rao (Eds), *Handbook of Statistics*, (Vol. 14, pp. 647–692). Amsterdam: Elsevier-North-Holland.

Hasbrouck, J., & Ho, T. (1987). Order arrival, quote behaviour, and the return-generating process. *Journal of Finance*, *42*, 1035–1048.

Lo, A., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, *1*, 41–66.

Lo, A., & MacKinlay, A. C. (1990). An econometric analysis of nonsynchronous-trading. *Journal of Econometrics*, *45*, 181–212.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, *3*, 205–258.

Newey, W., & West, K. (1987). A simple positive semi-definite, heteroskedasticity and auto-correlation consistent covariance matrix. *Econometrica*, *55*, 703–708.

O'Hara, M. (1995). *Market microstructure theory*. Cambridge: Blackwell Publishers.

Oomen, R. (2004). *Properties of bias corrected realized variance in calendar time and business time*. Manuscript. Warwick Business School.

Roll, R. (1984). A simple measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, *39*, 1127–1139.

Russell, J., & Bandi, F. (2004). *Microstructure Noise, Realized Volatility, and Optimal Sampling*. Econometric Society 2004 Latin American Meetings 220, Econometric Society.

Zhou, B. (1996). High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, *14*, 45–52.

This page intentionally left blank

PART III: UNIVARIATE VOLATILITY MODELS

This page intentionally left blank

MODELING THE ASYMMETRY OF STOCK MOVEMENTS USING PRICE RANGES

Ray Y. Chou

ABSTRACT

It is shown in Chou (2005). Journal of Money, Credit and Banking, 37, 561–582that the range can be used as a measure of volatility and the conditional autoregressive range (CARR) model performs better than generalized auto regressive conditional heteroskedasticity (GARCH) in forecasting volatilities of S&P 500 stock index. In this paper, we allow separate dynamic structures for the upward and downward ranges of asset prices to account for asymmetric behaviors in the financial market. The types of asymmetry include the trending behavior, weekday seasonality, interaction of the first two conditional moments via leverage effects, risk premiums, and volatility feedbacks. The return of the open to the max of the period is used as a measure of the upward and the downward range is defined likewise. We use the quasi-maximum likelihood estimation (QMLE) for parameter estimation. Empirical results using S&P 500 daily and weekly frequencies provide consistent evidences supporting the asymmetry in the US stock market over the period 1962/01/01–2000/08/25. The asymmetric range model also provides sharper volatility forecasts than the symmetric range model.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 231–257

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20009-9

1. INTRODUCTION

It's known for a long time in statistics that range is a viable measure of the variability of random variables. In the recent two decades, applications to finance issues discovered that ranges were useful to construct efficient volatility estimators; e.g., see Parkinson (1980); Garman and Klass (1980); Beckers (1983); Wiggins (1991); Rogers and Satchell(1991); Kunitomo (1992); Rogers (1998); Gallant, Hsu, and Tauchen (1999); Yang and Zhang (2000); and Alizadeh, Brandt, and Diebold (2002). In Chou (2005), we propose the conditional autoregressive range (CARR) model for range as an alternative to the modeling of financial volatilities. It is shown both theoretically and empirically that CARR models are worthy candidates in volatility modeling in comparison with the existing methodologies, say the generalized auto regressive conditional heteroskedasticity (GARCH) models. Empirically, the CARR model performs very satisfactory in forecasting volatilities of S&P 500 using daily and weekly observations. In all four cases with different measures of the "observed volatility", CARR dominates GARCH in the Mincer/Markovtz regression of forecasting evaluations. It's a puzzle (see Cox & Rubinstein, 1985) that despite the elegant theory and the support of simulation results, the range estimator has performed poorly in empirical studies. In Chou (2005), we argue that the failure of all the range-based models in the literature is due to its ignorance of the temporal movements of the range. Using a proper dynamic structure for the conditional expectation of range, the CARR model successfully resolves this puzzle and retains its superiority in empirical forecasting powers.

This paper focuses on an important feature in financial data: asymmetry. Conventionally, symmetric distributions are usually assumed in asset pricing models, e.g., normal distributions in CAPM and the Black/Sholes option pricing formula. Furthermore, in calculating various measures of risk, standard deviations (or equivalently, variances) are used frequently, which implicitly assume a symmetric structure of the prices. However, there are good reasons why the prices of speculative assets should behave asymmetrically. For investors, the more relevant risk is generated by the downward price moves rather than the upward price moves; the latter is important in generating the expected returns. For example, the consideration of the value-at-risk only utilizes the lower tail of the return distribution. There are also models of asset prices that utilize the third moment (an asymmetric characteristic feature), for example, Levy and Markowitz (1979). Furthermore, asymmetry can arise in a dynamic setting in models considering time-varying conditional moments. For example, the ARCH-M model of Engle, Lilien, and Robbins (1987) posits a linkage between the first sample moment and past second sample moments. This model has a theoretical interpretation in finance: the risk premium hypothesis (See Malkiel, 1978; Pindyck, 1984; Poterba & Summers, 1986; Chou, 1988). The celebrated leverage effect of Black (1976) and Christie (1982) is cast into a dynamic volatility model in the form of the linkage between the second sample moment and past first sample moments; See EGARCH of Nelson (1991) and NGARCH of Engle and Ng (1993). Furthermore, the asymmetry can arise in other forms such as the volatility feedback of Campbell (1997). Barberis and Huang (2000) give an example of loss aversion and mental account that would predict an asymmetric structure in the price movements. Tsay (2000) uses only observations of the downward, extreme movements in stock prices to model the crash probability.

Chou (2005) incorporated one form of asymmetry, the leverage effect, into the CARR model and it appeared to be more significant than reported in the literature of GARCH or Stochastic Volatility models. The nature of the CARR model is symmetric because range is used in modeling which treats the maximum and minimum symmetrically. In this paper, a more general form of asymmetry is considered by allowing the dynamic structure of the upward price movements to be different from that of the downward price movements. In other words, the maximum and the minimum of price movements in fixed intervals are treated in separate forms. It may be relevant to suspect that the information in the downward price movements are as relevant as the upward price movements in predicting the upward price movements in the future. Similarly, the opposite case is true. Hence it is worthy to model the CARR model asymmetrically.

The paper is organized as following. It proposes and develops the Asymmetric CARR (ACARR) model with theoretical discussions in section 2. In addition, discussions are given about some immediate natural extensions of the ACARR model. An empirical example is given in section 3 using the S&P 500 daily index. Section 4 concludes with considerations of future extensions.

2. MODEL SPECIFICATION, ESTIMATION, AND PROPERTIES

2.1. The Model Specification, Stochastic Volatilities, and the Range

Let P_t be the logarithmic price of a speculative asset observed at time t, $t = 1, 2, \ldots T$. P_t is a realization of a price process $\{P_t\}$, which is assumed to be a continuous processf.[1] We further assume that within each time interval, we observe P_t at every fixed time interval dt. Let n denote the number of intervals between each unit time, then $dt = 1/n$. There are hence, $n+1$ observations within each time interval between $t-1$ and t. Let P_t^o, P_t^c, P_t^{HIGH}, P_t^{LOW}, be the opening, closing, high and low prices, in natural logarithm, between $t-1$ and t. The closing price at time t will be identical to the opening price at time $t+1$ in considerations of markets that are operated continuously, say, some of the foreign exchange markets. Further, define UPR_t, the upward range, and $DWNR_t$, the downward range as the differences between the daily highs, daily lows, and the opening price respectively, at time t, in other words,

$$\begin{aligned} UPR_t &= P_t^{\text{HIGH}} - P_t^o \\ DWNR_t &= P_t^{\text{LOW}} - P_t^o \end{aligned} \tag{2.1}$$

Note that these two variables, UPR_t and $DWNR_t$, represent the maximum and the minimum returns respectively, over the unit time interval ($t-1$, t). This is related to the range variable in Chou (2005) that R_t, defined to be

$$R_t = P_t^{\text{HIGH}} - P_t^{\text{LOW}} \tag{2.2}$$

It's clear that the range is also the difference between the two variables, UPR_t, and $DWNR_t$, in other words,

$$R_t = UPR_t - DWNR_t \tag{2.3}$$

In Chou (2005) we propose a dynamic model, the CARR model, for the range. It's a conjecture that the extreme value theory can be used to show that the conditional range, or equivalently the disturbance term, has a limiting distribution that is governed by a shifted Brownian bridge on the unit interval.[2] In this paper, we propose a model for the one-sided range, UPR_t, and $DWNR_t$, to follow a similar dynamic structure. In particular,

$$\begin{aligned} &UPR_t = \lambda_t^u \varepsilon_t^u \\ &DWNR_t = -\lambda_t^d \varepsilon_t^d \\ &\lambda_t^u = \omega^u + \sum_{i=1}^{p} \alpha_i^u UPR_{t-i} + \sum_{j=1}^{q} \beta_j^u \lambda_{t-j}^u \\ &\lambda_t^d = \omega^d + \sum_{i=1}^{p} \alpha_i^d DWNR_{t-i} + \sum_{j=1}^{q} \beta_j^d \lambda_{t-j}^d \\ &\varepsilon_t^u \sim iid f^u(\cdot),\ \varepsilon_t^d \sim iid f^d(\cdot) \end{aligned} \tag{2.4}$$

Model (2.4) is called the asymmetric conditional autoregressive range (Asymmetric CARR or ACARR, henceforth) model. In the following discussions, we will disregard the super-scripts when there is no concern of confusion. In (2.4), λ_t is the conditional mean of the one-sided range based on all information up to time t. The distribution of the disturbance terms ε_t of the normalized one-sided-range, or $OSR_t(= UPR_t$ or $DWNR_t,)$, $\varepsilon_t = OSR_t/\lambda_t$, are assumed to be identically independent with density function $f^i(\cdot)$, where $i = u$ or d. Given that both the one-sided ranges UPR_t and $-DWNR_t$, and their expected values λ_t are both positive hence their disturbances ε_t, the ratio of the two, are also positively valued.

The asymmetric behavior between the market up and down movements can be characterized by different values for the pairs of parameters, (ω^u, ω^d), (α^u, α^d), (β^u, β^d), and from the error distributions $(f^u(\cdot), f^d(\cdot))$.

The equations specifying the dynamic structures for λ_t's characterize the persistence of shocks to the one-sided range of speculative prices or what is usually known as the volatility clustering. The parameters ω, α_i, β_j, characterize respectively, the inherent uncertainty in range, the short-term impact effect and the long-term effect of shocks to the range (or the volatility of return). The sum of the parameters $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j$, plays a role in determining the persistence of range shocks. See Bollerslev (1986) for a discussion of the parameters in the context of GARCH.

The model is called an asymmetric conditional autoregressive range model of order (p,q), or ACARR(p,q). For the process to be stationary, we require that the characteristic roots of the polynomial to be out side the unit circle, or $\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1$. The long-term range denoted ω-bar, is calculated as $\omega/(1 - (\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j))$. Further, all the parameters in the second equation, are assumed positive, i.e., ω, α_i, $\beta_j > 0$.

It is useful to compare this model with the CARR model of Chou (2005):

$$
\begin{aligned}
&R_t = \lambda_t \varepsilon_t \\
&\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} \\
&\varepsilon_t \sim iid f(.)
\end{aligned}
\tag{2.5}
$$

Ignoring the distribution functions, the ACARR model reduces to the CARR if all the parameters with superscript u and d are identical pair-wise. Testing these various types of model asymmetry will be of interest because asymmetry can arise in varieties, e.g., size of the range, i.e., level of the volatility $(\omega - bar = \omega/(1 - \alpha - \beta))$, the speed of mean-reversion $(\alpha + \beta)$, and the short-term (α) versus long-term (β) impact of shocks.

Eq. (2.4) is a reduced form for the one-sided ranges. It is straight forward to consider extending the model to include other explanatory variables, $X_{t-1,l}$ that are measurable with respect to the information set up to time $t-1$.

$$\lambda_t = \omega + \sum_{i=1}^{p} \alpha_i R_{t-i} + \sum_{j=1}^{q} \beta_j \lambda_{t-j} + \sum_{l=1}^{L} \gamma_l X_{t-1,l} \tag{2.6}$$

This model is called the ACARR model with exogenous variables, or ACARRX. Among others, some important exogenous variables are trading volume (see Lamoureux & Lastrapes, 1990; Karpoff, 1987), the lagged returns measuring the leverage effect of Christie (1982), Black (1976) and Nelson (1990) and some seasonal factor to characterize the seasonal pattern within the range interval.

Note that although we have not specified specifically, all the variables and parameters in (2.4) are all dependent on the parameter n, the number of intervals used in measuring the price within each range-measured interval. It is clear that all the range estimates are downward biased if we assume the true data-generating mechanism is continuous or if the sampling frequency is lower than that of the data generating process if the price is discrete. The bias of the size of the one-sided-range, whether upward range (UPR_t) or downward range ($DWNR_t$), like the total range, will be a a non-increasing function of n. Namely, the finer the sampling interval of the price path, the more accurate the measured ranges will be.

It is possible that the highest frequency of the price data is non-constant given the heterogeneity in the trading activities within each day and given the nature of the transactions of speculative assets. See Engle and Russell (1998) for a detailed analysis of the non-constancy of the trading intervals, or the durations. Extensions to the analyses of the ranges of non-fixed interval prices will be an interesting subject for future research.[3] However, some recent literature suggest that it is not desirable to work with the transaction data in estimating the price volatility given the consideration of microstructures such as the bid/ask bounces, the intra-daily seasonality, among others. See Andersen, Bollerslev, Diebold, and Labys, (2000); Bai, Russell, and Tiao (2000), and Chen, Russell, and Tsay (2000).

As is the case for the CARR model, the ACARR model mimics the ACD model of Engle and Russell (1998) for durations between trades. Nonetheless, there are important distinctions between the two models. First, duration is measured at some random intervals but the range is measured at fixed intervals, hence the natures of the variables of interest are different although they share the common property that all observations are positively valued.

Second, in the ACD model, the distribution of the disturbances is usually chosen arbitrarily – a feature also shared by all GARCH models. The ACARR model, on the contrary, has some natural choices from the results of extreme value theories in statistics.[4]

2.2. Properties of ACARR: Estimation and Relationships with Other Models

Given that the ACARR model has exactly the same form as the CARR model, all the statistical results in CARR apply to ACARR. Furthermore, the ACARR model has some unique properties of its own. We illustrate some of the important properties in this subsection. Given that the upward and the downward range evolutions are specified independently, the estimation can hence be performed separately. Further, consistent estimation of the parameters can be obtained by the quasi-maximum likelihood estimation (QMLE) method. The consistency property follows from the ACD model of Engle and Russell (1998) and Chou (2005). It indicates that the exponential distribution can be used in constructing the likelihood to consistently estimate the parameters in the conditional mean equation.

Specifically, given the exponential distribution for the error terms, we can perform the QMLE. Using $R_{t,}\ t = 1, 2, \ldots, T$ as a general notation of UPR_t and $DWNR_t$, the log-likelihood function for each of the one-sided range series is

$$L(\alpha_i,\ \beta_j;\ R_1,\ R_2,\ \ldots,\ R_T,) = -\sum_{t=1}^{T}\left[\log(\lambda_t) + \frac{R_t}{\lambda_t}\right].$$

The intuition of this property relies on the insight that the likelihood function in ACARR with an exponential density is identical to the GARCH model with a normal density function with some simple adjustments on the specification of the conditional mean. Furthermore, all asymptotic properties of GARCH apply to ACARR. Given that ACARR is a model for the conditional mean, the regularity conditions (e.g., the moment condition) are in fact, less stringent then in GARCH.

Note that although QMLE is consistent, it is not efficient. The efficiency can be obtained if the conditional density function is known. This leads us to the limiting distribution of the conditional density of range. The discussion will require a far more complicated theoretic framework, which is worthy of pursuing by an independent work. We hence do not pursue this route in this paper and follow the strategy of Chou (2005) in relying on the QMLE.[5] Again, it is an empirical question as to how substantial in efficiency such methods can generate. Engle and Russell (1998) reported that deviations from the exponential density function do not offer efficiency gain sufficiently high in justifying the extra computation burdens.

It is important to note that the direct application of QMLE will not yield consistent estimates for the covariance matrix of the parameters. The standard errors of the parameters are consistently estimated by the robust method of Bollerslev and Wooldridge (1992). The efficiency issue related to these estimates is a subject for future investigation.

Another convenient property for ACARR (due to its connection with ACD) is the ease of estimation. Specifically, the QMLE estimation of the ACARR model can be obtained by estimating a GARCH model with a particular specification: specifying a GARCH model for the square root of range without a constant term in the mean equation.[6] This property is related to the above QMLE property by the observation of the equivalence of the likelihood functions of the exponential distribution in ACARR and ACD and of the normal density in GARCH. It indicates that it is almost effortless to estimate the ACARR model if a GARCH software is available.

It will be interesting and important to investigate whether the ACARR model will satisfy a closure property, namely, whether the ACARR process is invariant to temporal and cross-sectional aggregations. This is important given the fact that in financial economics, aggregates are frequently encountered, e.g., portfolios are cross-sectional aggregates and monthly, weekly returns are temporal aggregates of daily returns. It is also a property that is stressed in the literature of time series econometrics.[7]

Another interesting property of the CARR model is the encompassing property. It is interesting that the square-root-GARCH model turns out to be a special case of CARR, and in fact, the least efficient member of the CARR model. This property does not apply to ACARR since there are no analogies of the open to maximum (minimum) in the GARCH model family.

2.3. Robust ACARR

It is suggested in statistics that range is sensitive to outliers. It is useful hence, to consider extension of ACARR to address such considerations. We consider robust measures of range to replace the standard range defined as the difference between the max and the min. A simple naive method is to use the next-to-max for max and the next-to-min for min. By doing so, the chance of using outliers created by typing errors will be greatly reduced. It will also reduce the impact of some true outliers.

A second alternative is to use the quantile range, for example, a 90% quantile range is defined as the difference between the 95% percentile and the 5% percentile. A frequently adopted robust range is the interquartile range (IQR) which is a 75% quantitle and it can be conveniently obtained by taking the difference of the medians of the top and lower halves of the sampling data. In measuring a robust maximum or minimum likewise, we can use the 75% quantile in both the upward price distribution and the downward price distribution.

Similarly other types of robust extreme values can be adopted like the next-*i*th-to-max (min) and the average of the top 5% observations and the bottom 5% observations, et.al. There are several important issues relevant in considerations such as the efficiency loss, e.g., the IQR discards 50% of the information while the next-to-max approach discards very little. Another issue is the statistical tractability of the new range measures. For example, the quantile range will have a more complicated distribution than the range and the statistical property for the next-*i*th-to-extreme approach is less known than the quantile range. Another consideration is the data feasibility. In most cases, none of the information other than the extreme observations are available. For example, the standard data sources such as CRSP, and the Wall Street Journal, the Financial Times only report the daily highs and lows. As a result, the robust range estimators are infeasible unless one uses the intra-daily data. Nonetheless, the robust range estimators are feasible if the target volatility is measured at lower frequency than a day. This is obvious since there are 20 some daily observations available in each given month hence the monthly volatility can be measured by a robust range if the outlier problem is of concern. Given the existence of intra-daily data, daily robust range model is still an important topic for future research.

3. AN EMPIRICAL EXAMPLE USING THE S&P 500 DAILY INDEX, 1962/01/03–2000/08/25

3.1. The Data Set

The daily index data of the Standard and Poor 500 (S&P 500) are used for empirical study in this paper to gauge the effectiveness of the ACARR model. The data set is downloaded from the finance subdirectory of the website "Yahoo.com". The sample period covered in this paper is 1962/01/03–2000/08/25. The models are estimated by using this daily data set, comparisons are made for various volatility models on the accuracy of the volatility predictions.

Table 1 gives the summary statistics of RANGE, *UPR*, and *DWNR* for the full sample and two sub-sample periods. Sub-samples are considered because there is an apparent shift in the level of the daily ranges roughly on the date 1982/04/20. As is shown in Table 1, the averaged range level was reduced by almost a half since this particular date. Reductions in the level of similar magnitude are seen for the max and min as well. It's likely an institutional change occurred at the above-mentioned date. From a telephone conversation with the Standard and Poor Incorporated, the source of this structural change was revealed. Before this stated date, the index high and index low were compiled by aggregating the highs and lows of individual firm prices for each day. This amounts to assume that the highs and lows for all 500 companies occur at the same time in each day. This is clearly an incorrect assumption and amounts to an overestimate of the highs and an underestimate of the lows. As a result, the ranges are over-estimated. The compiling process was corrected after April 1982.[8] The company computes the index value at some fixed (unknown to me, say, 5 min) intervals within each day and than select the maximum and minimum price levels to be the index highs and index lows.[9]

Figs. 1–4 give the plots of the daily max and min price movements. It is interesting that the upward and downward range are roughly symmetric from the closeness of summary statistics and the seemingly reflective nature of Fig. 1. Another interesting observation (see Figs. 2–4) is that excluding the outlier of the 1987 crash, the two measures have very similar unconditional distributions. Careful inspection of the Figures and Tables however, reveals important differences in these two measures of market movements in the two opposite directions. For example, although both one-sided ranges (henceforth OSRs) have clustering behaviors but their extremely large values occur at different times and with different magnitudes. Further, as Table 1 and Fig. 5 show, the magnitudes of the autocorrelation at some lags for the UPR seem to be substantially different from that of the DWNR indicating different level of persistence. This can be viewed as a primitive indicator of the difference in the dynamic structure of the two processes. The true comparison of the dynamic structures of the two range processes will be made in the next section.

Table 1. Summary Statistics of the Daily Range, Upward Range, and Downward Range of S&P 500 Index, 1/2/1962–8/25/2000.

		Nobs	Mean	Median	Max	Min	Std Dev	ρ_1	ρ_2	ρ_{12}	Q(12)
Full Sample											
RANGE	1/2/62-8/25/00	9700	1.464	1.407	22.904	0.145	0.76	0.629	0.575	0.443	30874
UPR	1/2/62-8/25/00	9700	0.737	0.636	9.053	0	0.621	0.308	0.147	0.172	3631
DWNR	1/2/62-8/25/00	9700	−0.727	−0.598	0	−22.9	0.681	0.326	0.181	0.162	4320
Before structural shift											
RANGE	1/2/62-4/20/82	5061	1.753	1.643	9.326	0.53	0.565	0.723	0.654	0.554	21802
UPR	1/2/62-4/20/82	5061	0.889	0.798	8.631	0	0.581	0.335	0.087	0.106	1199
DWNR	1/2/62-4/20/82	5061	−0.864	−0.748	0	−6.514	0.559	0.378	0.136	0.163	2427
After structural shift											
RANGE	4/21/82-8/25/00	4639	1.15	0.962	22.904	0.146	0.818	0.476	0.414	0.229	11847
UPR	4/21/82-8/25/00	4639	0.572	0.404	9.053	0	0.622	0.189	0.089	0.125	651
DWNR	4/21/82-8/25/00	4639	−0.578	−0.388	0	−22.9	0.767	0.247	0.147	0.101	994

Note: Summary statistics for the three variables, RANGE, UPR and DWNR, defined to be the differences between the max and min, the max and open, and the min and open of the daily index prices in logarithm are described. The structural shift refers to the day April 20, 1982, at which the Standard and Poor Inc. changed the ways of constructing the daily max and daily min prices. ρ_1, ρ_2 and ρ_{12} are autocorrelation coefficients for lags 1, 2, and 12 respectively, and Q(12) is the Ljung–Box statistics of lag 12.

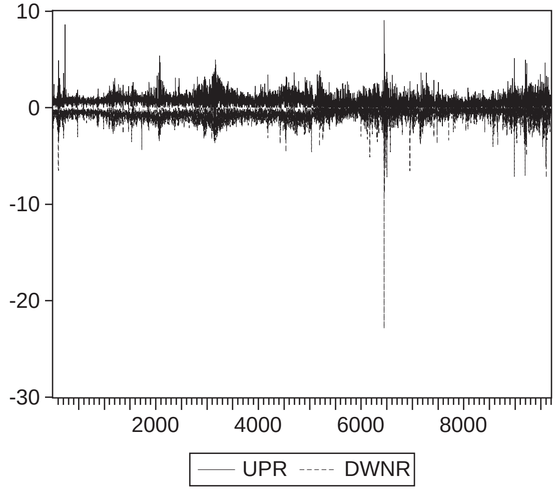


Fig. 1. Daily UPR and Daily DWNR, S&P 500, 1962/1–2000/8.

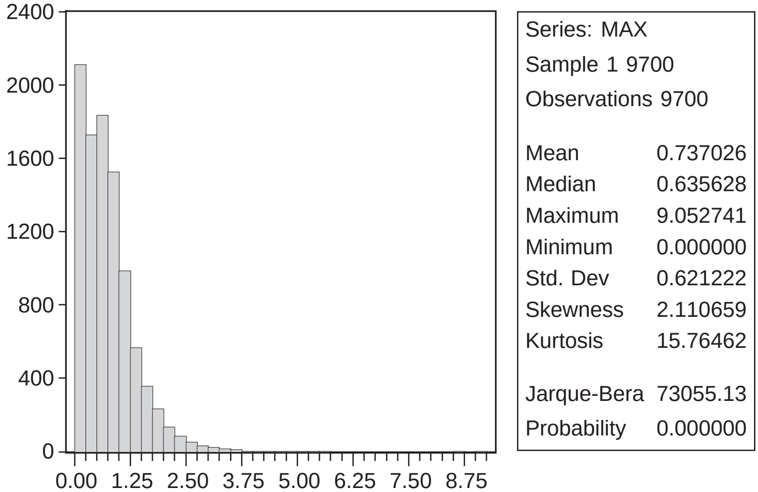


Fig. 2. Daily UPR of S&P 500 index, 1962/1–2000/8.

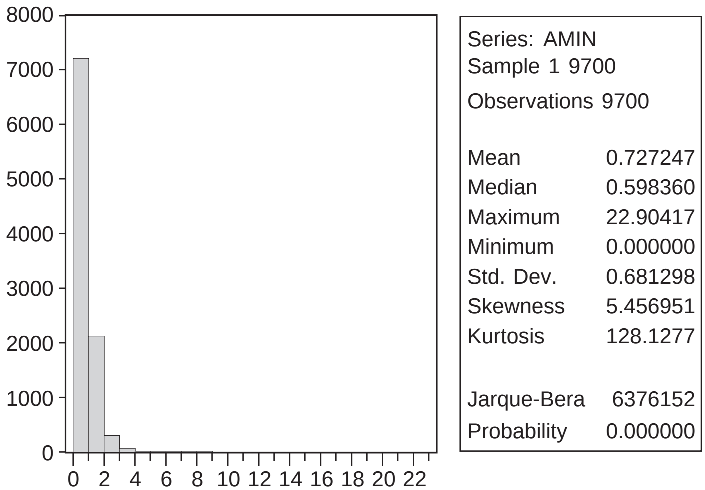


Fig. 3. Daily DWNR of S&P 500 index, Unsigned, 1962/1–2000/8.

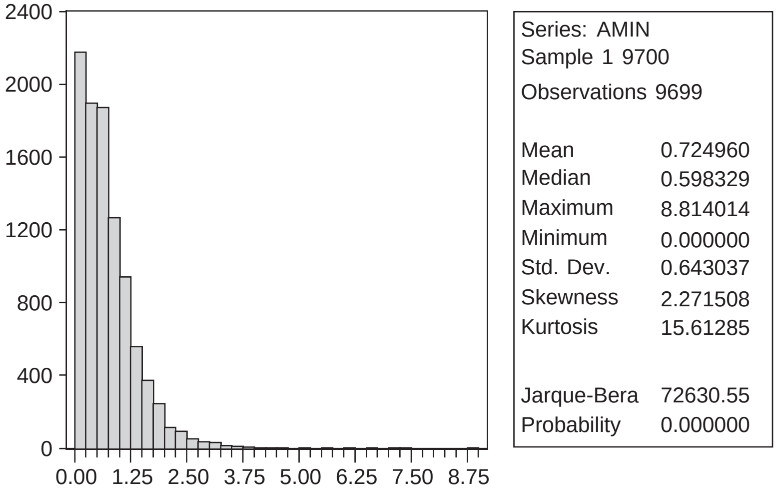


Fig. 4. Daily DWNR w/o crash, Unsigned, 1962/1–2000/8.

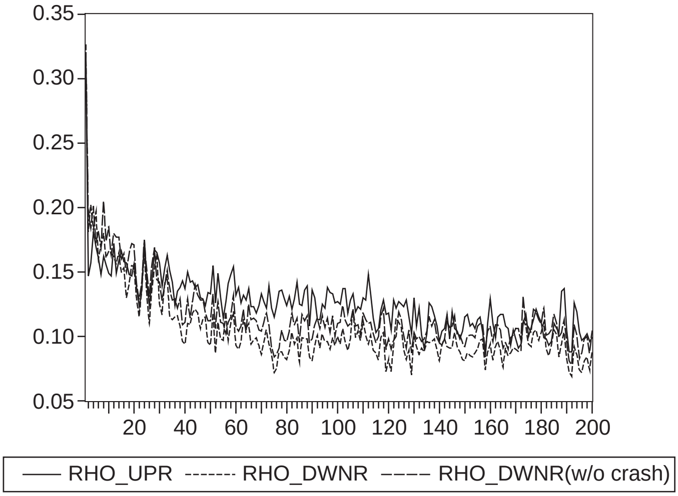


Fig. 5. Correlograms of Daily UPR and DWNR.

3.2. Estimating Results

We use QMLE to estimate the ACARR and ACARRX models with different dynamic specifications and exogenous variables. The exogenous variables considered are lagged return, r_{t-1}, for the leverage effect, a Tuesday (TUE) and a Wednesday dummy (WED), for the weekly seasonal pattern, a structural shift dummy (SD, 0 before 1982/4/20 and 1 otherwise) for capturing the shift in the data compiling method. We also include the lagged opposite range variable, i.e., DWNR in the UPR model and UPR in the DWNR model. This is for the consideration of the volatility clustering effect. Tables 2 and 3 give respectively, the model estimating results for UPR and for DWNR.

It is interesting that for both OSRs, a ACARR(2,1) clearly dominates the simpler alternative of ACARR(1,1) model, which is in contrast of the result in Chou (2005) using CARR to estimate the range variable.[10] This is shown clearly by the difference in the values of the log-likelihood function (LLF) reported for the two models, ACARR(1,1) vs. ACARR(2,1). The ACARR(2,1) model is consistent with the specification of the Component GARCH model of Engle and Kim (1999), in which the volatility dynamics is decomposed into two parts, a permanent component and a temporary component.

Table 2. QMLE Estimation of ACARR Using Daily Upward Range of S&P 500 Index 1/2/1962–8/25/2000.

	ACARR(1,1)	ACARR(2,1)	ACARRX(2,1)-a	ACARR(2,1)-b	ACARR(2,1)-c
LLF	−12035.20	−12011.86	−11955.78	−11949.64	−11950.32
Constant	0.002[3.216]	0.001[3.145]	−0.002[−0.610]	−0.003[−0.551]	−0.004[−0.973]
UPR(t−l)	0.03[8.873]	0.145[10.837]	0.203[14.030]	0.179[11.856]	0.186[12.845]
UPR(t−2)		−0.126[−9.198]	−0.117[−9.448]	−0.112[−8.879]	−0.115[−9.245]
λ(t−1)	0.968[267.993]	0.978[341.923]	0.903[69.643]	0.871[48.426]	0.877[52.942]
r(t−l)			−0.057[−8.431]	−0.018[−1.959]	−0.023[−2.734]
TUE			0.058[3.423]	0.059[3.475]	0.059[3.481]
WED				0.02[1.271]	
SD			0.0000[0.201]	−0.003[−1.647]	
DWNR(−l)				0.046[4.868]	0.042[4.803]
Q(12)	184.4[0.000]	22.346[0.034]	22.304[0.034]	20.282[0.062]	20.503[0.053]

$$UPR_t = \lambda_t \varepsilon_t$$

$$\lambda_t^u = \omega^u + \sum_{i=1}^{p} \alpha_i^u UPR_{t-i} + \sum_{j=1}^{q} \beta_j^u \lambda_{t-j}^u + \sum_{l=1}^{L} \gamma_l X_{t-1}^l$$

$$\varepsilon_t \sim iid\, f(\cdot)$$

Note: Estimation is carried out using the QMLE method hence it is equivalent to estimating an exponencial ACARR(X) (p,q) or and EACARR(X) (p,q) model. Numbers in parentheses are t-ratios (p-values) with robust standard errors for the model coefficients (Q statistics). LLF is the log-likelihood function.

Table 3. QMLE Estimation of ACARR Using Daily Downward Range of S&P500 Index 1/2/1962–8/25/2000.

	ACARR(1,1)	ACARR(2,1)	ACARRX(2,1)-a	ACARRX(2,1)-b	ACARRX(2,1)-c
LLF	−11929.39	−11889.61	−11873.54	−11868.55	−11870.14
Constant	0.014[5.905]	0.004[4.088]	0.017[4.373]	0.017 [3.235]	0.017[4.417]
DWNR(*t*-l)	0.084[11.834]	0.229[16.277]	0.252[16.123]	0.233[14.770]	0.239[16.364]
DWNR(*t*-2)		−0.195[−13.489]	−0.189[−12.811]	−0.185[−12.594]	−0.186[−12.897]
λ(*t*−1)	0.897[101.02]	0.961[199.02]	0.927[87.639]	0.906[61.212]	0.911[63.893]
r(*t*−1)			0.023[4.721]	−0.009[−1.187]	
TUE				−0.008[0.503]	
WED			−0.051[−3.582]	−0.053[−3.617]	−0.052[−3.587]
SD			−0.002[−2.124]	0.001[0.904]	
UPR(−1)				0.037[4.164]	0.028[5.084]
Q(12)	192.8[0.000]	18.94[0.009]	22.227[0.035]	14.422[0.275]	14.774[0.254]

$$DWNR_t = \lambda_t \varepsilon_t$$

$$\lambda_t^d = \omega^d + \sum_{i=1}^{p} \alpha_i^d DWNR_{t-i} + \sum_{j=1}^{q} \beta_j^d \lambda_{t-j}^d + \sum_{l=1}^{L} \gamma_l X_{t-1}^l$$

$$\varepsilon_t \sim iid\, f(\cdot)$$

Note: Estimation is carried out using the QMLE method hence it is equivalent to estimating an Exponential ACARR(X)(p,q) or and EACARR(X)(p,q) model. Numbers in parentheses are *t*-ratios(*p*-values) with robust standard errors for the model coefficients (*Q* statistics). LLF is the log-likelihood function.

Another conjecture for the inadequacy of the (1,1) dynamic specification is related to the volatility clustering effect. It is known that volatility clusters over time and in the original words of Mandelbrot (1963), "large changes tend to be followed by large changes and small by small, of either sign…". Given that range can be used as a measure of volatility, both UPR and DWNR can be viewed as "signed" measure of volatility. It is hence not surprising that a simple dynamic structure offered by the ACARR(1,1) model is not sufficient to capture the clustering effect. This conjecture is supported by the result of the model specification of ACARRX(2,1)-b where the opposite OSR are included and the coefficients significantly different from zero.

The dynamic structures for the UPR and DWNR variable are different as is revealed in comparing the values of the coefficients. The coefficient of $\beta 1$, measuring the long-term persistence effect, is (0.927, 0.906, 0.911) respectively, for the three different ACARRX specifications for DWNR in Table 3. They are all higher than their corresponding elements (0.903, 0.871, 0.877) in the ACARRX models for UPR in Table 2. This suggests that volatility shocks in the downside are more long-lived than in the upside. Further the impact coefficient $\alpha 1$ is equal to (0.252, 0.233, 0.239) in the DWNR models and is (0.203, 0.179, 0.186) in the UPR models. Volatility shock effects in the short-run are also higher for the downside shocks than for the upward surges. Both of these findings are new in the literature of financial volatility models as all existing literatures do not distinguish the shock asymmetry in this fashion.

Another interesting comparison between the two OSR models is on the leverage effect. This coefficient is statistically negative (positive) for the ACARRX(2,1)-a specifications for the UPR (DWNR). It is however, less significant or insignificant in models ACARRX(2,1)-b and ACARRX(2,1)-c, when the lagged opposite OSR is included. My conjecture is that the opposite sided ranges are correlated with the returns and hence multicollinearity reduces some explanatory power of the leverage effect. It remains, however, to be explained why such a phenomenon is more severe for the DWNR model than the UPR models. We leave this issue for future studies.

A different weekly seasonality also emerges from the comparison of the estimation result of the two OSRs. For reasons unknown to me, a positive Tuesday effect is found for the upward range while a negative Wednesday effect is present for the downward range. The dummy variable SD, measuring the effect of the structuring change in the data compiling method, are not significant for UPR models but are negatively significant for one of the DWNR models. It is not clear why there should be such difference in the results. Again leave these as empirical puzzles to be explored in future studies.

Model specification tests are carried out in two ways, the Ljung–Box-Q statistics and the Q–Q plots. The Ljung–Box Q statistics measure the overall significance of the autocorrelations in the residuals for the fitted models.[11] The evidence shown in the two tables are consistent that a pure ACARR model is not sufficient and exogenous variables are necessary to warrant the model to pass the model misspecification tests. Using a 5% significance level for the test, the model is satisfactory once the lagged returns, the weekly dummies and the opposite-sided range are included in the specifications.

Figs. 6 and 7 provide the expected and observed daily UPR and DWNR respectively. It is interesting to note that the ACARR model gives smoother yet very adaptive estimates of the two one-sided ranges. Figs. 8–11 are histograms and Q–Q plots of the estimated residuals in the two models. It seems to indicate that the exponential distribution is more satisfactory for the UPR than for the DWNR as the degree of fitness of fit can be measured by the deviations of the Q–Q plot from the 45 degree lines. This fact further indicates the difference in the characteristics of the two variables in addition to the results reported above. Whether a different error distribution will be more useful warrants more investigation. For example in Chou (2005) I found that a more general error distribution such as Weibull might improve the goodness of fit substantially in the CARR model.

The message from this section is clear: the market dynamics for the upward swing and the downward plunge are different. They are different in their dynamics of the volatility shocks, i.e., the short-term impact and long-term persistence. They are also different in the forces that have effects on them, the leverage effect, the weekly seasonal effect and the volatility clustering effect. Finally, even the error structures of the two variables are different.

3.3. Comparing ACARR and CARR in Forecasting Volatility

Although the above results shows important differences in the models for the upward and the downward range, we further ask a question on the value of the modeling of asymmetries. How much difference does this modeling consideration make to improve the power of the model in forecasting volatilities? In Chou (2005) we proposed the CARR model, where the upward and downward movements of the stock price are treated symmetrically. We showed that the CARR model provides a much sharper tool in forecasting

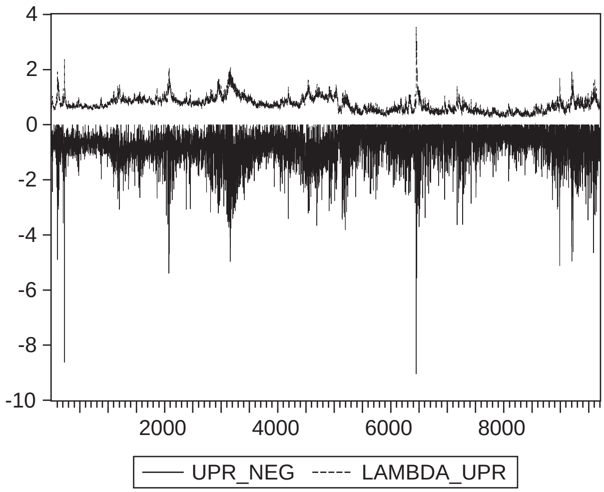


Fig. 6. Expected and Observed Daily UPR, 1962/1–2000/8.

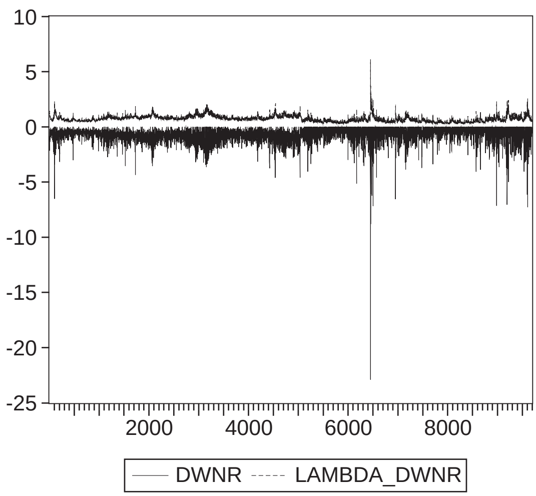


Fig. 7. Expected and Observed Daily DWNR, 1962/1–2000/8.

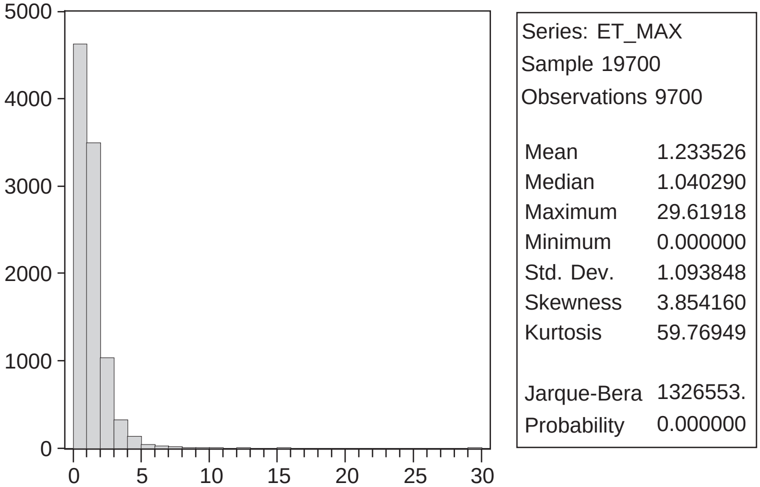


Fig. 8. Histogram of Daily et_UPR, 1962/1–2000/8.

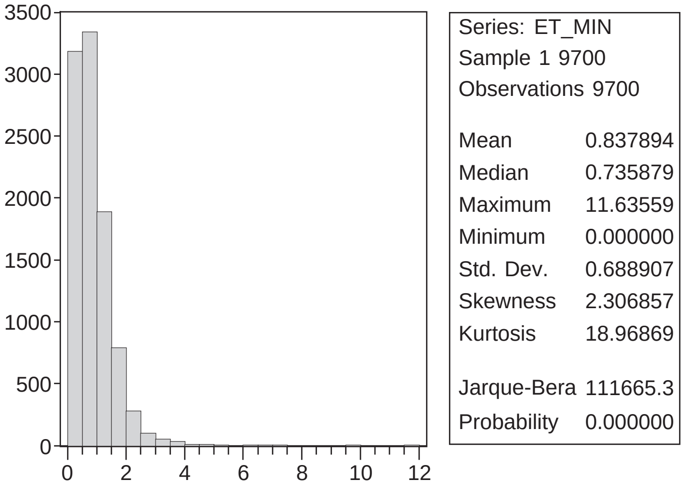


Fig. 9. Histogram of Daily et_DWNR, 1962/1–2000/8.

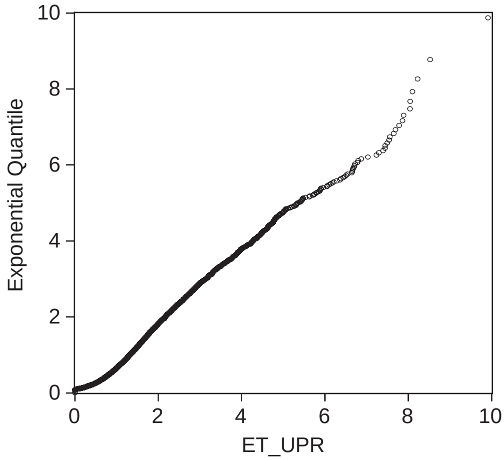


Fig. 10. Q–Q Plot of et_UPR.

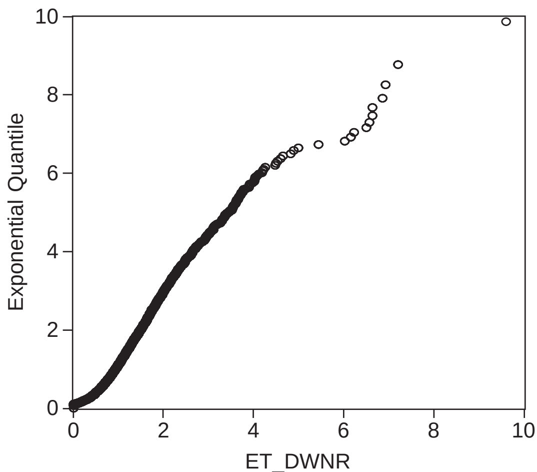


Fig. 11. Q–Q Plot of et_DWNR.

volatility than the GARCH model. In this section, we further compare the forecasting power for volatilities of the CARR model, which ignores the asymmetry, and the ACARR model which give explicit considerations to the asymmetric structures. Given our finding of the importance of modeling asymmetry in the above section, we would expect the ACARR model to provide more accurate volatility forecast comparing with the CARR model.

Since volatility is an unobservable variable, we employ three proxies as measures of volatility (henceforth MVs). They are the daily high/low range (RNG) as defined in (2.2), the daily return squared (RETSQ) as is commonly used in the literature of volatility forecast comparisons and the absolute value of the daily returns (ARET) which is more robust to outliers than the second measure. We then use the following regressions to gauge the forecasting powers of the CARR and the ACARR models.

$$\mathrm{MV}_t = a + b\,\mathrm{FV}_t(\mathrm{CARR}) + u_t \tag{3.1}$$

$$\mathrm{MV}_t = a + b\,\mathrm{FV}_t(\mathrm{ACARR}) + u_t \tag{3.2}$$

$$\mathrm{MV}_t = a + b\,\mathrm{FV}_t(\mathrm{CARR}) + c\,\mathrm{FV}(\mathrm{ACARR}) + u_t \tag{3.3}$$

$\mathrm{FV}_t(\mathrm{CARR})$ is the forecasted volatility using the CARR model in (2.5). $\mathrm{FV}_t(\mathrm{ACARR})$ is computed as the sum of the forecasted UPR and forecasted DWNR as is shown in (2.4). Proper transformations are made to adjust the difference between a variance estimator and a standard deviation estimator. Table 4 gives the estimation result.

The results are consistent for the three measures of volatility. In all cases, the forecasted volatility using ACARR dominates the forecasted volatility using CARR. In the three measures, the corresponding t-ratios for the two models are (21.83, 0.46) using RNG, (7.61, −2,32) using RETSQ and (8.09, −1.91) using ARET. Once the forecasted volatility using ACARR is included, CARR provides no additional explanatory power. Another interesting observation is that the results using range as the measured volatility look particularly favorable for the ACARR model and the absolute results are a bit weaker. In other words, the adjusted R^2 of the regression using these two measures are much smaller than that using RNG. This result is consistent with the observation in Chou (2005) that both RETSQ and ARET are based on close-to-close return data and are much more noisier than RNG which is based on the extreme values of the price.

Table 4. ACARR versus CARR.

Measured Volatility	Explanatory Variables			Adj. R^2	S.E.
	Constant	FV(CARR)	FV(ACARR)		
RNG	−0.067[−0.366]	1.005[96.29]		0.489	0.543
RNG	−0.006[−4.148]		1.047[101.02]	0.513	0.531
RNG	−0.067[0.632]	0.021[0.46]	1.026[21.83]	0.513	0.531
RETSQ	−1.203[−1.35]	0.397[14.25]		0.02	5.725
RETSQ	−0.265[−2.94]		0.459[16.02]	0.026	5.709
RETSQ	−0.249[−2.76]	−0.191[−2.32]	0.644[7.61]	0.026	5.708
ARET	1.142[7.41]	0.334[27.07]		0.07	0.642
ARET	0.113[5.85]		0.354[28.28]	0.076	0.639
ARET	0.115[5.95]	−0.106[−1.91]	0.458[8.09]	0.076	0.639

Note: In-sample Volatility Forecast Comparison Using Three Measured Volatilities as Benchmarks. The three measures of volatility are RNG, RETSQ and ARET: respectively, daily ranges, squared-daily-returns, and absoulte daily return. ACARR(1,1) model is fitted for the range series and a ACARR models are fitted for the upward range and the downward range series. FV(CARR) (FV(ACARR)) is the forecasted volatility using CARR (ACARR). FV(ACARR) is the forecasted range using the sum of the forcasted upward range and downward range. Proper transformations are made for adjusting the difference between a variance estimator and a standard-deviation estimator. Numbers in parentheses are *t*-ratios.

$$\mathrm{MV}_t = a + b\,\mathrm{FV}_t(\mathrm{CARR}) + u_t$$
$$\mathrm{MV}_t = a + c\,\mathrm{FV}_t(\mathrm{ACARR}) + u_t$$
$$\mathrm{MV}_t = a + b\,\mathrm{FV}_t(\mathrm{CARR}) + c\,\mathrm{FV}_t(\mathrm{ACARR}) + u_t$$

4. CONCLUSION

The ACARR model provides a simple, yet efficient and natural framework to analyze the asymmetry of the price movement in financial markets. Applications can be used in computing the option prices where the upward (downward) range (or the maximum (minimum) return) is more relevant for computing the price of a call (put) option. Value-at-Risk is another important area for applications using the downward range dynamic model. The ACARR model is related to studies like the duration between a threshold high or low price level. Further more, the ACARR model can be used to forecast volatilities comparing with the symmetric model, CARR GARCH, and SV models, or other asymmetric volatility models like EGARCH, GJR-GARCH models. Further Monte Carlo analysis will be useful as well as applications to other financial markets such as foreign exchanges, bonds, and commodities. Applications of the ACARR model to other frequency of range interval, say every 30 min, every hour, or every quarter, and other frequencies, will provide further understanding of the usefulness/limitation of the model. Other generalization of the ACARR model will be worthy subjects of future research, for example, the generalization of the univariate to a multivariate framework, models simultaneously treat the price return and the range data, long memory ACARR models.[12]

The ACARR model in this paper can be seen as an example of an emerging literature: applications of extreme value theory in finance. Embrecht, Kluppelberg, and Mikosch (1999) and Smith (1999), among others, are strong advocates of such an approach in studying many important issues in financial economics. Noticeable examples are Embrechts, McNeil, and Straumann (2002) for correlation of market extreme movements, McNeil and Frey (2000) for volatility forecasts, and Tsay (2000) for modeling crashes. In fact, all the static range literature (Parkinson, 1980) and the long-term dependence literature using rescaled range (Mandelbrot, 1972; Lo, 1991) can be viewed as earlier examples of this more general broader approach to the study of empirical finance.

NOTES

1. A general data generating process for P_t can be written as

$$dP_t = \mu_t + \sigma_t dW_t$$
$$d\sigma_t = \theta_t + \kappa dV_t$$

where W_t and V_t are two independent standard Wiener processes, or Brownian motions.

2. See Lo (1991) for a similar case and a proof.

3. It is not clear to me yet how the daily highs/lows of asset prices are compiled reported on the public or private data sources such as the Wall Street Journal, Financial Times, and in CRSP. They may be computed from a very high, fixed frequency. Alternatively, they may be computed directly from the transaction data, a sampling frequency with non-fixed intervals.

4. Although in this paper we follow the approach of Engle and Russell (1998) in relying on the QMLE for estimation, it is important to recognize the fact that the limiting distribution of CARR is known while it is not the case for ACD. This issue is dealt within the later section.

5. It is of course a worthy topic for future research as to how much of efficiency gain can be obtained by utilizing the FIML with the limiting distribution to estimate the parameters. Alternatively, one can estimate the density function using nonparametric methods.

6. See Engle and Russell (1998) for a proof.

7. It is noteworthy that the closure property holds only for the weak-GARCH processes. Namely, in general, the GARCH process is not closed under aggregation. See Drost and Nijman (1993) for the discussion of the closure property of GARCH process.

8. The exact date is unknown since this change in compiling process was not documented by the company. However, from a detailed look at the data, the most likely date is April 20, 1982.

9. Mathematically, these two compiling methods are respectively, index of the highs (lows) and highs (lows) of the index. The Jensen inequality tells us that these two operations are not interchangeable.

10. For the range variable, it is consistently found that a CARR(1,1) model is sufficient to capture the dynamics for daily and weekly and for different sub-sample periods.

11. In the GARCH literature a Q-statistics for the squared normalized residuals is usually included as well to account for the remaining ARCH effect in the residual. Here we do not include such a statistics because range is by itself a measure of volatility and this statistics will be measuring the persistence of the volatility of volatility. For formal tests of the distribution, some tests can be incorporated to complement the Q–Q plots.

12. In the daily ACARR models, as is suggested by the Portmanteau statistics, the memory in range (hence in the volatility) seems to be longer than can be accounted for using the simple ACARR(1,1) or ACARR(2,1) models with short memories. However, such a phenomenon disappears in the weekly model. Given our empirical results, it is questionable whether such an attempt is useful in practice.

ACKNOWLEDGMENTS

I would like to thank an anonymous referee and the editor Dek Terrell for helpful comments and suggestions. This paper has also benefited from conversations with Jeff Russell, George Tiao, Wen-Jen Tsai, Ruey Tsay and other colleagues in the Academia Sinica. Financial support for this project is provided by the National Science Council of Taiwan under contract NSC 90-2415-H-001-032.

REFERENCES

Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *Journal of Finance*, *57*, 1047–1092.

Andersen, T., Bollerslev, T., Diebold, F., & Labys, P. (2000). The distribution of exchange rate volatility. *Journal of American Statistical Association*, *96*, 42–55.

Bai, X., Russell, J., & Tiao, G. (2000). *Beyond Merton's Utopia: Effects on non-normality and dependence on the precision of variance estimates using high-frequency financial data.* GSB working paper, University of Chicago.

Barberis, N., & Huang, M. (2000). Mental accounting, loss aversion, and individual stock returns. GSB working paper, University of Chicago.

Beckers, S. (1983). Variance of security price return based on high, low and closing prices. *Journal of Business*, *56*, 97–112.

Black, F. (1976). Studies of stock, price volatility changes. *Proceedings of the 1976 meetings of the business and economics statistics section* (pp. 177–181). American Statistical Association.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time varying covariances. *Econometric Reviews*, *11*, 143–172.

Campbell, J. (1997). Asset prices, consumption and the business cycle. In: T. Taylor & W. Woodford (Eds), *The handbook of macroeconomics*. Amsterdam: North-Holland.

Chen, M., Russell, J., & Tsay, R. (2000). *Information determinants of bid and ask quotes: Implications for market liquidity and volatility*. Working paper, GSB, University of Chicago.

Chou, R. (1988). Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics*, *3*, 279–294.

Chou, R. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, *37*, 561–582.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, *10*, 407–432.

Cox, J., & Rubinstein, M. (1985). *Options markets*. Englewood cliffs, NJ: Prentice-Hall.

Drost, F. C., & Nijman, T. E. (1993). Temporal aggregation of GARCH processes. *Econometrica*, *61*, 909–927.

Embrecht, P., Kluppelberg, C., & Mikosch, T. (1999). *Modelling extremal events for insurance and finance*. Berlin: Springer.

Embrechts, P., McNeil, A., & Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In: M. Dempster (Ed.), *Risk management: Value at risk and beyond* (pp. 176–223). Cambridge: Cambridge University Press.

Engle, R. F., Lilien, D. M., & Robbins, R. P. (1987). Estimating time varying risk premia in the term structure. The ARCH-M model. *Econometrica*, *55*, 391–407.

Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, *48*, 1749–1778.

Engle, R. F., & Russell, J. (1998). Autoregressive conditional duration: A new model for irregular spaced transaction data. *Econometrica*, *66*(5), 1127–1162.

Gallant, A., Hsu, C., & Tauchen, G. (1999). Calculating volatility diffusions and extracting integrated volatility. *Review of Economics and Statistics*, *81*, 159–192.

Garman, M., & Klass, M. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, *53*, 67–78.

Karpoff, J. (1987). The relation between price change and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, *22*, 109–126.

Kunitomo, N. (1992). Improving the Parkinson method of estimating security price volatilities. *Journal of Business*, *65*, 295–302.

Lamoureux, C., & Lastrapes, W. (1990). Heteroskedasticity in stock return data: Volume versus GARCH effect. *Journal of Finance*, *45*, 221–229.

Levy, H., & Markowitz, H. M. (1979). *Approximating expected utility by a function of mean and variance. American Economic Review*, *69*, 308–317.

Lo, A. (1991). Long memory in stock market prices. *Econometrica*, *59*(5), 1279–1313.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*, 394–419.

Mandelbrot, B. (1972). Statistical methodology for non-periodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement*, *1*, 259–290.

McNeil, A., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*, 271–300.

Nelson, D. (1990). ARCH models as diffusion approximations. *Journal of Econometrics*, *45*, 7–39.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, *53*, 61–65.

Pindyck, R. (1984). Risk, inflation, and the stock market. *American Economic Review*, *74*, 335–351.

Poterba, J., & Summers, L. (1986). The persistence of volatility and stock market fluctuations. *American Economic Review*, *76*, 1142–1151.

Rogers, C. (1998). Volatility estimation with price quanta. *Mathematical Finance*, *8*(3), 277–290.

Rogers, C., & Satchell, S. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability*, *1*(4), 504–512.

Smith, R. (1999). *Measuring risk with extreme value theory*. Working paper, department of Statistics, University of North Carolina at Chapel Hill.

Tsay, R. (2000). *Extreme value analysis of financial data*. Working paper, GSB, University of Chicago.

Wiggins, J. (1991). Empirical tests of the bias and efficiency of the extreme-value variance estimator for common stocks. *Journal of Business*, *64*(3), 417–432.

Yang, D., & Zhang, Q. (2000). Drift independent volatility estimation based on high, low, open, and close prices. *Journal of Business*, *73*(3), 477–491.

This page intentionally left blank

ON A SIMPLE TWO-STAGE CLOSED-FORM ESTIMATOR FOR A STOCHASTIC VOLATILITY IN A GENERAL LINEAR REGRESSION

Jean-Marie Dufour and Pascale Valéry

ABSTRACT

In this paper, we consider the estimation of volatility parameters in the context of a linear regression where the disturbances follow a stochastic volatility (SV) model of order one with Gaussian log-volatility. The linear regression represents the conditional mean of the process and may have a fairly general form, including for example finite-order autoregressions. We provide a computationally simple two-step estimator available in closed form. Under general regularity conditions, we show that this two-step estimator is asymptotically normal. We study its statistical properties by simulation, compare it with alternative generalized method-of-moments (GMM) estimators, and present an application to the S&P composite index.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 259–288

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20010-5

1. INTRODUCTION

Modelling conditional heteroscedasticity is one of the central problems of financial econometrics. The two main families of models for that purpose consist of generalized auto-regressive conditional heteroskedasticity (GARCH)-type processes, originally introduced by Engle (1982), and stochastic volatility (SV) models proposed by Taylor (1986). Although the latter may be more attractive – because they are directly connected to diffusion processes used in theoretical finance – GARCH models are much more popular because they are relatively easy to estimate; for reviews, see Gouriéroux (1997) and Palm (1996). In particular, evaluating the likelihood function of GARCH models is simple compared to SV models for which it is very difficult to get a likelihood in closed form; see Shephard (1996), Mahieu and Schotman (1998) and the review of Ghysels, Harvey, and Renault (1996). Due to the high dimensionality of the integral defining the likelihood function, this is a general feature of almost all nonlinear latent variable models. As a result, maximum likelihood methods are prohibitively expensive from a computational viewpoint, and alternative methods appear to be required for applying such models.

Since the first discrete-time SV models were proposed by Taylor (1986) as an alternative to ARCH models, much progress has been made regarding the estimation of nonlinear latent variable models in general and SV models in particular. The methods suggested include quasi maximum likelihood estimation (see Nelson, 1988; Harvey, Ruiz, & Shephard, 1994; Ruiz, 1994), generalized method-of-moments (GMM) procedures (Melino & Turnbull, 1990; Andersen & Sørensen, 1996), sampling simulation-based techniques– such as simulated maximum likelihood (Danielsson & Richard, 1993; Danielsson, 1994), indirect inference and the efficient method of moments (Gallant, Hsieh, & Tauchen, 1997; Andersen, Chung, & Sørensen, 1999) – and Bayesian approaches (Jacquier, Polson, & Rossi, 1994; Kim, Shephard, & Chib, 1998; Wong, 2002a, SVb). Note also that the most widely studied specification in this literature consists of an SV model of order one with Gaussian log-volatility and zero (or constant) conditional mean. The most notable exception can be found in Gallant et al. (1997) who allowed for an autoregressive conditional mean and considered a general autoregressive process on the log-volatility. It is remarkable that all these methods are highly nonlinear and computer-intensive. Implementing them can be quite complicated and get more so as the number of parameters increases (e.g., with the orders of the autoregressive conditional mean and log-volatility).

In this paper, we consider the estimation of SV parameters in the context of a linear regression where the disturbances follow an SV model of order one with Gaussian log-volatility. The linear regression represents the conditional mean of the process and may have a fairly general form, which includes for example finite-order autoregressions. Our objective is to develop a computationally inexpensive estimator that can be easily exploited within simulation-based inference procedures, such as Monte Carlo and bootstrap tests.[1] So we study here a simple two-step estimation procedure which can be described as follows: (1) the conditional mean model is first estimated by a simple consistent procedure that takes into account the SV structure; for example, the parameters of the conditional mean can be estimated by ordinary least squares (although other estimation procedures can be used); (2) using residuals from this preliminary regression, the parameters of the SV model are then evaluated by a method-of-moment estimator based on three moments (2S-3 M) for which a simple closed-form expression can be derived. Under general regularity conditions, we show the two-stage estimator is asymptotically normally distributed. Following recent results on the estimation of autoregressive models with SV (see, for example, Gonçalves & Kilian, 2004, Theorem 3.1), this entails that the result holds for such models.

An interesting and potentially useful feature of the asymptotic distribution stems from the fact that its covariance matrix does not depend on the distribution of the conditional mean estimator, i.e., the estimation uncertainty on the parameters of the conditional mean does not affect the distribution of the volatility parameter estimates (asymptotically). The properties of the 2S-3M estimator are also studied in a small Monte Carlo experiment and compared with GMM estimators proposed in this context. We find that the 2S-3M estimator has quite reasonable accuracy with respect to the GMM estimators: indeed, in several cases, the 2S-3M estimator has the lowest root mean-square error. With respect to computational efficiency, the 2S-3M estimator always requires less than a second while GMM estimators may take several hours before convergence is obtained (if it does). Finally, the proposed estimator is illustrated by applying it to the estimation of an SV model on the Standard and Poor's Composite Price Index (1928–1987, daily data).

The paper is organized as follows. Section 3 sets the framework and the main assumptions used. The closed-form estimator studied is described in Section 3. The asymptotic distribution of the estimator is established in Section 4. In Section 5, we report the results of a small simulation study on the performance of the estimator. Section 6 presents the application to the Standard and Poor's Composite Price Index return series. We conclude in Section 7. All proofs are gathered in the Appendix.

2. FRAMEWORK

We consider here a regression model for a variable y_t with disturbances that follow an SV process, which is described below following a notation similar to the one used by Gallant et al. (1997). $\mathbb{N}_0$ refers to the nonnegative integers.

Assumption 2.1. *Linear regression with stochastic volatility*. The process $\{y_t : t \in \mathbb{N}_0\}$ follows an SV model of the type:

$$y_t = x_t'\beta + u_t \tag{2.1}$$

$$u_t = \exp(w_t/2) r_y z_t\,, \quad w_t = \sum_{j=1}^{L_w} a_j w_{t-j} + r_w v_t \tag{2.2}$$

where x_t is a $k \times 1$ random vector independent of the variables $\{x_{\tau-1},\ z_\tau,\ v_\tau,\ w_\tau : \tau \leq t\}$, and $\beta,\ r_y,\ \{a_j\}_{j=1}^{L_w},\ r_w$ are fixed parameters.

Typically y_t denotes the first difference over a short time interval, a day for instance, of the log-price of a financial asset traded on security markets. The regression function $x_t'\beta$ represents the conditional mean of y_t (given the past) while the SV process determines a varying conditional variance. A common specification here consists in assuming that $x_t'\beta$ has an autoregressive form as in the following restricted version of the model described in Assumption 2.1.

Assumption 2.2. *Autoregressive model with stochastic volatility*. The process $\{y_t : t \in \mathbb{N}_0\}$ follows an SV model of the type:

$$y_t - \mu_y = \sum_{j=1}^{L_y} c_j (y_{t-j} - \mu_y) + u_t \tag{2.3}$$

$$u_t = \exp(w_t/2) r_y z_t\,, \quad w_t = \sum_{j=1}^{L_w} a_j w_{t-j} + r_w v_t \tag{2.4}$$

where $\beta, \{c_j\}_{j=1}^{L_y}, r_y, \{a_j\}_{j=1}^{L_w}$ and r_w are fixed parameters.

We shall refer to the latter model as an AR-SV(L_y, L_w) model. The lag lengths of the autoregressive specifications used in the literature are typically short, e.g.: $L_y = 0$ and $L_w = 1$ (Andersen & Sørensen, 1996; Jacquier, Polson & Rossi, 1994; Andersen et al., 1999), $0 \leq L_y \leq 2$ and $0 \leq L_w \leq 2$ (Gallant et al., 1997). In particular, we will devote special attention to the AR-SV(1,1) model:

$$y_t - \mu_y = c(y_{t-1} - \mu_y) + \exp(w_t/2) r_y z_t \ , \quad |c| < 1 \tag{2.5}$$

$$w_t = a w_{t-1} + r_w v_t \ , \quad |a| < 1 \tag{2.6}$$

so that

$$cov(w_t, w_{t+\tau}) = a^\tau \gamma \tag{2.7}$$

where $\gamma = r_w^2/(1 - a^2)$. The basic assumptions described above will be completed by a Gaussian distributional assumption and stationarity condition.

Assumption 2.3. *Gaussian noise*. The vectors $(z_t, v_t)'$, $t \in \mathbb{N}_0$, are *i.i.d.* according to a $\mathrm{N}[0, I_2]$ distribution.

Assumption 2.4. *Stationarity*. The process $s_t = (y_t, w_t)'$ is strictly stationary.

The process defined above is Markovian of order $L_s = \max(L_y, L_w)$. Under these assumptions, the AR-SV (L_y, L_w) is a parametric model with parameter vector

$$\delta = (\mu_y, c_1, \ldots, c_{L_y}, r_y, a_1, \ldots, a_{L_w}, r_w)'. \tag{2.8}$$

Due to the fact that the model involves a latent variable (w_t), the joint density of the vector of observations $y(T) = (y_1, \ldots, y_T)$ is not available in closed form because the latter would involve evaluating an integral with dimension equal to the whole path of the latent volatilities.

3. CLOSED-FORM METHOD-OF-MOMENTS ESTIMATOR

In order to estimate the parameters of the volatility model described in the previous section, we shall consider the moments of the residual process in (2.1), which can be estimated relatively easily from regression residuals.

Specifically, we will focus on SV of order one ($L_w = 1$). Set

$$\theta = (a,\ r_y,\ r_w)' \tag{3.1}$$

$$v_t(\theta) \equiv \exp\Big(\frac{aw_{t-1} + r_w v_t}{2}\Big) r_y z_t\ ,\ \ \forall t \tag{3.2}$$

Models (2.1) and (2.2) may then be conveniently rewritten as the following identity:

$$y_t - x_t'\beta\ = v_t(\theta)\ ,\ \ \forall t\ . \tag{3.3}$$

The estimator we will study is based on the moments of the process $u_t \equiv v_t(\theta)$. The required moments are given in the following lemma.[2]

Lemma 3.1. *Moments and cross-moments of the volatility process.* Under the assumptions 2.1, 2.3 and 2.4 with $Lw = 1$, the moments and cross-moments of $u_t = \exp(w_t/2) r_y z_t$ are given by the following formulas: for k, l and $m \in \mathbb{N}_0$,

$$\begin{aligned}\mu_k(\theta) \equiv \mathrm{E}(u_t^k) &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!} \exp\left[\frac{k^2}{8} r_w^2/(1-a^2)\right], \quad \text{if } k \text{ is even} \\ &= 0\ , \quad \text{if } k \text{ is odd}\end{aligned} \tag{3.4}$$

$$\begin{aligned}\mu_{k,l}(m|\theta) &\equiv \mathrm{E}(u_t^k u_{t+m}^l) \\ &= r_y^{k+l} \frac{k!}{2^{(k/2)}(k/2)!} \frac{l!}{2^{(l/2)}(l/2)!} \exp[\frac{r_w^2}{8(1-a^2)}(k^2 + l^2 + 2kla^m)]\end{aligned} \tag{3.5}$$

if k and l are even, and $\mu_{k,l}(m|\theta) = 0$ if k or l is odd.

On considering $k = 2$, $k = 4$ or $k = l = 2$ and $m = 1$, we get:

$$\mu_2(\theta) = \mathrm{E}(u_t^2) = r_y^2 \exp[r_w^2/2(1-a^2)] \tag{3.6}$$

$$\mu_4(\theta) = \mathrm{E}(u_t^4) = 3r_y^4 \exp[2r_w^2/(1-a^2)] \tag{3.7}$$

$$\mu_{2,2}(1|\theta) = \mathrm{E}[u_t^2 u_{t-1}^2] = r_y^4 \exp[r_w^2/(1-a)] \tag{3.8}$$

An important observation here comes from the fact that the above equations can be explicitly solved for a, r_y and r_w. The solution is given in the following lemma.

Lemma 3.2. *Moment equations solution.* Under the assumptions of Proposition 3.1, we have:

$$a = \left[\frac{\log[\mu_{2,2}(1|\theta)] + \log[\mu_4(\theta)/(3\mu_2(\theta)^4)]}{\log[\mu_4(\theta)/(3\mu_2(\theta)^2)]}\right] - 1 \tag{3.9}$$

$$r_y = \frac{3^{1/4}\mu_2(\theta)}{\mu_4(\theta)^{1/4}} \tag{3.10}$$

$$r_w = [(1 - a^2)\log[\mu_4(\theta)/(3\mu_2(\theta)^2)]]^{1/2} \tag{3.11}$$

From Lemmas 3.1 and 3.3, it is easy to derive higher-order autocovariance functions. In particular, for later reference, we will find useful to spell out the second and fourth-order autocovariance functions.

Lemma 3.3. *Higher-order autocovariance functions.* Under the assumptions of Proposition 3.1, let $X_t = (X_{1t}, X_{2t}, X_{3t})'$ with

$$X_{1t} = u_t^2 - \mu_2(\theta),\ \ X_{2t} = u_t^4 - \mu_4(\theta),\ \ X_{3t} = u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)\,. \tag{3.12}$$

Then the covariances $\gamma_i(\tau) = \mathrm{Cov}(X_{i,t},\ X_{i,t+\tau})$, $i = 1, 2, 3$, are given by:

$$\gamma_1(\tau) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1] \tag{3.13}$$

$$\gamma_2(\tau) = \mu_4^2(\theta)[\exp(4\gamma a^\tau) - 1]\,,\ \ \forall\tau \geq 1 \tag{3.14}$$

$$\gamma_3(\tau) = \mu_{2,2}^2(1|\theta)[\exp(\gamma(1 + a)^2 a^{\tau-1}) - 1]\,,\ \ \forall\tau \geq 2 \tag{3.15}$$

where $\gamma = r_w^2/(1 - a^2)$.

Suppose now we have a preliminary estimator $\hat{\beta}$ of β. For example, for the autoregressive models (2.3)–(2.4), estimation of Eq. (2.3) yields consistent asymptotically normal estimators of β;see Gonçalves and Kilian (2004, Theorem 3.1) and Kuersteiner (2001). Of course, other estimators of the regression coefficients may be considered. Given the residuals

$$\hat{u}_t = y_t - x_t'\hat{\beta}\,,\ \ t = 0, 1,\ \ldots\ , T \tag{3.16}$$

it is then natural to estimate $\mu_2(\theta)$, $\mu_4(\theta)$, and $\mu_{2,2}(1|\theta)$ by the corresponding empirical moments:

$$\hat{\mu}_2 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\,,\quad \hat{\mu}_4 = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^4\,,\quad \hat{\mu}_2(1) = \frac{1}{T}\sum_{t=1}^{T}\hat{u}_t^2\hat{u}_{t-1}^2$$

This yields the following estimators of the SV coefficients:

$$\hat{a} = \left[\frac{\log[\hat{\mu}_2(1)] + \log[\hat{\mu}_4/(3\hat{\mu}_2^4)]}{\log[\hat{\mu}_4/(3\hat{\mu}_2^2)]}\right] - 1 \tag{3.17}$$

$$\hat{r}_y = \frac{3^{1/4}\hat{\mu}_2}{\hat{\mu}_4^{1/4}} = \left(\frac{3\hat{\mu}_2^4}{\hat{\mu}_4}\right)^{1/4} \tag{3.18}$$

$$\hat{r}_w = [(1 - \hat{a}^2)\log[\hat{\mu}_4/(3\hat{\mu}_2^2)]]^{1/2} \tag{3.19}$$

Clearly, it is straightforward to compute the latter estimates as soon as the estimator $\hat{\beta}$ used to compute the residuals $\hat{u}_t = y_t - x_t'\hat{\beta}$ is easy to obtain (e.g., $\hat{\beta}$ could be a least squares estimator).[3]

4. ASYMPTOTIC DISTRIBUTION

We will now study the asymptotic distribution of the moment estimator defined in (3.17)–(3.19). For that purpose, it will be convenient to view the latter as a special case of the general class of estimators obtained by minimizing a quadratic form of the type:

$$M_T(\theta) = [\bar{g}_T(\hat{U}_T) - \mu(\theta)]'\hat{\Omega}_T[\bar{g}_T(\hat{U}_T) - \mu(\theta)] \tag{4.1}$$

where $\mu(\theta)$ is a vector of moments, $\bar{g}_T(\hat{U}_T)$ the corresponding vector of empirical moments based on the residual vector $\hat{U}_T = (\hat{u}_1, \ldots, \hat{u}_T)'$, and $\hat{\Omega}_T$ a positive-definite (possibly random) matrix. Of course, this estimator belongs to the general family of moment estimators, for which a number of general asymptotic results do exist; see Hansen (1982), Gouriéroux and Monfort (1995b, Volume 1, Chapter 9) and Newey and McFadden (1994). However, we need to account here for two specific features, namely: (1) the disturbances in (2.1) follow an SV model, and the satisfaction of the relevant regularity conditions must be checked; (2) the two-stage nature of the procedure where the estimator of the parameter β of the conditional mean equation is obtained separately and may not be based on the same objective function as the one used to estimate θ. In particular, it is important to know whether the estimator of the conditional mean parameter β has an effect on the asymptotic distribution of the estimator of θ. It is worth noting at this stage that Andersen and Sørensen (1996) did refer to the asymptotic distribution of the usual GMM estimator as derived in Hansen (1982), but without checking the suitable regularity conditions for the SV model, which we will do here.

To spell out the properties of the estimator $\hat{\theta}_T(\hat{\Omega}_T)$ obtained by minimizing $M_T(\theta)$, we will consider first the following *generic assumptions*, where θ_0 denotes the "true" value of the parameter vector θ.

Assumption 4.1. *Asymptotic normality of empirical moments.*

$$\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta_0)] \xrightarrow{D} N[0,\ \Omega_*] \tag{4.2}$$

where $U_T \equiv (u_1,\ \ldots\ ,\ u_T)'$ and

$$\Omega_* = \lim_{T\to\infty} \mathrm{E}\{T[\bar{g}_T(U_T) - \mu(\theta_0)][\bar{g}_T(U_T) - \mu(\theta_0)]'\} \tag{4.3}$$

Assumption 4.2. *Asymptotic equivalence for empirical moments.* The random vector $\sqrt{T}[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)]$ is asymptotically equivalent to $\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta_0)]$, i.e.

$$\underset{T\to\infty}{\mathrm{plim}}\{\sqrt{T}[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)] - \sqrt{T}[\bar{g}_T(U_T) - \mu(\theta_0)]\} = 0 \tag{4.4}$$

Assumption 4.3. *Asymptotic nonsingularity of weight matrix.* $\underset{T\to\infty}{\mathrm{plim}}\,(\hat{\Omega}_T) = \Omega$ where $\det(\Omega) \neq 0$.

Assumption 4.4. *Asymptotic nonsingularity of weight matrix.* $\mu(\theta_0)$ is twice continuously differentiable in an open neighborhood of θ_0 and the Jacobian matrix $\mathrm{P}(\theta_0)$ has full rank, where $P(\theta) = \frac{\partial \mu'}{\partial \theta}$.

Given these assumptions, the asymptotic distribution of $\hat{\theta}_T(\hat{\Omega}_T)$ is determined by a standard argument on method-of-moments estimation.

Proposition 4.1. *Asymptotic distribution of method-of-moments estimator.* Under the assumptions 4.1 to 4.4,

$$\sqrt{T}[\hat{\theta}_T(\Omega) - \theta_0] \xrightarrow{D} N[0,\ V(\theta_0|\Omega)] \tag{4.5}$$

where

$$V(\theta|\Omega) = \left[P(\theta)\Omega P(\theta)'\right]^{-1} P(\theta)\Omega\Omega_*\Omega P(\theta)'\left[P(\theta)\Omega P(\theta)'\right]^{-1} \tag{4.6}$$

$P(\theta) = \frac{\partial \mu'}{\partial \theta}$. If, furthermore, (*i*) $P(\theta)$ is a square matrix, or (*ii*) Ω_* is nonsingular and $\Omega = \Omega_*^{-1}$, then

$$V(\theta|\Omega) = \left[P(\theta)\Omega_*^{-1} P(\theta)'\right]^{-1} \equiv V_*(\theta) \tag{4.7}$$

As usual, $V_*(\theta_0)$ is the smallest possible asymptotic covariance matrix for a method-of-moments estimator based on $M_T(\theta)$. The latter, in particular, is reached when the dimensions of μ and θ are the same, in which case the estimator is obtained by solving the equation

$$\bar{g}_T(\hat{U}_T) = \mu(\hat{\theta}_T)$$

Consistent estimators $V(\theta_0|\Omega)$ and $V_0(\theta_0)$ can be obtained on replacing θ_0 and Ω_* by consistent estimators.

A consistent estimator of Ω_* can easily be obtained (see Newey & West, 1987) by a Bartlett kernel estimator, i.e.:

$$\hat{\Omega}_* = \hat{\Gamma}_0 + \sum_{k=1}^{K(T)} \left(1 - \frac{k}{K(T)+1}\right)\left(\hat{\Gamma}_k + \hat{\Gamma}'_k\right) \tag{4.8}$$

where

$$\hat{\Gamma}_k = \frac{1}{T} \sum_{t=k+1}^{T} [g_{t-k}(\hat{u}) - \mu(\theta)][g_t(\hat{u}) - \mu(\theta)]' \tag{4.9}$$

with θ replaced by a consistent estimator $\tilde{\theta}_T$ of θ. The truncation parameter $K(T) = \tilde{c}\, T^{1/3}$ is allowed to grow with the sample size such that:

$$\lim_{T\to\infty} \frac{K(T)}{T^{1/2}} = 0 \tag{4.10}$$

see White and Domowitz (1984). A consistent estimator of $V_*(\theta_0)$ is then given by

$$\hat{V}_* = [P(\hat{\theta}_T)\hat{\Omega}_*^{-1} P(\hat{\theta}_T)']^{-1}. \tag{4.11}$$

The main problem here consists in showing that the relevant regularity conditions are satisfied for the estimator $\hat{\theta} = (\hat{a}, \hat{r}_y, \hat{r}_w)'$ given by (3.17)–(3.19) for the parameters of an SV model of order one. In this case, we have $\mu(\theta) = [\mu_2(\theta), \mu_4(\theta), \mu_{2,2}(1|\theta)]'$,

$$\bar{g}_T(\hat{U}_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(\hat{U}_T) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^4 \\ \frac{1}{T}\sum_{t=1}^{T} \hat{u}_t^2 \hat{u}_{t-1}^2 \end{pmatrix} \tag{4.12}$$

$$\bar{g}_T(U_T) = \frac{1}{T}\sum_{t=1}^{T} g_t(U_T) = \begin{pmatrix} \frac{1}{T}\sum_{t=1}^{T} u_t^2 \\ \frac{1}{T}\sum_{t=1}^{T} u_t^4 \\ \frac{1}{T}\sum_{t=1}^{T} u_t^2 u_{t-1}^2 \end{pmatrix} \tag{4.13}$$

where $g_t(\hat{U}_T) = [\hat{u}_t^2,\ \hat{u}_t^4,\ \hat{u}_t^2\hat{u}_{t-1}^2]'$, and $g_t(U_T) = [u_t^2,\ u_t^4,\ u_t^2 u_{t-1}^2]'$.

Since the number of moments used is equal to the number of parameters (three), the moment estimator can be obtained by taking $\hat{\Omega}_T$ equal to an identity matrix so that Assumption 4.3 automatically holds. The main problem then consists in showing that the assumptions 4.1 and 4.2 are satisfied.

Assumption 4.5. *Existence of moments.*

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t x_t' = \sigma_{2,x}(0) \tag{4.14}$$

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t u_t^2 x_t' = \sigma_{2,x,u}(0,0) \tag{4.15}$$

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_t u_{t-1}^2 x_t' = \sigma_{2,x,u}(0,1) \tag{4.16}$$

$$\operatorname*{plim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} x_{t-1} u_t^2 x_{t-1}' = \sigma_{2,x,u}(1,0) \tag{4.17}$$

where the $k \times k$ matrices $\sigma_{2,x}(0)$, $\sigma_{2,x,u}(0,0)$, $\sigma_{2,x,u}(0,1)$ and $\sigma_{2,x,u}(1,0)$ are bounded.

Proposition 4.2. *Asymptotic distribution for empirical moments.* Under the assumptions 2.1, 2.3 and 2.4 with $L_w = 1$, we have:

$$\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta_0)] \xrightarrow{D} N[0,\ \Omega_*] \tag{4.18}$$

where $\bar{g}_T(U_T) = \Sigma_{t=1}^{T} g_t/T, g_t = [u_t^2,\ u_t^4,\ u_t^2 u_{t-1}^2]'$, and

$$\Omega_* = V[g_t] = \mathrm{E}[g_t g_t'] - \mu(\theta_0)\mu(\theta_0)'. \tag{4.19}$$

Proposition 4.3. *Asymptotic equivalence for empirical moments*. Suppose the assumptions 2.1, 2.3, 2.4 and 4.5 hold with $L_w = 1$, let $\hat{\beta}$ be an estimator of β such that

$$\sqrt{T}(\hat{\beta} - \beta) \text{ is asymptotically bounded,} \tag{4.20}$$

and let $\hat{u}_t = y_t - x_t'\hat{\beta}$. Then $\sqrt{T}[\bar{g}_T(\hat{U}_T) - \mu(\theta_0)]$ is asymptotically equivalent to $\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta_0)]$.

The fact that condition (4.20) is satisfied by the least squares estimator can be easily seen from earlier published results on the estimation of regression models with SV; see Gonçalves and Kilian (2004, Theorem 3.1) and Kuersteiner (2001). Concerning equation (4.14) it holds in particular for the AR(p) case with $x_t = Y_{t-1} = (y_{t-1}, \ldots, y_{t-p})'$; see the proofs of Gonç alves and Kilian (2004, Theorems 3.1).

On assuming that the matrices Ω_* and $P(\theta_0)$ have full rank, the asymptotic normality of $\hat{\theta}_T$ follows as described in Proposition 4.1. Concerning the latter, it is interesting and potentially useful to note that this asymptotic distribution does not depend on the asymptotic distribution of the first step estimator of the autoregressive coefficient ($\hat{\beta}$) in the conditional mean equation.

5. SIMULATION STUDY

In this section, we study by simulation the properties of the 2S-3M estimator in terms of bias, variance and root mean square error (RMSE). We consider two different sets of parameters: one set with low serial dependence in the autoregressive dynamics of both processes ($c = 0.3$, $a = 0$), while the other one has high dependence ($c = 0.95$ and $a = 0.95$). For both sets, the scale parameters are fixed at $r_y = 0.5$. and $r_w = 0.5$ Under these designs, the 2S-3M estimator is compared with the GMM estimators of Andersen and Sørensen (1996) based on 5 and 24 moments. The number of replications used is 1000. The results are presented in Tables 1–3.

Globally, there is no uniform ranking between the different estimators. In several cases, the simple 3-moment estimator exhibits smaller bias and RMSE than the computationally expensive GMM estimator based on 24 moments. The GMM estimator with 5 moments is also clearly dominated by the 2S-3M estimator. In terms of variance, the GMM estimator with 24 moments performs better than the 2S-3M estimator, but its bias is higher.

Table 1. Moment Estimators: Bias.

$c = 0.3$, $a = 0$, $r_y = 0.5$, $r_w = 0.5$

	$T = 100$			$T = 200$			$T = 500$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	−0.2106	−0.0767	0.0780	−0.1554	−0.0522	0.0901	−0.0805	−0.0233	0.0717
r_y	0.0047	−0.0117	−0.0152	0.0044	−0.0021	−0.0064	0.0023	0.0017	−0.0012
r_w	−0.2988	−0.4016	−0.3315	−0.2384	−0.3643	−0.3070	−0.1360	−0.3210	−0.2218
	$T = 1,000$			$T = 2,000$			$T = 5,000$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	−0.0332	0.0052	0.0186	−0.0204	0.0149	0.0186	−0.0062	0.0191	0.0186
r_y	0.0012	0.0026	0.0009	0.0006	0.0019	0.0009	0.0003	0.0012	0.0009
r_w	−0.0685	−0.3097	−0.0485	−0.0328	−0.3026	−0.0485	−0.0127	−0.2074	−0.0485
	$c = 0.95$, $a = 0.95$, $r_y = 0.5$, $r_w = 0.5$								
	$T = 100$			$T = 200$			$T = 500$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	−0.2490	−0.2904	−0.3400	−0.1576	−0.2652	−0.1327	−0.0921	−0.3209	−0.0257
r_y	0.2063	0.0801	0.0178	0.1754	0.0422	0.0339	0.1379	0.0124	0.0284
r_w	−0.1240	−0.3307	−0.3024	−0.0817	−0.2240	−0.3146	−0.0687	−0.0843	−0.3215
	$T = 1,000$			$T = 2,000$			$T = 5,000$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	−0.0610	−0.3391	−0.0156	−0.0480	−0.3593	0.0071	−0.0299	−0.3813	0.0256
r_y	0.1149	0.0056	0.0253	0.0890	0.0061	0.0262	0.0639	0.0141	0.0305
r_w	−0.0746	−0.0104	−0.3105	−0.0583	0.0676	−0.2856	−0.0683	0.1988	−0.2461

It is interesting to note that Andersen and Sørensen (1996) studied the choice of the number of moments to include in the overidentified estimation procedure and found that it depends critically on sample size. According to these authors, one should exploit additional moment restrictions when the sample size increases. However, this advice does not appear to be compelling in view of the fact that the 2S-3M estimator can exhibit a better performance even for large samples, such as $T = 1000, 2000, 5000$. This may be related to theoretical and simulation findings on IV-based inference suggesting that large numbers of instruments may adversely affect estimator precision and test power; see Buse (1992), Chao and Swanson (2000) and Dufour and Taamouti (2003). In particular, overidentification increases the bias of IV and GMM estimators in finite samples. When 24 moments are used, one needs to estimate $24(24 + 1)/2$ separate entries in the weighting matrix along

Table 2. Moment Estimators: Variance.

$c = 0.3$, $a = 0$, $r_y = 0.5$, $r_w = 0.5$

	$T = 100$			$T = 200$			$T = 500$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.6482	0.3712	0.2914	0.5434	0.3819	0.2986	0.3346	0.3373	0.2947
r_y	0.0019	0.0056	0.0024	0.0010	0.0018	0.0008	0.0005	0.0004	0.0003
r_w	0.0572	0.0423	0.0360	0.0593	0.0557	0.0321	0.0436	0.0827	0.0233
	$T = 1,000$			$T = 2,000$			$T = 5,000$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.1686	0.2103	0.0354	0.0862	0.1027	0.0354	0.0276	0.0304	0.0354
r_y	0.0002	0.0001	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
r_w	0.0200	0.1119	0.0030	0.0092	0.1432	0.0030	0.0029	0.1252	0.0030
	$c = 0.95$, $a = 0.95$, $r_y = 0.5$, $r_w = 0.5$								
	$T = 100$			$T = 200$			$T = 500$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.1796	0.3538	0.3019	0.0751	0.3217	0.1634	0.0343	0.3339	0.0426
r_y	0.1184	0.0815	0.0691	0.0647	0.0458	0.0497	0.0284	0.0177	0.0225
r_w	0.1574	0.0607	0.0633	0.1679	0.0979	0.0481	0.1649	0.1254	0.0325
	$T = 1,000$			$T = 2,000$			$T = 5,000$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.0210	0.3336	0.0414	0.0143	0.3309	0.0172	0.0093	0.2911	0.0003
r_y	0.0143	0.0089	0.0115	0.0073	0.0047	0.0056	0.0040	0.0020	0.0021
r_w	0.1522	0.1484	0.0213	0.1432	0.1546	0.0189	0.1312	0.1709	0.0108

with the sample moments, and the GMM estimator becomes computationally cumbersome. Furthermore, when the values of the autoregressive parameters get close to the boundaries of the domain, this creates numerical instability in estimating the weight matrix, and the situation gets worse in small samples ($T = 100, 200$). When the sample size is small ($T = 100, 200$), RMSE is critically high especially for the autoregressive parameter a; this may be due to the poor behavior of sample moments in small samples.

6. APPLICATION TO STANDARD AND POOR'S PRICE INDEX

In this section, we apply our moment estimator to the daily data on Standard and Poor's Composite Price Index (S&P), over the period

Table 3. Moment Estimators: RMSE.

$c = 0.3,\ a = 0,\ r_y = 0.5,\ r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.8318	0.6138	0.5459	0.7530	0.6205	0.5536	0.5837	0.5818	0.5475
r_y	0.0439	0.0759	0.0513	0.0320	0.0434	0.0295	0.0226	0.0203	0.0199
r_w	0.3827	0.4512	0.3822	0.3408	0.4335	0.3555	0.2491	0.4313	0.2694
	$T = 1,000$			$T = 2,000$			$T = 5,000$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.4118	0.4590	0.4759	0.2942	0.3211	0.3561	0.1662	0.1754	0.1891
r_y	0.0155	0.0140	0.0137	0.0113	0.0101	0.0098	0.0078	0.0070	0.0068
r_w	0.1571	0.4559	0.2000	0.1014	0.4852	0.1393	0.0556	0.4100	0.0732
$c = 0.95,\ a = 0.95,\ r_y = 0.5,\ r_w = 0.5$									
	$T = 100$			$T = 200$			$T = 500$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.4914	0.6617	0.6459	0.3159	0.6351	0.4252	0.2069	0.6607	0.2079
r_y	0.4010	0.2964	0.2634	0.3089	0.2180	0.2255	0.2178	0.1338	0.1527
r_w	0.4155	0.4123	0.3933	0.4176	0.3847	0.3835	0.4116	0.3638	0.3686
	$T = 1,000$			$T = 2,000$			$T = 5,000$		
	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm	3 mm	5 mm	24 mm
a	0.1573	0.6696	0.2041	0.1291	0.6780	0.1314	0.1014	0.6605	0.0312
r_y	0.1659	0.0944	0.1102	0.1234	0.0686	0.0797	0.0900	0.0460	0.0553
r_w	0.3970	0.3852	0.3431	0.3828	0.3988	0.3170	0.3685	0.4586	0.2673

1928–1987. This time series was used by Gallant et al. (1997) to estimate a standard SV model with the efficient method of moments. The data comprise 16,127 daily observations $\{\tilde{y}_t\}_{t=1}^{16,127}$ on adjusted movements of the Standard and poor's Composite Price Index, 1928–1987. The raw series is the Standard and Poor's Composite Price Index (SP), daily, 1928–1987. The raw series is converted to a price movements series, $100[\log(SP_t) - \log(SP_{t-1})]$, and then adjusted for systematic calendar effects, that is, systematic shifts in location and scale due to different trading patterns across days of the week, holidays, and year-end tax trading. This yields a variable we shall denote y_t. The unrestricted estimated value of δ from the data is:

$$\hat{\delta}_T = (0.129,\ 0.926,\ 0.829,\ 0.427)'$$

$$\hat{\sigma}_T = [0.007, 8.10, 1.91, 8.13]'$$

where the method-of-moments estimated value of a corresponds to $\hat{a}_T = 0.926$. We may conjecture that there is some persistence in the data during the period 1928–1987, which has been statistically checked by performing the three standard tests in a companion paper (see Dufour and Valéry, 2005).

7. CONCLUSION

In this paper, we have provided a computationally simple moment estimator available in closed form and derived its asymptotic distribution for the parameters of an SV in a linear regression. Compared with the GMM estimator of Andersen and Sørensen (1996), it demonstrates good statistical properties in terms of bias and RMSE in many situations. Further, it casts doubt on the advice that one should use a large number of moments. In this respect, our just identified estimator underscores that one should not include too many instruments increasing thereby the chance of including irrelevant ones in the estimation procedure. This assertion is documented in the literature on asymptotic theory; see for example, Buse (1992), Chao and Swanson (2000). In particular, overidentification increases bias of IV and GMM estimators in finite samples. Concurring evidence based on finite-sample optimality results and Monte Carlo simulations is also available in Dufour and Taamouti (2003). Further, our closed-form estimator is especially convenient for use in the context of computationally costly inference techniques, such as simulation-based inference methods when asymptotic approximations do not provide reliable inference; see Dufour and Valéry (2005).

SUMMARY

In this paper, we consider the estimation of volatility parameters in the context of a linear regression where the disturbances follow an SV model of order one with Gaussian log-volatility. The linear regression represents the conditional mean of the process and may have a fairly general form, including for example, finite-order autoregressions. We provide a computationally simple two step estimator available in closed form which can be described as follows: (1) the conditional mean model is first estimated by a simple consistent procedure that does take into account the SV structure; (2) using residuals from this preliminary regression, the volatility parameters are then evaluated by a method-of-moment estimator based on three moments for which a simple closed-form expression is derived. Under

general regularity conditions, we show that this two-step estimator is asymptotically normal with the covariance matrix that does not depend on the distribution of the conditional mean estimator. We then study its statistical properties by simulation and compare it with alternative GMM estimators based on more moments. Finally, we provide an application to Standard and Poor's Composite Price Index (1928–1987).

NOTES

1. This feature is exploited in a companion paper (Dufour & Valéry, 2005) where various simulation-based test procedures are developed and implemented.
2. Expressions for the autocorrelations and autocovariances of u_t^2were derived by Taylor (1986, Section 3.5) and Jacquier et al. (1994). The latter authors also provide the higher-order moments $E[|u_t^m|]$, while general formulas for the higher-order cross-moments of an SV process are reported (without proof) by Ghysels et al. (1996). For completeness, we give a relatively simple proof in the Appendix.
3. Andersen and Sørensen (1996) did also consider a moment estimator based on a different set of 3 moments, namely $\mathrm{E}(|u_t|)$, $\mathrm{E}(u_t^2)$ and $\mathrm{E}(|u_t u_{t-1}|)$. A closed-form solution for this estimator was not provided.

ACKNOWLEDGMENTS

The authors thank Silvia Gonçalves, Christian Gouriéroux, Nour Meddahi, an anonymous referee, and the editors Thomas Fomby and Dek Terrell for several useful comments. This work was supported by the Canada Research Chair Program (Chair in Econometrics, Université de Montréal), the Alexander-von-Humboldt Foundation (Germany), the Institut de finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence (program on Mathematics of Information Technology and Complex Systems (MITACS)), the Canada Council for the Arts (Killam Fellowship), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

REFERENCES

Andersen, T. G., Chung, H.-J., & Sørensen, B. E. (1999). Efficient method of moments estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Econometrics*, *91*, 61–87.

Andersen, T. G., & Sørensen, B. E. (1996). GMM estimation of a stochastic volatility model: A Monte Carlo study. *Journal of Business and Economic Statistics*, *14*(3), 328–352.

Buse, A. (1992). The bias of instrumental variables estimators. *Econometrica*, *60*, 173–180.

Carrasco, M., & Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, *18*(1), 17–39.

Chao, J., & Swanson, N. R. (2000). *Bias and MSE of the IV estimators under weak identification*. Technical Report, Department of Economics, University of Maryland.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood. *Journal of Econometrics*, *61*, 375–400.

Danielsson, J., & Richard, J.-F. (1993). Accelerated gaussian importance sampler with application to dynamic latent variable models. *Journal of Applied Econometrics*, *8*, S153–S173.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians* (2nd ed.). Oxford, UK: Oxford University Press.

Dufour, J. -M., & Taamouti, M. (2003). *Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models*. Technical Report, C.R.D.E., Université de Montréal.

Dufour, J. -M., & Valéry, P. (2005). Exact and asymptotic tests for possibly non-regular hypotheses on stochastic volatility models. Technical Report, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de recherche en économie quantitative (CIREQ), Université de Montréal.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1008.

Fuller, W. A. (1976). *Introduction to statistical time series*. New York: Wiley.

Gallant, A. R., Hsieh, D., & Tauchen, G. (1997). Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics*, *81*(1), 159–192.

Ghysels, E., Harvey, A. C., & Renault, E. (1996). Stochastic volatility. In: G. S. Maddala & C. R. Rao (Eds), *Handbook of statistics 14: Statistical method in finance* (pp. 119–183). North-Holland: Amsterdam.

Gonçalves, S., & Kilian, S. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, *123*, 89–120.

Gouriéroux, C. (1997). *ARCH models and financial applications Springer Series in Statistics*. New York: Springer.

Gouriéroux, C., & Monfort, A. (1995a). *Simulation based econometric methods CORE lecture series*. Louvain-la-Neuve: CORE Foundation.

Gouriéroux, C., & Monfort, A. (1995b). Statistics and econometric models, Volumes One and Two, Cambridge, UK: Cambridge University Press.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.

Harvey, A., Ruiz, E., & Shephard, N. (1994). Multivariate stochastic variance models. *Review of Economic Studies*, *61*, 247–264.

Jacquier, E., Polson, N., & Rossi, P. (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Economics and Business Statistics*, *12*, 371–417.

Kim, S., Shephard, N., & Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, *65*, 361–393.

Kuersteiner, G. M. (2001). Optimal instrumental variables estimation for ARMA models. *Journal of Econometrics*, *104*(2), 349–405.

Mahieu, R. J., & Schotman, P. C. (1998). An empirical application of stochastic volatility models. *Journal of Applied Econometrics*, *13*, 333–360.

Melino, A., & Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, *45*, 239–265.

Nelson, D. (1988). *Times series behavior of stock market volatility and returns. PhD thesis*, Department of Economics, Massachusetts: Massachusetts Institute of Technology Cambridge.

Newey, W. K., & McFadden, D. (1994). Large sample estimation and hypothesis testing. In: R. F. Engle & D. L. McFadden (Eds), *Handbook of econometrics*, (Vol. 4, pp. 2111–2245). Amsterdam: North-Holland (chapter 36).

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Palm, F. C. (1996). GARCH models of volatility. In: G. S. Maddala & C. R. Rao (Eds), *Handbook of statistics 14: Statistical methods in finance,* (pp. 209–240). North-Holland: Amsterdam.

Rudin, W. (1976). *Principles of mathematical analysis* (3rd Ed.). New York: McGraw-Hill.

Ruiz, E. (1994). Quasi-maximum likelihood estimation of stochastic variance models. *Journal of Econometrics*, *63*, 284–306.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: D. R. Cox, O. E. Barndorff-Nielsen & D. V. Hinkley (Eds), *Time series models in econometrics, finance and other fields*. London, UK: Chapman & Hall.

Taylor, S. (1986). *Modelling financial time series*. New York: Wiley.

White, H., & Domowitz, I. (1984). Nonlinear regression with dependent observations. *Econometrica*, *52*, 143–161.

Wong, C. K. J. (2002a). *Estimating stochastic volatility: New method and comparison*. PhD thesis, Oxford, UK.

Wong, C. K. J. (2002b). *The MCMC method for an extended stochastic volatility model*. Technical Report, Oxford University, Oxford, UK.

APPENDIX. PROOFS

Proof of Lemma 3.1. If $U \sim N(0,\ 1)$, then $\mathrm{E}(U^{2p+1}) = 0,\ \forall p \in \mathbb{N}$ and $\mathrm{E}(U^{2p}) = (2p)!/[2^p p!]\ \forall\ p \in \mathbb{N}$; see Gouriéroux and Monfort (1995a, Volume 2, p. 518). Under Assumption 2.1,

$$\begin{aligned}\mathrm{E}(u_t^k) &= r_y^k \mathrm{E}(z_t^k)\mathrm{E}[\exp(kw_t/2)]\\ &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!}\exp\left[\frac{k^2}{4} r_w^2/2(1-a^2)\right]\\ &= r_y^k \frac{k!}{2^{(k/2)}(k/2)!}\exp\left[\frac{k^2}{8} r_w^2/(1-a^2)\right] \qquad \text{(A.1)}\end{aligned}$$

where the second equality uses the definition of the Gaussian Laplace transform of $w_t \sim N[0, r_w^2/(1-a^2)]$ and of the moments of the $N(0,1)$ distribution. Let us now calculate the cross-product:

$$\begin{aligned}\mathrm{E}[u_t^k u_{t+m}^l] &= \mathrm{E}\left[r_y^{k+l} z_t^k z_{t+m}^l \exp\left(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}\right)\right] \\ &= r_y^{k+l}\mathrm{E}(z_t^k)\mathrm{E}(z_{t+m}^l)\mathrm{E}\left[\exp\left(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}\right)\right] \\ &= r_y^{k+l}\frac{k!}{2^{(k/2)}(k/2)!}\frac{l!}{2^{(l/2)}(l/2)!}\exp\left[\frac{r_w^2}{8(1-a^2)}(k^2+l^2+2kla^m)\right] \qquad \text{(A.2)}\end{aligned}$$

where $E(w_t) = 0$, $Var(w_t) = r_w^2/(1-a^2)$ and

$$\begin{aligned}\mathrm{Var}\left(k\frac{w_t}{2} + l\frac{w_{t+m}}{2}\right) &= \frac{k^2}{4}\mathrm{Var}(w_t) + \frac{l^2}{4}\mathrm{Var}(w_{t+m}) + 2\frac{k}{2}\frac{l}{2}\mathrm{Cov}(w_t,\ w_{t+m}) \\ &= \frac{r_w^2}{4(1-a^2)}(k^2+l^2+2kla^m) \qquad \text{(A.3)}\end{aligned}$$

Proof of Lemma 3.2. By equations (3.6) and (3.7), we get:

$$\frac{\mathrm{E}(u_t^4)}{[\mathrm{E}(u_t^2)]^2} = 3\exp\left[r_w^2/\left(1-a^2\right)\right] \qquad \text{(A.4)}$$

hence

$$r_w^2/(1-a^2) = \log\left(\frac{\mathrm{E}(u_t^4)}{3(\mathrm{E}(u_t^2))^2}\right) \equiv Q \qquad \text{(A.5)}$$

Inserting $Q \equiv r_w^2/(1-a^2)$ in (3.6) yields

$$r_y = \left(\frac{\mathrm{E}(u_t^2)}{\exp(Q/2)}\right)^{1/2} = \frac{3^{1/4}\mathrm{E}(u_t^2)}{\mathrm{E}(u_t^4)^{1/4}} \qquad \text{(A.6)}$$

From equation (3.8), we have

$$\exp\left(\frac{r_w^2}{(1-a)}\right) = \frac{\mathrm{E}[u_t^2 u_{t-1}^2]}{r_y^4} \qquad \text{(A.7)}$$

which, after a few manipulations, yields

$$1 + a = \frac{[\log(\mathrm{E}[u_t^2 u_{t-1}^2]) - 4\log(r_y)]}{Q} \qquad \text{(A.8)}$$

and

$$a = \frac{[\log(\mathrm{E}[u_t^2 u_{t-1}^2]) - \log(3) - 4\log(\mathrm{E}[u_t^2]) + \log(\mathrm{E}[u_t^4])]}{\log\left(\frac{\mathrm{E}[u_t^4]}{3(\mathrm{E}[u_t^2])^2}\right)} - 1 \tag{A.9}$$

Finally, from (A.5), we get:

$$r_w = \left[(1 - a^2)\log\left(\frac{\mathrm{E}[u_t^4]}{3(\mathrm{E}[u_t^2])^2}\right)\right]^{1/2} \tag{A.10}$$

Proof of Lemma 3.3. Here we derive the covariances of the components of

$$\begin{aligned}\gamma_1(\tau) &= \mathrm{Cov}(X_{1t},\ X_{1,t+\tau}) = \mathrm{E}\{[u_t^2 - \mu_2(\theta)][u_{t+\tau}^2 - \mu_2(\theta)]\}\\ &= \mathrm{E}(u_t^2 u_{t+\tau}^2) - \mu_2^2(\theta) = r_y^4 \mathrm{E}[\exp(w_t + w_{t+\tau})] - \mu_2^2(\theta)\\ &= r_y^4 \exp\left[\frac{r_w^2}{1-a^2}(1 + a^\tau)\right] - \mu_2^2(\theta) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1]\end{aligned} \tag{A.11}$$

where $\gamma = r_w^2/(1 - a^2)$. Similarly,

$$\begin{aligned}\gamma_2(\tau) &= \mathrm{Cov}(X_{2t},\ X_{2,t+\tau}) = \mathrm{E}\{[u_t^4 - \mu_4(\theta)][u_{t+\tau}^4 - \mu_4(\theta)]\}\\ &= \mathrm{E}(u_t^4 u_{t+\tau}^4) - \mu_4^2(\theta) = 9 r_y^8 \mathrm{E}\{\exp[2(w_t + w_{t+\tau})]\} - \mu_4^2(\theta)\\ &= 9 r_y^8 \exp\left[4\frac{r_w^2}{1-a^2}(1 + a^\tau) - \mu_4^2(\theta)\right] = \mu_4^2(\theta)[\exp(4\gamma a^\tau) - 1]\end{aligned} \tag{A.12}$$

Finally,

$$\begin{aligned}\gamma_3(\tau) &= \mathrm{Cov}(X_{3t},\ X_{3,t+\tau}) = \mathrm{E}\{[u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)][u_{t+\tau}^2 u_{t+\tau-1}^2 - \mu_{2,2}(1|\theta)]\}\\ &= \mathrm{E}[u_t^2 u_{t-1}^2 u_{t+\tau}^2 u_{t+\tau-1}^2] - \mu_{2,2}^2(1|\theta)\\ &= r_y^8 \mathrm{E}\exp(w_{t+\tau} + w_{t+\tau-1} + w_t + w_{t-1}) - \mu_{2,2}^2(1|\theta)\\ &= r_y^8 \exp[2(1+a)\gamma]\exp[\gamma(a^{\tau-1} + 2a^\tau + a^{\tau+1})] - \mu_{2,2}^2(1|\theta)\\ &= \mu_{2,2}^2(1|\theta)\{\exp[\gamma(a^{\tau-1} + 2a^\tau + a^{\tau+1})] - 1\}\\ &= \mu_{2,2}^2(1|\theta)\{\exp[\gamma(1+a)^2 a^{\tau-1}] - 1\}\end{aligned} \tag{A.13}$$

for all $\tau \geq 2$.

Proof of Proposition 4.1. The method-of-moments estimator $\hat{\theta}_T(\Omega)$ is solution of the following optimization problem:

$$\min_\theta M_T(\theta) = \min_\theta [\mu(\theta) - \bar{g}_T(\hat{U}_T)]' \hat{\Omega}_T [\mu(\theta) - \bar{g}_T(\hat{U}_T)] \tag{A.14}$$

The first order conditions (F.O.C) associated with this problem are:

$$\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}_T[\mu(\hat{\theta}_T) - \bar{g}_T(\hat{U}_T)] = 0 \tag{A.15}$$

An expansion of the F.O.C above around the true value θ yields

$$\frac{\partial \mu'}{\partial \theta}(\hat{\theta}_T)\hat{\Omega}_T[\mu(\theta) + P(\theta)'(\hat{\theta}_T - \theta) - \bar{g}_T(\hat{U}_T)] = O_p(T^{-1}) \tag{A.16}$$

where, after rearranging the equation,

$$\sqrt{T}[\hat{\theta}_T(\Omega) - \theta] = [P(\theta)\Omega P(\theta)']^{-1}P(\theta)\Omega\sqrt{T}[\bar{g}_T(\hat{U}_T) - \mu(\theta)] + O_p(T^{-1/2}) \tag{A.17}$$

Using Assumptions 4.1 – 4.4, we get the asymptotic normality of $\hat{\theta}_T(\Omega)$ with asymptotic covariance matrix $V(\Omega)$ as specified in Proposition 4.1.

Proof of Proposition 4.2. In order to establish the asymptotic normality of $\sqrt{T}[\bar{g}_T(U_T) - \mu(\theta))]$, we shall use a central limit theorem (C.L.T) for dependent processes (see Davidson, 1994, Theorem 24.5, p. 385). For that purpose, we first check the conditions under which this C.L.T holds. Setting

$$X_t \equiv \begin{pmatrix} u_t^2 - \mu_2(\theta) \\ u_t^4 - \mu_4(\theta) \\ u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta) \end{pmatrix} = g_t(\theta) - \mu(\theta) \tag{A.18}$$

$$S_T = \sum_{t=1}^{T} X_t = \sum_{t=1}^{T} [g_t(\theta) - \mu(\theta)] \tag{A.19}$$

and the subfields $F_t = \sigma(s_t, s_{t-1}, \ldots)$ where $s_t = (y_t, w_t)'$, we need to check three conditions:

(a) $\{X_t, F_t\}$ is stationary and ergodic,

(b) $\{X_t, F_t\}$ is a L_1-mixingale of size -1,

(c) $$\limsup_{T\to\infty} T^{-1/2}\mathrm{E}|S_T| < \infty \tag{A.20}$$

in order to get that $T^{-1/2}S_T = \sqrt{T}[\bar{g}_T(U_T) - \mu(\theta)] \xrightarrow{D} N[0, \Omega_*]$

(a) By propositions 5 and 17 from Carrasco and Chen (2002), we can say that:

(i) if $\{w_{\mathrm{t}}\}$ is geometrically ergodic, then $\{(w_t, \ln|v_t|)\}$ is Markov geometrically ergodic with the same decay rate as the one of $\{w_{\mathrm{t}}\}$;

(ii) if $\{w_{\mathrm{t}}\}$ is stationary β-mixing with a certain decay rate, then $\{\ln|v_t|\}$ is β-mixing with a decay rate at least as fast as the one of $\{w_{\mathrm{t}}\}$.

If the initial value v_0 follows the stationary distribution, $\{\ln|v_t|\}$ is strictly stationary β-mixing with an exponential decay rate. Since this property is preserved by any continuous transformation, $\{v_{\mathrm{t}}\}$ and hence $\{v_t^k\}$ and $\{v_t^k v_{t-1}^k\}$ are strictly stationary and exponential β-mixing. We can then deduce that x_t is strictly stationary and exponential β-mixing.

(b) A mixing zero-mean process is an adapted L_1-mixingale with respect to the subfields F_{t} provided it is bounded in the L_1-norm (see Davidson, 1994, Theorem 14.2, p. 211). To see that $\{X_{\mathrm{t}}\}$ is bounded in the L_1-norm, we note that:

$$\mathrm{E}|v_t^2 - \mu_2(\theta)| \leq \mathrm{E}(|v_t^2| + |\mu_2(\theta)|) = 2\mu_2(\theta) < \infty \tag{A.21}$$

$$\mathrm{E}|v_t^4 - \mu_4(\theta)| \leq 2\mu_4(\theta) < \infty \tag{A.22}$$

$$\mathrm{E}|v_t^2 v_{t-1}^2 - \mu_{2,2}(1|\theta)| \leq 2\mu_{2,2}(1|\theta) < \infty \tag{A.23}$$

We now need to show that the L_1-mixingale $\{X_t, F_t\}$ is of size -1. Since X_{t} is β-mixing, it has mixing coefficients of the type $\beta_n = c\rho^n$, $c > 0$, $0 < \rho < 1$. In order to show that $\{X_t\}$ is of size-1, we need to show that its mixing coefficients $\beta_n = O(n^{-\phi})$, with $\phi > 1$. Indeed,

$$\frac{\rho^n}{n^{-\phi}} = n^{\phi}\exp(n\log\rho) = \exp(\phi\log n)\exp(n\log\rho) = \exp(\phi\log n + n\log\rho) \tag{A.24}$$

It is known that $\lim_{n\to\infty}\phi\log n + n\log\rho = -\infty$ which yields

$$\lim_{n\to\infty} \exp(\phi\log n + n\log\rho) = 0 \tag{A.25}$$

This holds in particular for $\phi > 1$ (see Rudin 1976, Theorem 3.20(d), p. 57).

(c) By the Cauchy–Schwarz inequality, we have:

$$\mathrm{E}|T^{-1/2}S_T| \leq T^{-1/2}||S_T||_2 \tag{A.26}$$

so that (A.20) can be proven by showing that $\lim\ \sup_{T\to\infty} T^{-1}\mathrm{E}(S_T S_T') < \infty$ We shall prove that:

$$\lim_{T\to\infty}\sup\ T^{-1}\mathrm{E}(S_T S_T') = \lim_{T\to\infty}\sup\ \ \mathrm{Var}\left[\frac{1}{\sqrt{T}}S_T\right] < \infty \tag{A.27}$$

(i) First and second components of S_T. $S_{T1} = \Sigma_{t=1}^{T} X_{1,t}$ where $X_{1,t} \equiv u_t^2 - \mu_2(\theta)$. We compute:

$$\begin{aligned}
\mathrm{Var}\left[\frac{1}{\sqrt{T}}S_{T1}\right] &= \tfrac{1}{T}\left[\sum_{t=1}^{T} \mathrm{Var}(X_{1,t}) \ + \sum_{\substack{t=1\\ s\neq t}}^{T} \mathrm{Cov}(X_{1,s},\ X_{1,t})\right] \\
&= \tfrac{1}{T}\left[T\gamma_1(0) + 2\sum_{\tau=1}^{T}(T-\tau)\gamma_1(\tau)\right] \\
&= \gamma_1(0) + 2\sum_{\tau=1}^{T}\left(1-\frac{\tau}{T}\right)\gamma_1(\tau)
\end{aligned} \tag{A.28}$$

where $\gamma \equiv r_w^2/(1-a^2)$. We must prove that $\Sigma_{\tau=1}^{T}\,(1-\frac{\tau}{T})\gamma_1(\tau)$ converge as $T \to \infty$. By Lemma 3.1.5 in Fuller (1976, p. 112), it is sufficient to show that $\Sigma_{\tau=1}^{\infty}\,\gamma_1(\tau)\Sigma_{\tau=1}^{\infty}\,\gamma_1(\tau)$ converge. Using Lemma 3.3. we have

$$\begin{aligned}
\gamma_1(\tau) = \mu_2^2(\theta)[\exp(\gamma a^\tau) - 1] &= \mu_2^2(\theta)\left[1 + \sum_{k=1}^{\infty}\frac{(\gamma a^\tau)^k}{k!} - 1\right] \\
&= \mu_2^2(\theta)\left[\gamma a^\tau \sum_{k=1}^{\infty}\frac{(\gamma a^\tau)^{k-1}}{k!}\right] \\
&= \mu_2^2(\theta)\left[\gamma a^\tau \sum_{k=0}^{\infty}\frac{(\gamma a^\tau)^{k}}{(k+1)!}\right] \leq \mu_2^2(\theta)\gamma a^\tau \sum_{k=0}^{\infty}\frac{(\gamma a^\tau)^k}{k!} \\
&= \mu_2^2(\theta)\gamma a^\tau \exp(\gamma a^\tau)
\end{aligned} \tag{A.29}$$

Therefore, the series

$$\sum_{\tau=1}^{\infty} \gamma_1(\tau) \leq \mu_2^2(\theta)\gamma \sum_{\tau=1}^{\infty} a^{\tau} \exp(\gamma a^{\tau}) \leq \mu_2^2(\theta)\gamma \exp(\gamma a) \sum_{\tau=1}^{\infty} a^{\tau}$$
$$= \mu_2^2(\theta) \frac{a\gamma \exp(\gamma a)}{1-a} < \infty \tag{A.30}$$

converges. We deduce by the Cauchy–Schwarz inequality that

$$\limsup_{T\to\infty} T^{-1/2} \mathrm{E} \left| \sum_{t=1}^{T} [u_t^2 - \mu_2(\theta)] \right| < \infty \tag{A.31}$$

The proof is very similar for the second component of S_T.

(ii) Third component of S_T. Set $S_{T3} = \Sigma_{t=1}^{T} X_{3,t}$ where $X_{3,t} \equiv u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)$. Likewise, we just have to show that $\Sigma_{\tau=1}^{\infty} \gamma_3(\tau) < \infty$ in order to prove that

$$\limsup_{T\to\infty} T^{-1/2} \mathrm{E} \left| \sum_{t=1}^{T} [u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)] \right| < \infty \tag{A.32}$$

By Lemma 3.3 we have for all $\tau \geqslant 2$:

$$\begin{aligned}
\gamma_3(\tau) &= \mu_{2,2}^2(1|\theta)[\exp(\gamma(1+a)^2 a^{\tau-1}) - 1] \\
&= \mu_{2,2}^2(1|\theta) \left\{ 1 + \sum_{k=1}^{\infty} \frac{[\gamma(1+a)^2 a^{\tau-1}]^k}{k!} - 1 \right\} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2 a^{\tau-1}] \sum_{k=1}^{\infty} \frac{[\gamma(1+a)^2 a^{\tau-1}]^{k-1}}{k!} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2 a^{\tau-1}] \sum_{k=0}^{\infty} \frac{[\gamma(1+a)^2 a^{\tau-1}]^{k}}{(k+1)!} \\
&\leq \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2 a^{\tau-1}] \sum_{k=0}^{\infty} \frac{[\gamma(1+a)^2 a^{\tau-1}]^{k}}{k!} \\
&= \mu_{2,2}^2(1|\theta)[\gamma(1+a)^2 a^{\tau-1}] \exp[\gamma(1+a)^2 a^{\tau-1}]
\end{aligned} \tag{A.33}$$

hence,

$$\begin{aligned}\sum_{\tau=1}^{\infty}\gamma_3(\tau) &\leq \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\sum_{\tau=2}^{\infty}a^{\tau-1}\exp[\gamma(1+a)^2a^{\tau-1}]\\ &\leq \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\sum_{\tau=2}^{\infty}a^{\tau-1}\\ &= \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\sum_{\tau=1}^{\infty}a^{\tau}\\ &= \gamma_3(1)+\mu_{2,2}^2(1|\theta)\gamma(1+a)^2\exp[\gamma(1+a)^2a]\frac{a}{1-a}<\infty \qquad \text{(A.34)}\end{aligned}$$

Since $\limsup_{T\to\infty}T^{-1/2}\mathrm{E}|\Sigma_{t=1}^{T}X_t|<\infty$, we can therefore apply Theorem 24.5 of Davidson (1994) to each component S_{Ti}, $i=1,2,3$ of $\mathrm{S_T}$ to state that: $T^{-1/2}S_{Ti}\xrightarrow{D}\mathrm{N}(0,\ \lambda_i)$ and then by the the Cramér–Wold theorem, establish the limiting result for the 3×1-vector S_T using the stability property of the Gaussian distribution, i.e.,

$$T^{-1/2}S_T=T^{-1/2}\sum_{t=1}^{T}X_t=\sqrt{T}[\bar{g}_T(U_T)-\mu(\theta)]\xrightarrow{D}\mathrm{N}_3(0,\ \Omega_*) \qquad \text{(A.35)}$$

where

$$\Omega_*=\lim_{T\to\infty}\mathrm{E}[(T^{-1/2}S_T)^2]=\lim_{T\to\infty}\mathrm{E}\{T[\bar{g}_T(U_T)-\mu(\theta)][\bar{g}_T(U_T)-\mu(\theta)]'\}$$

Proof of Proposition 4.3. The asymptotic equivalence of

$$\sqrt{T}\,[\bar{g}_T(\hat{U}_T)-\mu(\theta)]=\begin{pmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2-\mu_2(\theta)]\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^4-\mu_4(\theta)]\\ \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)]\end{pmatrix} \qquad \text{(A.36)}$$

with $\sqrt{T}\,[\bar{g}_T(U_T)-\mu(\theta)]$ can be established by looking at each component.

1. Component $\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}\,[\hat{u}_t^2-\mu_2(\theta)]$. We have:

$$\begin{aligned}\hat{u}_t^2-u_t^2&=(y_t-x_t'\hat{\beta})^2-(y_t-x_t'\beta)^2\\ &=(\hat{\beta}-\beta)'x_tx_t'(\hat{\beta}-\beta)-2(\hat{\beta}-\beta)'x_tu_t \qquad \text{(A.37)}\end{aligned}$$

We deduce after aggregation:

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2-\mu_2(\theta)] = {} & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^2-\mu_2(\theta)] \\ & + \frac{1}{\sqrt{T}}[\sqrt{T}(\hat{\beta}-\beta)]'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]\sqrt{T}(\hat{\beta}-\beta) \\ & - 2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_t \end{aligned} \tag{A.38}$$

By (4.20), Assumption 4.5 and the Law of Large Numbers (LLN), we have

$$\frac{1}{T}\sum_{t=1}^{T} x_t u_t \to \mathrm{E}(x_t u_t) = \mathrm{E}(x_t)\mathrm{E}(u_t) = 0 \tag{A.39}$$

and equation (A.38) is equivalent to

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2-\mu_2(\theta)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^2-\mu_2(\theta)] + o_p(1) \tag{A.40}$$

asymptotically.

2. Component $\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}[\hat{u}_t^4-\mu_4(\theta)]$. Noting that

$$\begin{aligned}\hat{u}_t^4-u_t^4 = {} & -4(\hat{\beta}-\beta)'x_t u_t^3 + 6(\hat{\beta}-\beta)'x_t u_t^2 x_t'(\hat{\beta}-\beta) \\ & -4(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t u_t \\ & +(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)'x_t x_t'(\hat{\beta}-\beta)\end{aligned} \tag{A.41}$$

we get after aggregation:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^4-\mu_4(\theta)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^4-\mu_4(\theta)] + R_T \tag{A.42}$$

where

$$
\begin{aligned}
R_T \equiv & -4\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_t^3 + \frac{6}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t^2 x_t'\right]\sqrt{T}(\hat{\beta}-\beta) \\
& -4\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t u_t\right] \\
& +\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right]\sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\left[\frac{1}{T}\sum_{t=1}^{T} x_t x_t'\right] \\
& \times \sqrt{T}(\hat{\beta}-\beta)
\end{aligned}
$$

Since $\sqrt{T}(\hat{\beta}-\beta)$, $\frac{1}{T}\Sigma_{t=1}^{T} x_t u_t^2 x_t'$ and $\frac{1}{T}\Sigma_{t=1}^{T} x_t x_t'$ are asymptotically bounded, while by the LLN, $\frac{1}{T}\Sigma_{t=1}^{T} x_t u_t^3$ and $\frac{1}{T}\Sigma_{t=1}^{T} x_t u_t$ converge to zero, we can conclude that R_T is an $o_p(1)$ variable, which yields

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^4-\mu_4(\theta)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^4-\mu_4(\theta)] + o_p(1)
$$

3. Component $\frac{1}{\sqrt{T}}\Sigma_{t=1}^{T}\,[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)]$. Since

$$
\begin{aligned}
\hat{u}_t^2\hat{u}_{t-1}^2 - u_t^2 u_{t-1}^2 = & -2(\hat{\beta}-\beta)'[x_t u_t u_{t-1}^2 + x_{t-1}u_t^2 u_{t-1}] + (\hat{\beta}-\beta)' x_t u_{t-1}^2 x_t'(\hat{\beta}-\beta) \\
& + (\hat{\beta}-\beta)' x_{t-1} u_t^2 x_{t-1}'(\hat{\beta}-\beta) + 4(\hat{\beta}-\beta)' x_t u_t u_{t-1} x_{t-1}'(\hat{\beta}-\beta) \\
& - 2(\hat{\beta}-\beta)' x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)' x_{t-1} u_{t-1} \\
& - 2(\hat{\beta}-\beta)' x_{t-1} x_{t-1}'(\hat{\beta}-\beta)(\hat{\beta}-\beta)' x_t u_t \\
& + (\hat{\beta}-\beta)' x_t x_t'(\hat{\beta}-\beta)(\hat{\beta}-\beta)' x_{t-1} x_{t-1}'(\hat{\beta}-\beta)
\end{aligned} \tag{A.43}
$$

we have:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2\hat{u}_{t-1}^2-\mu_{2,2}(1|\theta)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^2 u_{t-1}^2-\mu_{2,2}(1|\theta)] + R_T \tag{A.44}
$$

where

$$
\begin{aligned}
R_T \equiv & -2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T}[x_t u_t u_{t-1}^2 + x_{t-1}u_t^2 u_{t-1}] \\
& + \frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_{t-1}^2 x_t' \sqrt{T}(\hat{\beta}-\beta) \\
& + \frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_{t-1} u_t^2 x_{t-1}' \sqrt{T}(\hat{\beta}-\beta) \\
& + 4\frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_t u_{t-1} x_{t-1}' \sqrt{T}(\hat{\beta}-\beta) \\
& - 2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_{t-1}u_{t-1} \\
& - 2\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_{t-1} x_{t-1}' \sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t u_t \\
& + \frac{1}{\sqrt{T}}\sqrt{T}(\hat{\beta}-\beta)'\frac{1}{T}\sum_{t=1}^{T} x_t x_t' \sqrt{T}(\hat{\beta}-\beta)\sqrt{T}(\hat{\beta}-\beta)' \\
& \times \frac{1}{T}\sum_{t=1}^{T} x_{t-1} x_{t-1}' \sqrt{T}(\hat{\beta}-\beta)
\end{aligned}
$$

By (4.20), Assumption 4.5 and the LLN applied to $\frac{1}{T}\Sigma_{t=1}^{T}\, x_{t-1}u_t^2 u_{t-1}$, which converges to

$$
\mathrm{E}[x_{t-1}u_t^2 u_{t-1}] = \mathrm{E}[u_t^2 \mathrm{E}(u_{t-1}|F_{t-2})\mathrm{E}(x_{t-1}|F_{t-2})] = 0 \qquad \text{(A.45)}
$$

we deduce that R_T is an $o_p(1)$ variable. Therefore, we have the asymptotic equivalence:

$$
\frac{1}{\sqrt{T}}\sum_{t=1}^{T}[\hat{u}_t^2\hat{u}_{t-1}^2 - \mu_{2,2}(1|\theta)] = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}[u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)] + o_p(1) \qquad \text{(A.46)}
$$

Hence,

$$\begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\hat{u}_t^2 - \mu_2(\theta)] \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\hat{u}_t^4 - \mu_4(\theta)] \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\hat{u}_t^2 \hat{u}_{t-1}^2 - \mu_{2,2}(1|\theta)] \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [u_t^2 - \mu_2(\theta)] \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [u_t^4 - \mu_4(\theta)] \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [u_t^2 u_{t-1}^2 - \mu_{2,2}(1|\theta)] \end{pmatrix} + o_p(1)$$

and

$$\sqrt{T}\,[\bar{g}_T(\hat{U}_T) - \mu(\theta)] = \sqrt{T}\,[\bar{g}_T(U_T) - \mu(\theta)] + o_p(1) \tag{A.47}$$

with $\bar{g}_T(U_T)$ defined as in Eq. (4.13).

THE STUDENT'S *T* DYNAMIC LINEAR REGRESSION: RE-EXAMINING VOLATILITY MODELING

Maria S. Heracleous and Aris Spanos

ABSTRACT

This paper proposes the Student's t Dynamic Linear Regression (St-DLR) model as an alternative to the various extensions/modifications of the ARCH type volatility model. The St-DLR differs from the latter models of volatility because it can incorporate exogenous variables in the conditional variance in a natural way. Moreover, it also addresses the following issues: (i) apparent long memory of the conditional variance, (ii) distributional assumption of the error, (iii) existence of higher moments, and (iv) coefficient positivity restrictions. The model is illustrated using Dow Jones data and the three-month T-bill rate. The empirical results seem promising, as the contemporaneous variable appears to account for a large portion of the volatility.

1. INTRODUCTION

Modeling and forecasting dynamic volatility has been the subject of extensive research among academics and practitioners since the introduction

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 289–319

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20011-7

of the autoregressive conditional heteroskedasticity (ARCH) model by Engle in 1982. Volatility plays a major role in risk management, derivative pricing and hedging, portfolio selection, and the formation of monetary policy. Engle (1982) was the first systematic attempt to capture the phenomenon of volatility clustering, which had been documented in the literature by numerous researchers.[1] This seminal paper gave rise to a tremendous body of research seeking to model volatility by modifying/extending the ARCH model, the most important extension being the Generalized ARCH (GARCH) model; see Li, Ling, and McAleer (2002) and Bauwens, Laurent, and Rombouts (2003) for recent surveys of the literature.

The on-going research on volatility modeling gave rise to a number of interesting issues and problems concerning the modeling of the conditional variance: (i) the choice of an appropriate functional form, (ii) the temporal dependence of the underlying observable process, (iii) the distribution of the error, (iv) existence of moments, (v) extensions to the multivariate framework, and (vi) the inclusion of relevant exogenous variables. Numerous papers have attempted to address these issues and have led to some important contributions in the volatility literature (see Pagan & Schwert, 1990; Bollerslev, 1986; Engle & Bollerslev, 1986; Bollerslev, 1987).

In this paper we focus primarily on the issue of including relevant 'exogenous' variables in the conditional variance, by utilizing the Probabilistic Reduction (PR) approach to specifying statistical models (see Spanos, 1989, 1995). It turns out, however, that the resulting models deal effectively with some of the issues raised above. The potential influence of exogenous variables has been recognized in the literature as early as the mid-1980s. Granger, Robins, and Engle (1984) use price series and not just the past errors squared in the specification of the conditional variance. Also Weiss (1986) suggests a more general form for the conditional variance than the original ARCH, which includes lagged errors, lagged dependent variables and exogenous variables. In recent years the issue of exogenous variables has received new attention. Granger (2002) suggests that "volume or perhaps daily trade would be interesting explanatory variables to include in a model of volatility particularly as both of these variables appear to be forecastable". Also Engle and Patton (2001) point out that financial asset prices do not evolve independently of the market, and expect that other variables may contain information pertaining to the volatility of the series. Evidence in this direction can be found in papers by Engle, Ito, and Lin (1990), Engle and Mezrich (1996), and Glosten, Jagannathan and Runkle (1993).[2] Thus, while the literature has recognized the importance of exogenous variables in explaining volatility, it largely remains an unexplored issue. This is primarily because the ARCH-type models capture volatility through the error term, making the inclusion of other contemporaneous variables hard to justify.

The primary objective of this paper is to develop the Student's *t* Dynamic Linear Regression (St-DLR) model for univariate volatility, in direct analogy to the Normal, DLR model (NDLR) (see Spanos, 1986). It is shown that by viewing the specification of the St-DLR model in the context of the PR approach the inclusion of current exogenous variables in the conditional variance arises naturally. The overall advantage of the PR approach is that it provides a systematic way of specifying volatility models that can simultaneously address the issues (i)–(vi) raised in the financial econometrics literature. The St-DLR is illustrated using daily returns of the Dow Jones industrial price index and the three-month Treasury bill rate.

The remainder of the paper is organized as follows. In the next section, we provide an overview of the PR methodology and highlight the advantages it offers for volatility modeling. In Section 3 we develop the specification of the St-DLR model and discuss its maximum likelihood estimation. To illustrate the St-DLR we provide an empirical example using daily returns of the Dow Jones industrial index and the T-bill rate in Section 4. The estimated St-DLR model is compared to the traditional NDLR, and the Normal GARCH-X (which allows for exogenous variables in the conditional variance). We conclude in Section 5 by summarizing the main theoretical points, the empirical results and their implications.

2. SPECIFICATION OF STATISTICAL MODELS

Statistical model specification is one of the most neglected areas of research in empirical modeling (see Lehmann, 1990; Cox, 1990). As a result, new models are usually specified by modifying the assumptions of the error term of existing models, such as the linear regression model and its many variants. The ARCH model can be viewed as an extension of the Normal, Linear Regression/Autoregression models by modifying the conditional variance to include lagged square errors. The main problem with such an approach is that there is an *infinite* number of ad hoc modifications one can think of, but no *effective strategy* to choose among these possible alternatives. The voluminous literature on ARCH modifications/extensions provides a testimony to the unlimited number of modifications one can think of.[3] Moreover, as the literature on common factor restrictions (see

Hendry, 1995) testifies, the modeling of the error term often imposes unappetizing (and gratuitous) restrictions on the probabilistic structure of observable random variables involved.

How does one choose among the myriad of models thought up in this way? The traditional way to justify 'yet another ARCH model' has been on 'goodness-of-fit' grounds. This is clearly unsatisfactory, however, because showing that one statistical model fits better than another, when both are statistically inadequate, provides very weak grounds for deducing that the model adequately captures the statistical regularities in the data. The only grounds one can empirically justify the adoption of a statistical model is in terms of its *statistical adequacy*; thorough *misspecification testing* of the probabilistic assumptions comprising the model reveal no departures. Without statistical adequacy no reliable inference and thus, no learning from data, can take place. Statistical adequacy, however, pre-supposes that the modeler has a complete set of internally consistent probabilistic assumptions comprising the model. The primary danger with ad hoc modifications of existing models is that one often ends up with: (a) internally inconsistent probabilistic assumptions, (b) hidden assumptions, and (c) many implicit and explicit unappetizing parameter restrictions. For instance, as shown in Spanos (1995), the Normal Autoregressive ARCH model suffers from all three problems.

The PR approach to statistical model specification was designed to give rise to new models which are by construction internally consistent, all assumptions are explicitly stated and need no additional parameter restrictions. In addition, the PR approach renders explicit the connection between statistical model specification and the information conveyed by the data plots of the data in order to aid the choice among different models. This connection turns out to be particularly useful for misspecification testing as well as respecification (choosing an alternative model) purposes. The primary objective of the PR approach is to capture the *systematic features* of the observable phenomenon of interest by modeling the systematic (recurring) information in the observed data. Intuitively, *statistical information* is any *recurring chance regularity pattern* (stylized facts) which can be modeled via probabilistic concepts from the three broad categories described below.

The starting point of the PR approach is the set P of all possible statistical models that could have given rise to the observed data in question, say $\mathbf{Z} := (\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_T)$, where $\mathbf{Z}_t$ is a $m \times 1$ vector. Given that the only complete description of the probabilistic structure of the vector stochastic process $\{\mathbf{Z}_t, t \in \mathbb{T}\}$ can be provided by the joint distribution

$D(\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_T; \boldsymbol{\phi})$ (see Kolmogorov, 1933; Doob, 1953), one can characterize all the elements of P in relation to this distribution. This characterization comes in the form of imposing *reduction assumptions* on $D(\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_T; \boldsymbol{\phi})$ of the process $\{\mathbf{Z}_t, t \in \mathbb{T}\}$, which give rise to the statistical model in question. The reduction (probabilistic) assumptions come from three broad categories:

(D) Distribution (M) Dependence (H) Heterogeneity.

By combining different reduction assumptions one can generate a wealth of statistical models that would have been impossible to 'dream up' otherwise. In addition to ending up with models which are by definition internally consistent and have no additional parameter restrictions, the PR reduction amounts to imposing probabilistic assumptions which 'reduce' the space of models by *partitioning*, (see Spanos, 2005).

The quintessential example is the Linear Regression model, which is characterized by the *reduction assumptions* that the process $\{\mathbf{Z}_t, t \in \mathbb{T}\}, \mathbf{Z}_t := (y_t, \mathbf{X}_t)$ is: (D) *Normal*, (M) *Independent*, (H) *Identically Distributed*; Diagram 1 illustrates the partitioning whose overlap gives rise to the Linear Regression model based on $D(y_t|\mathbf{X}_t; \boldsymbol{\theta})$:

$$D(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_n; \boldsymbol{\phi}) \overset{\text{NIID}}{\rightsquigarrow} \Pi_{t=1}^{n} D(y_t|\mathbf{X}_t; \boldsymbol{\theta}) \tag{1}$$

see Spanos (1989) for the details.

The efficiency of partitioning P should be contrasted with the traditional way of statistical model specification, which attempts to exhaust P using *ad hoc modifications* of, say, the Linear Regression model, i.e. introducing some arbitrary non-linearity and/or heteroskedasticity as well as 'modeling' the error! The PR perspective has some distinct advantages over the traditional

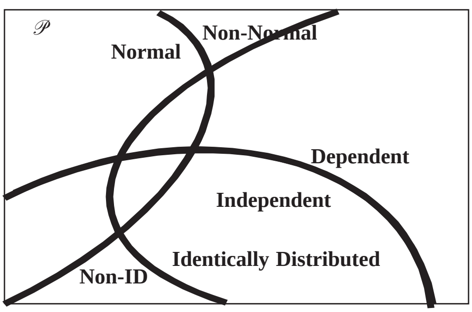


Diagram 1. Model Specification by Partitioning.

way of modifying particular components of existing models without worrying if there exists a stochastic process $\{\mathbf{Z}_t, t \in \mathbb{T}\}$ with a proper joint distribution $D(\mathbf{Z}_1, \mathbf{Z}_2, \cdots, \mathbf{Z}_T; \boldsymbol{\phi})$ underlying such constructs.

The PR approach has been utilized to shed light on a number of issues in econometric modeling, such as simultaneity (Spanos, 1990a), error-autocorrelation (Spanos, 1987; McGuirk & Spanos, 2004), data mining (Spanos, 2000), multicollinearity (Spanos & McGuirk, 2002), and unit root testing (Andreou & Spanos, 2003).

As a prelude to the specification of the St-DLR model we first present the Student's *t* Autoregressive (St-AR) model developed by Spanos (1990b) using the PR perspective; see also McGuirk, Robertson and Spanos (1993). The main reason for presenting this model is to illustrate the PR approach in a simpler case. It also provides us with the opportunity to compare it to the traditional ARCH specification.

2.1. Student's t Autoregressive Model with Dynamic Heteroskedasticity

Speculative price data, such as interest rates, stock returns, and exchange rates often exhibit the well-documented features of: (a) thick tails, (b) non-linear dependence, and (c) bell-shaped symmetry. These features suggest a number of different ways such chance regularity patterns can be modeled, but a most natural way seems to be to use a symmetric, leptokurtic distribution that can accommodate non-linear dependence. The distribution that suggests itself is the Student's *t* distribution. Using the PR approach as the backdrop we can proceed to impose additional probabilistic assumptions on a multivariate Student's *t* distribution of the observables in an attempt to give rise to statistical models appropriate for speculative price data.

In view of the fact that the Normal AR(p) model is characterized by the reduction assumptions that $\{y_t,\ t \in \mathbb{T}\}$ is a (i) (D) Normal, (ii) (M) Markov(p), and (iii) (H) Stationary process (see Spanos, 2001); the reduction assumptions that suggest themselves are that $\{y_t,\ t \in \mathbb{T}\}$ is a (a) (D) Student's t, (b) (M) Markov(p), and (c) (H) Stationary process. The St-AR model with dynamic heteroskedasticity, St-AR($p, p; \nu$), can be specified in terms of the first two conditional moments as follows:

$$y_t = \beta_0 + \sum_{i=1}^{p} \beta_i y_{t-i} + u_t, \quad p > 0, \quad t \in \mathbb{N} \tag{2}$$

where $u_t = y_t - E(y_t|\mathfrak{F}_{t-1})$ is distributed $St(0, \omega_t^2; v+p)$. The conditional variance, ω_t^2 is given by

$$\omega_t^2 = \left[\frac{v\sigma^2}{v+p-2}\right]\left[1 + \sum_{i=1}^{p}\sum_{j=1}^{p}\delta_{ij}\left[y_{t-i}-\mu\right]\left[y_{t-j}-\mu\right]\right] \tag{3}$$

where $\delta_{ij} = 0$ for all $|i-j|>p, \delta_{ij} = \delta_{|i-j|}$, and $\mu = E(y_t)$. Note that $v>2$ is the degrees of freedom parameter and $\mathfrak{F}_{t-1} = \sigma(Y_{t-1}^0)$ the conditioning information set generated by the past history of y_t, (see Spanos, 1990b, 1994 for details).

The reduction assumptions (a)–(c) imply the model assumptions [1]–[5] as shown in Table 1. It is interesting to note that the *autoregressive* and *autoskedastic* functions are: (a) explicitly viewed as based on the first two conditional moments of the same distribution, (b) both are functions of the same information set, Y_{t-1}^0, and (c) the coefficients $(\beta_0, \beta_1, \ldots, \beta_p)$ and $(\delta_0, \delta_1, \ldots, \delta_{p-1})$ are *interrelated* because they are functions of the same variance–covariance matrix of $(y_1, y_2, \ldots, y_T)$. These features can be contrasted to the ARCH family of models.

In order to bring out the crucial differences between the St-AR model and the ARCH model, let us compare the former with the simplest case of AR(1)–ARCH(1):

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t, \quad (u_t|\mathfrak{F}_{t-1}) \sim \mathrm{N}(0, \sigma_t^2)$$

where

$$\sigma_t^2 = \alpha_0 + a_1 u_{t-1}^2$$

Assumptions [2], [4]-[5] have their almost identical counterparts in the ARCH model, but there are crucial differences concerning Assumption [1] and [3]. In Spanos (1995), it was shown that the conditional Normality in

Table 1. Reduction and Probability Model Assumption: Student's *t* AR.

Reduction $\{Y_t\ t \in \mathbb{T}\}$		Model $\{(y_t|\mathfrak{F}_{t-1}), t \in \mathbb{T}\}$
Student's *t*	→	[1] $D(y_t|\mathfrak{F}_{t-1}; \varphi)$ is student's *t*
		[2] $E(y_t|\mathfrak{F}_{t-1}; \varphi) = \beta_0 + \Sigma_{i=1}^{p}\beta_i y_{t-1}$ is linear in y_{t-1}
		[3] $Cov(y_t|\mathfrak{F}_{t-1}) = \omega_t^2$ is heteroskedastic
Markov	→	[4] $\{(u_t|\mathfrak{F}_{t-1}),\ t \in T\}$ is a martingale difference process
Stationary	→	[5] The parameter $(\{\beta_i\}_{i=0}^{p},\ \{\delta_i\}_{i=0}^{p-1},\ \sigma^2)$ are *t*-invariant

the ARCH model leads to internal inconsistencies.[4] Moreover, although both conditional variances are quadratic functions of lagged y_t, the ARCH formulation implicitly imposes some unappetizing restrictions. To see this, let us substitute out the error term:

$$\begin{aligned}\sigma_t^2 &= \alpha_0 + a_1 u_{t-1}^2 = \alpha_0 + a_1(y_{t-1} - \beta_0 - \beta_1 y_{t-2})^2 \\ &= \alpha_0 + a_1((y_{t-1} - \mu) - \beta_1(y_{t-2} - \mu))^2 \\ &= a_0 + a_1(y_{t-1} - \mu)^2 + a_1\beta_1^2(y_{t-2} - \mu)^2 - 2a_1\beta_1(y_{t-1} - \mu)(y_{t-2} - \mu)\end{aligned}$$

When the ARCH is compared with the St-AR(2,2;v) conditional variance, we can see that, in addition to the positivity restrictions $a_0 > 0, a_1 > 0$, it also imposes the implicit parameter restriction:

$$\delta_{11}\delta_{22} = -\delta_{12}^2$$

This restriction is analogous to the common factor restrictions associated with modeling the AR(1) error (see Hendry, 1995). It goes without saying that for $p > 1$ the number of implicit parameter restrictions increases rapidly; (see Spanos 1990b).

The St-AR model was applied by McGuirk et al. (1993) for modeling exchange rates. The authors show that this model dominates the alternative GARCH-type formulations on statistical adequacy grounds. These results have provided the motivation for using the PR approach and the Student's t distribution to develop models that are more complex and more realistic than the St-AR. The aim in this paper is to propose a model that can explain volatility not only in terms of the past history of the series itself but also in terms of other relevant contemporaneous variables, rendering the connection to economic theory much more fruitful. Next, we use the PR approach to develop the St-DLR model.

3. STUDENT'S *T* DLR MODEL

The main objective of this section is twofold. First, to explicitly specify the form and the probabilistic assumptions underlying the St-DLR model. Second, to discuss the maximum likelihood estimation of this model. We begin with the derivation of the St-DLR model.

3.1. Specification

In the context of the PR approach, the operational St-DLR model is specified by imposing certain reduction assumptions on the joint distribution of the vector stochastic process $\{\mathbf{Z}_t, t \in \mathbb{T}s\}$ where $\mathbf{Z}_t := (y_t, \mathbf{X}_t^\top)^\top$. Here y_t is the dependent variable and $\mathbf{X}_t$ is a $(k \times 1)$ vector of exogenous variables. The reduction assumptions are derived from the probabilistic features of the underlying stochastic process and as before can be summarized into three categories: Distributional assumption, Dependence, and (time) Heterogeneity assumptions. For the St-DLR model the reduction assumptions are given by: (1) (D): Student's t, (2) (M): Markov of order p (M(p)), and (3) (H): Second-order stationarity (SS). Using these assumptions the reduction can be performed in the following way:

$$\begin{aligned} D(\mathbf{Z}_{1,}\ldots\mathbf{Z}_T;\boldsymbol{\psi}) &= D_1(\mathbf{Z}_1;\boldsymbol{\phi}_1)\prod_{t=2}^{T} D(\mathbf{Z}_t|\mathbf{Z}_{t-1,}\mathbf{Z}_{t-2},\ldots\mathbf{Z}_1;\boldsymbol{\phi}(t)) \\ &\overset{M(p)+SS}{=} D(\mathbf{Z}_p^1;\boldsymbol{\phi}_1)\prod_{t=p+1}^{T} D(\mathbf{Z}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\phi}_2) \\ &= D(\mathbf{Z}_p^1;\boldsymbol{\phi}_1)\prod_{t=p+1}^{T} D(y_t|\mathbf{Z}_{t-1}^{t-p},\mathbf{X}_t;\boldsymbol{\theta}_1)\cdot D(\mathbf{X}_t|\mathbf{Z}_{t-1}^{t-p};\boldsymbol{\theta}_2), \end{aligned} \tag{4}$$

where $\mathbf{Z}_{t-1}^{t-p} := \left(\mathbf{Z}_{t-1,}\mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-p}\right)$ denotes the past history of $\mathbf{Z}_t$ and $\mathbf{Z}_p^1 := \left(\mathbf{Z}_1, \ldots, \mathbf{Z}_p\right)$ denotes the initial conditions. Also $\boldsymbol{\psi}$, $\boldsymbol{\phi}_{1,}$ and $\boldsymbol{\phi}_2$ denote the parameters in the joint, marginal, and conditional distributions, respectively. The first equality shows that the joint distribution can be decomposed into a product of $(T-1)$ conditional distributions and one marginal distribution. The assumption of Markov(p) changes the conditioning information set to $\left(\mathbf{Z}_{t-1,}\mathbf{Z}_{t-2}, \ldots, \mathbf{Z}_{t-p}\right)$. Also, under the assumption of second-order stationarity, the statistical parameters $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are time invariant. Taken together these assumptions allow us to express the St-DLR model in terms of the first two moments as shown below.

The Student's t Dynamic Linear Regression Model

[A] The conditional mean takes the form

$$y_t = \beta_0^\top \mathbf{X}_t + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i^\top \mathbf{X}_{t-i} + u_t, \quad t \in \mathbb{T} \tag{5}$$

where $u_t = y_t - E(y_t|\mathfrak{F}_{t-1})$ is distributed $St(0, h_t^2; v+m), m = k(p+1)+p$, and $\mathfrak{F}_{t-1} = \sigma(\mathbf{Y}_{t-1}^0, \mathbf{X}_t^0)$. $\mathbf{Y}_{t-1}^0 := (y_{t-1}, y_{t-2}, \ldots, y_1)$, and $\mathbf{X}_t^0 := (\mathbf{X}_t, \mathbf{X}_{t-1}, \ldots, \mathbf{X}_1)$.

[B] The conditional variance h_t^2 is given by

$$h_t^2 := \left[\frac{v\sigma^2}{v+m-2}\right]\left[1 + \left(\mathbf{X}_t^* - \mathbf{M}_t^*\right)^\top \sum_{22}^{-1}\left(\mathbf{X}_t^* - \mathbf{M}_t^*\right)\right] \tag{6}$$

where $\mathbf{X}_t^* = (\mathbf{X}_t^\top, y_{t-1}, \ldots y_{t-p}, \mathbf{X}_{t-1}^\top, \ldots, \mathbf{X}_{t-p}^\top)^\top$, $\mathbf{M}_t^* = (\boldsymbol{\mu}_x^\top, \mu_y \mathbf{1}_{t-p}, \boldsymbol{\mu}_x \mathbf{1}_{t-p})^\top$, are both $(m \times 1)$ vectors when the number of conditioning variables is m. Σ_{22} is the variance–covariance matrix of $\mathbf{X}_t^*$ and v the degree of freedom parameter. The St-DLR can be fully specified in terms of five testable assumptions as shown in Table 2.

Observe that the conditional mean is linear in the conditioning variables as in the case of the NDLR model. The assumption of Student's t however, leads to a conditional variance function which is different from that of the N DLR model. It accommodates both static and dynamic heteroskedasticity. In particular static heteroskedasticity enters the conditional variance specification through the presence of contemporaneous variables, and dynamic heteroskedasticity enters through the lags of the conditioning variables. It is also useful to provide the link between the five model assumptions and the three reduction assumptions discussed earlier. This is shown in Table 3. Note that the reduction assumptions are imposed on the joint distribution and give rise to the model assumptions which relate to the conditional distribution. The relationship between the two sets of

Table 2. The PR Approach: Student's t DLR Specification.

The Student's t Dynamic Linear Regression Model	
$y_t = \beta_0^{\mathrm{T}} \mathrm{X}_t + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i^{\mathrm{T}} \mathrm{X}_{t-i} + u_t, \quad t>p$	
[1] Student's t	$D(y_t \| \sigma(\mathrm{Y}_{t-1}^0, \mathrm{X}_t^0); \theta^*)$ is Student's t
[2] Linearity	$E(y_t \| \sigma(\mathrm{Y}_{t-1}^0, \mathrm{X}_t^0); \theta^*) = \beta^* \mathrm{X}_t^*$ is linear in X_t^*
[3] Heteroskedasticity	$\left[\frac{v\sigma^2}{v+m-2}\right]\left[1 + \left(\mathrm{X}_t^* - \mathrm{M}_t^*\right)^{\mathrm{T}} \Sigma_{22}^{-1} \left(\mathrm{X}_t^* - \mathrm{M}_t^*\right)\right]$ is heteroskedastic
[4] Martingale	$\left\{\left(u_t \| \sigma(\mathrm{Y}_{t-1}^0, \mathrm{X}_t^0)\right), \ t \in \mathbb{T}\right\}$ is a martingale difference process
[5] t-homogeneity	$\theta^* := \left(\alpha_1, \alpha_2, , \ldots, \alpha_p, \beta_0, \beta_1, \ldots \beta_p, \sigma^2, \Sigma_{22}^{-1}\right)$ are not function of $t \in \mathbb{T}$

Table 3. Reduction and Probability Model Assumptions: Student's t DLR.

Reduction $\{Z_t\ t \in \mathbb{T}\}$		Model $\{(y_t \vert \sigma(Y^0_{t-1}), X^0_t = x^0_t;\ \theta^*), t \in \mathbb{T}\}$
Student's t	$\rightarrow$	(i) $D(y_t \vert \sigma(Y^0_{t-1}), X^0_t = x^0_t;\ \theta^*)$ is Student's t (ii) $E(y_t \vert \sigma(Y^0_{t-1}), X^0_t = x^0_t;\ \theta^*) = \beta^* X^*_t$ is linear in X^*_t (iii) $Cov\,(y_t \vert \sigma(Y^0_{t-1}), X^0_t = x^0_t;\ \theta^*) = h^2_t$ is heteroskedastic
Markov	$\rightarrow$	$\{(u_t \vert \sigma(Y^0_{t-1}), X^0_t = x^0_t), t \in \mathbb{T}\}$ is a martingale difference process
Stationary	$\rightarrow$	$\theta^* := (\alpha_1, \alpha_2, \ldots, \alpha_p, \beta_0, \beta_1, \ldots \beta_p, \sigma^2, \sum^{-1}_{22})$ are t-invariant

assumptions is particularly important for misspecification testing and respecification. We now discuss the estimation of this model.

3.2. Maximum Likelihood Estimation

To estimate the St-DLR model typically, one should substitute the functional form of $D(y_t|\mathbf{Z}^{t-p}_{t-1}\ \mathbf{X}_t;\theta_1)$ and $D(X_t|\mathbf{Z}^{t-p}_{t-1};\theta_2)$ in Eq. (4) to obtain the likelihood function of the St-DLR model. The logarithmic form of the likelihood function can then be differentiated to obtain the first order conditions. These can be solved to get the estimators of the parameters of interest. However, instead of following this approach we use an easier technique based on the reparametrization of the joint density. When $\mathbf{Z}_t$ is distributed multivariate Student's t its density takes the form:

$$D(\mathbf{Z}_t;\ \theta) = \frac{\Gamma\left[\frac{1}{2}(v+m+1)\right](\det\Sigma)^{-\frac{1}{2}}}{(\pi v)^{\frac{(m+1)}{2}}\Gamma\left[\frac{1}{2}(v)\right]}\left[1+\frac{1}{v}(\mathbf{Z}_t-\boldsymbol{\mu})^\top\Sigma^{-1}(\mathbf{Z}_t-\boldsymbol{\mu})\right]^{\frac{1}{2}(v+m+1)} \tag{7}$$

Using results from Searle (1982, pp. 258–260) the above equation can be rewritten as

$$D(\mathbf{Z}_t;\ \boldsymbol{\theta}) = \frac{\Gamma\left[\frac{1}{2}(v+m+1)\right]}{(\pi v)^{\frac{m+1}{2}}\Gamma\left[\frac{1}{2}(v)\right]}\sigma^2(\det\Sigma_{22})^{-\frac{1}{2}}\times$$
$$\left[1+\frac{1}{v}\left(\mathbf{X}_t-\boldsymbol{\mu}_2\right)^\top\boldsymbol{\Sigma}^{-1}_{22}\left(\mathbf{X}_t-\boldsymbol{\mu}_2\right)+\frac{1}{v\sigma^2}\left(y_t-\beta_0^\top-\beta^\top\mathbf{X}_t\right)\right]^{\frac{1}{2}(v+m+1)} \tag{8}$$

The advantage of this formulation is that all the parameters of interest appear explicitly. It also facilitates the St-DLR model in the case of Markov variance. By applying Searle's results we can now write the log-likelihood function as:

$$\begin{aligned}\ln L(\theta;\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_T) = {} & T\ln\Gamma\left[\frac{1}{2}(v+m+1)\right] - T\ln(\pi v) \\ & - T\Gamma\left[\frac{1}{2}(v)\right] + \frac{1}{2}T\ln(\det(\mathbf{L}^\top\mathbf{L})) \\ & - \frac{1}{2}T\ln(\sigma^2) - \frac{1}{2}(v+m+1)\sum_{t=2}^{T}\ln\det(\gamma_t) \quad (9)\end{aligned}$$

where

$$\gamma_t = \left[1+\frac{1}{v}(\mathbf{X}_t^* - \mathbf{M}_t^*)^\top\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_t^* - \mathbf{M}_t^*) + \frac{1}{v\,\sigma^2}(y_t - c_0 - \boldsymbol{\beta}_0^\top\mathbf{x}_t - \boldsymbol{\alpha}^\top\mathbf{y}_{t-p} - \boldsymbol{\beta}^\top\mathbf{X}_{t-p}\right],$$

$\mathbf{y}_{t-p} = (y_{t-1},\ \ldots\ y_{t-p})^\top$, $\mathbf{X}_{t-p} := (\mathbf{x}_{t-1},\ \ldots\ \mathbf{x}_{t-p})^\top$ and $\mathbf{L}^\top\mathbf{L} = \boldsymbol{\Sigma}_{22}^{-1}$.

The first-order conditions are given in the appendix, and as we can see, they are non-linear, requiring the use of a numerical procedure. It is worth pointing out at this stage that numerical algorithms are likely to encounter problems with the positive definiteness of $\boldsymbol{\Sigma}_{22}^{-1}$ and σ^2. To address these problems we follow the approach used by Paczkowski (1997) for the St-AR model. To obtain the derivative of σ^2 we take the derivative with respect to σ instead. This ensures that regardless of the value of σ the value of σ^2 will always be positive. We use the same technique for $\boldsymbol{\Sigma}_{22}^{-1}$. One can factorize $\boldsymbol{\Sigma}_{22}^{-1}$ as the product of two matrices, that is $\boldsymbol{\Sigma}_{22}^{-1} = \mathbf{L}^\top\mathbf{L}$ where $\mathbf{L}$ can be a symmetric matrix or a Cholesky factorization. In this case, we assume that $\mathbf{L}$ is a symmetric matrix. These simple modifications allow us to obtain an operational model. The model is estimated using procedures written in GAUSS, (Aptech, 2002).

Before we proceed further we have a few observations about the standard errors of the estimates. Maximum likelihood estimation of this model yields estimates of $\beta_0, \beta, \boldsymbol{\alpha}, \mu_y, \boldsymbol{\mu}_x, \sigma,$ and $\mathbf{L}$. Asymptotic standard errors for these estimates are obtained from the inverse of the final Hessian. To obtain estimates of $\boldsymbol{\Sigma}_{22}^{-1}$ and their standard errors we rely on the invariance property of the maximum likelihood estimators. Note that $\hat{\boldsymbol{\Sigma}}_{22}^{-1}$ can be calculated using the fact that $\hat{\boldsymbol{\Sigma}}_{22}^{-1} = \hat{\mathbf{L}}^\top\hat{\mathbf{L}}$. The standard errors of this matrix can be derived using the δ-method.[5] The Eq. of the estimated variance–covariance matrix (*Cov*) for the distinct elements of $\hat{\boldsymbol{\Sigma}}_{22}^{-1}$ is given by:

$$Cov(\mathrm{vech}\Sigma_{22}^{-1}) = \left[\mathbf{H}\left(\mathbf{I}_{m^2} + \mathbf{K}_{m,m}\right)(\mathbf{I}_m \otimes \hat{\mathbf{L}}^{\top})\mathbf{G}\right]\hat{\Delta}\left[\mathbf{H}\left(\mathbf{I}_{m^2} + \mathbf{K}_{m,m}\right)(\mathbf{I}_m \otimes \hat{\mathbf{L}}^{\top})\mathbf{G}\right]^{\top} \tag{10}$$

where $\hat{\Delta}$ is the estimated variance–covariance matrix of the distinct elements of $\mathbf{L}$ and $\mathbf{K}_{m,m}$ a commutation matrix.[6]

4. EMPIRICAL RESULTS

To illustrate the differences between the traditional ARCH–GARCH formulations and the St-DLR developed in the previous section we now present an empirical example. We begin by describing the data set followed by the empirical results.

4.1. Data

The data used in this section come from a paper by Engle and Patton (2001), consisting of daily closing price data for the Dow Jones Industrial Index over the period August 23, 1988 to August 22, 2000; $T = 3,131$ observations. Log differences of the value of the index are used to compute continuously compounded returns. This data set also includes the three-month US Treasury bill rate over the same period. Following Engle and Patton, the T-bill rate is in percentage form while the Dow Jones returns are expressed in decimal form. Both variables are important for investment and their interrelationship is of interest to policy makers as well as ordinary investors. The Dow Jones returns and its volatility is of interest to a large number of economic agents since it represents about a fifth of the total value of the US stock market, while the T-bill rate is one of the most important indicators of the investment environment in the US economy.

There is a fairly extensive literature in financial economics which investigates the relationship between stock market excess returns and interest rates. It essentially involves two main lines of research. The first one suggests that short-term interest rates can be good proxies for expected inflation, and thus they can be used to indirectly investigate the relationship between inflation and stock market returns. The second one argues that since short-term interest rates embody expectations about inflation, they can be used to predict stock returns as well as volatility in stock returns (see

Breen, Glosten, & Jagannathan, 1989; Giovannini & Jorion, 1989; & Campbell, 1987, inter alia). Another reason for using the T-bill rate in a volatility model is that it can potentially offer some structural or economic explanation for the volatility. It is believed that the T-bill rate is correlated with the cost of borrowing to firms, and may contain information that is relevant to the volatility of the Dow Jones Industrial Index. This idea was explored by Glosten et al. (1993) and Engle and Patton (2001).

As mentioned in Section 2, data plots play a major role in the PR approach. We argue that since t-plots and scatter plots convey information about the marginal and joint distributions of the observables, they can provide one with good insights about the nature of the conditional model that might be appropriate. In order to explore the probabilistic features of the data, we use graphical techniques to bring out the chance regularity patterns exhibited by the data. Fig. 1 represents the standardized time series plot for the Dow Jones returns, and as can be seen, this plot exhibits the common features, which we expect to see in speculative price data. The series seems to be stationary in that the mean over time seems to be constant and the variation around the mean appears to be relatively constant as well. We observe a slight deviation from bell-shaped symmetry. Also, compared to the Normal distribution there seems to be a larger concentration of points

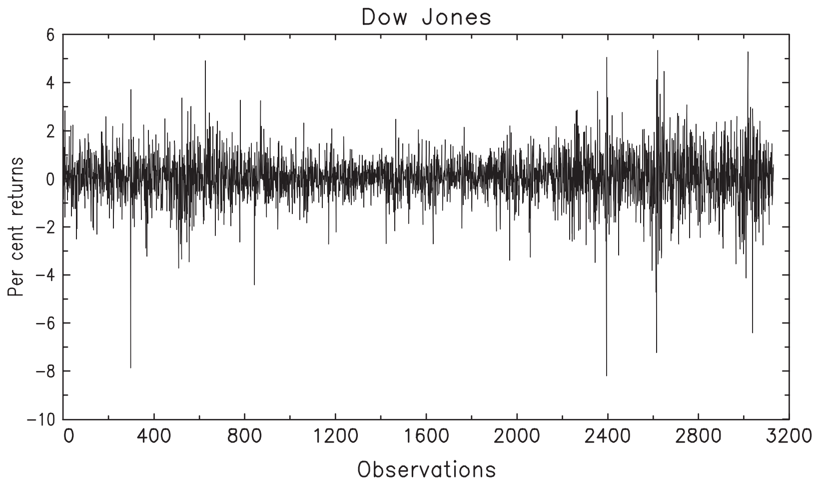


Fig. 1. Standardized *t* Plot (DJ).

around the mean and more 'outliers'. We also observe that there are clusters of small changes followed by clusters of big changes which is symptomatic of second-order temporal dependence.

To further investigate the departure from Normality we also utilize the standardized P–P plots, (for more details on P–P plots see Spanos, 1999, chap. 5). Fig. 2 represents the standardized Normal P–P plot for the Dow Jones daily returns with Cauchy as the reference curve (inverted S). Given the leptokurtic nature of our data we choose the Cauchy distribution as the reference curve because it represents the Student's *t* distribution with one degree of freedom. Fig. 2 clearly indicates that Normality cannot describe the data as shown by the departures in the direction of the Cauchy distribution. The degree of leptokurtosis in the Student's *t* distribution is reflected in the degree of freedom parameter. The difficulty lies in choosing the most appropriate degrees of freedom for the particular data set. One way to approaching this problem is by using the standardized Student's *t* P–P plot (see Spanos, 1999, chap. 5).

Fig. 3 shows the P–P plot for the Student's *t* distribution with three degrees of freedom and the Cauchy distribution as the reference curve. Increasing the degrees of freedom to four seems to provide a much better fit as shown in Fig. 4. To check if this is the best fit we repeat the same exercise for five degrees of freedom in Fig. 5. Through this visual inspection of the data, an educated conjecture is that the Dow Jones returns might be best described by a Student's *t* distribution with four degrees of freedom.

Table 4 reports certain descriptive statistics for the Dow Jones returns. The results seem to strengthen the conjecture regarding non-Normality. In particular, the kurtosis coefficient appears to be significantly greater than 3, which indicates that the unconditional distribution has much thicker tails than the Normal. The skewness coefficient also indicates some degree of asymmetry. We also report some tests on temporal dependence. The χ^2 portmanteau autocorrelation test (Ljung & Box (LB), 1978); indicates the presence of linear temporal dependence. To examine the presence of second-order dependence we use the χ^2 portmanteau test (McLeod & Li (ML), 1983). The *p*-values of this test suggest that there is also considerable second-order dependence in the Dow Jones returns data. In contrast, we note that Engle and Patton (2001) use correlograms to investigate the nature of dependence in the data. They conclude that there is substantial dependence in the volatility of returns but not in the mean. Based on these results they assume a constant conditional mean for their GARCH specification.

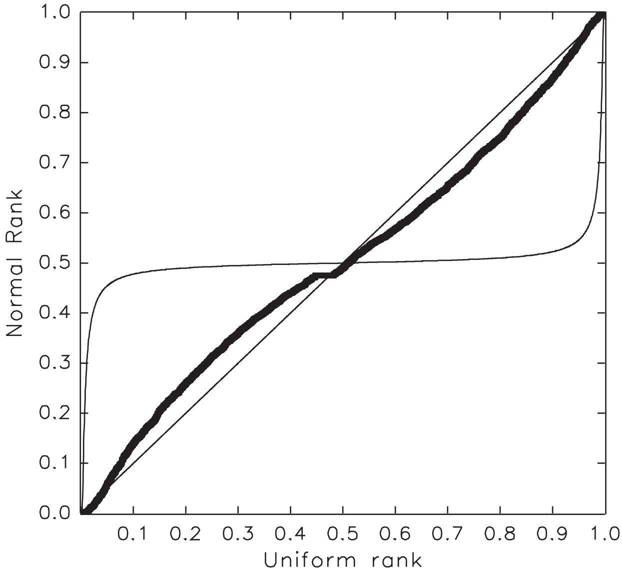


Fig. 2. Normal P–P: Dow Jones Returns data.

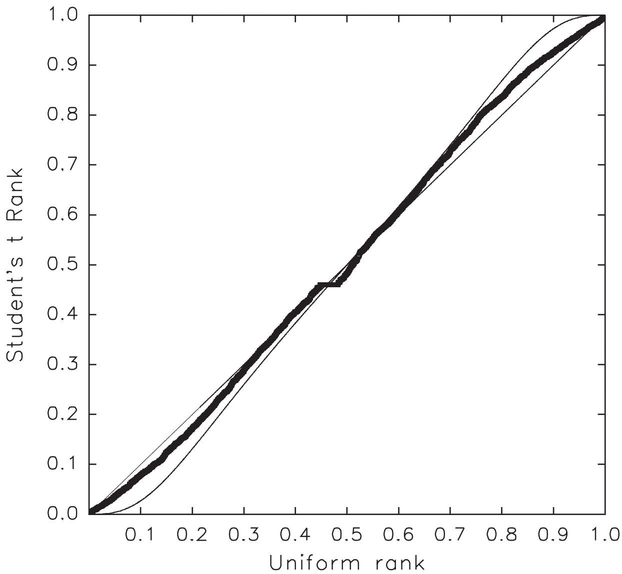


Fig. 3. Student's t P–P plot with df $= 3$.

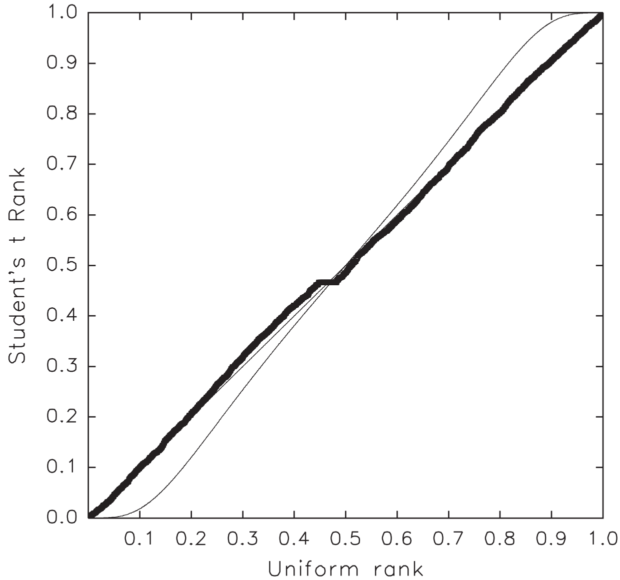


Fig. 4. Student's *t* P–P plot with df = 4

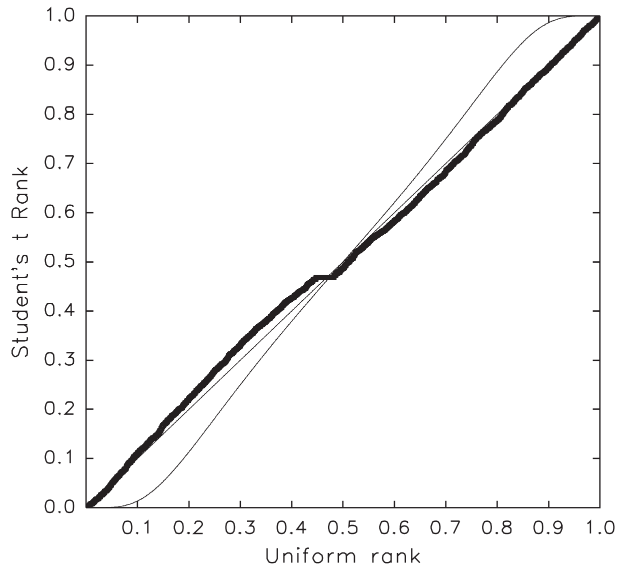


Fig. 5. Student's *t* P–P plot with df = 5.

Table 4. Descriptive Statistics: Dow Jones ($T = 3130$).

	Dow Jones
Mean	0.05504
Std	0.01624
Skewness	−0.52662
Kurtosis	9.04738
P-values for Sample Statistics	
LB (16)	(0.000)*
LB (26)	(0.000)*
ML (16)	(0.000)*
ML (26)	(0.000)*

*Rejection of the null hypothesis at 5% level of significance.

4.2. Empirical Specification and Results

We now provide estimation results based on the traditional ARCH, NGARCH, St-GARCH as well as the St-AR model discussed above. These specifications are representative of the type used in the literature on stock returns and model volatility by using only information on the past history of the series itself.

4.2.1. Volatility Models with no Exogenous Variables

[1] Conditional Normal ARCH (ARCH)

The ARCH conditional variance takes the form:

$$h_t^2 = a_0 + \sum_{i=1}^{m} a_i u_{t-i}^2, \qquad m \geq 1, \qquad \left(u_t/\mathfrak{F}_{t-1}\right) \sim N(0, h_t^2) \tag{11}$$

where the parameter restrictions $a_0 > 0, a_i \geq 0$, are required for ensuring that the conditional variance is always positive. Also $\Sigma_{i=1}^{m} a_i < 1$ is required for the convergence of the conditional variance. The maximum likelihood estimates for the AR(3)–ARCH(5) model are presented in Table 5. It is worth pointing out that an AR(3) – ARCH(9) model was also estimated for the Dow Jones returns. The ARCH parameters in the conditional variance were significant even at the 9th lag. This indicates the problem of long lags commonly observed in the ARCH specification. To overcome this problem the more parsimonious GARCH(p,q) formulation is also estimated.

Table 5. Volatility Models with no Exogenous Variables.

Estimation Results					
	ARCH(5)	NGARCH	t-GARCH		St- AR(3,3,4)
$\hat{\beta}_0$	0.068	0.059	0.076		
	(0.015)*	(0.014)*	(0.013)*		
$\hat{\beta}_1$	0.055	0.030	0.010	$\hat{\beta}_1$	0.028
	(0.020)*	(0.019)	(0.017)		(0.023)
$\hat{\beta}_2$	0.037	0.0119	−0.027	$\hat{\beta}_2$	−0.022
	(0.020)	(0.020)	(0.017)		(0.034)
$\hat{\beta}_3$	−0.047	−0.042	−0.049	$\hat{\beta}_3$	−0.052
	(0.020)*	(0.019)*	(0.017)*		(0.020)*
$\hat{\alpha}_0$		0.007	0.006		
		(0.003)*	(0.002)*		
$\hat{\alpha}_1$	0.448	0.037	0.038	δ_{11}	0.249
	(0.025)*	(0.007)*	(0.008)*		(0.005)*
$\hat{\alpha}_2$	0.101			δ_{21}	−0.007
	(0.021)*				(0.005)
$\hat{\alpha}_3$	0.126			δ_{31}	0.004
	(0.022)*				(0.005)
$\hat{\alpha}_4$	0.042				
	(0.019)*				
$\hat{\alpha}_5$	0.075				
	(0.022)*				
$\hat{\gamma}_1$		0.954	0.956		
		(0.009)*	(0.009)*		
v			5.061		
			(0.474)		

Note: The numbers in the brackets refer to standard errors.
*Rejection of the null hypothesis at 5% level of significance.

[2] Conditional Normal GARCH (NGARCH)

The GARCH conditional variance specification can be written as

$$h_t^2 = a_0 + \sum_{i=1}^{p} a_i u_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-j}^2, \qquad p \geq 1, \qquad q \geq 1 \tag{12}$$

where $(u_t/\mathfrak{F}_{t-1}) \sim N(0, h_t^2)$. Also $a_0 > 0, a_i \geq 0, \gamma_j \geq 0$, and $\Sigma_{i=1}^{p} a_i + \Sigma_{j=1}^{q} \gamma_j < 1$ are required for the positivity of the variance and its stability. We also estimate the traditional GARCH(1,1) model under the assumption that the error follows the Student's *t* distribution (see Bollerslev, 1987).

[3] Conditional Student's *t* GARCH (t-GARCH)

The conditional variance specification takes the same form as in Eq. (12) shown above. Now however, we replace the conditional Normality of the error term with that of the conditional Student's *t* distribution. According to Bollerslev (1987) the distribution of the error term for this specification takes the following form:

$$f(u_t/Y_{t-1}^p) = \frac{\Gamma\left[\frac{1}{2}(v+1)\right]}{\pi^{\frac{1}{2}}\Gamma\left[\frac{1}{2}v\right]}\left[(v-2)h_t^2\right]^{-\frac{1}{2}}\left[1+\frac{u_t^2}{(v-2)h_t^2}\right]^{-\frac{1}{2}(v+1)} \tag{13}$$

[4] Student's *t* AR Model (St-$AR(p,p,v)$)

The St-AR model is specified as shown in Section 2. A St-AR(3,3,4) was found to be statistically adequate for the Dow Jones returns. We should point out that we have chosen an AR(3) representation for the conditional mean so as to capture any weak form of linear dependence that might be present in the Dow Jones returns data. The maximum likelihood estimates for the AR(3)–ARCH(5), AR(3)–NGARCH, and AR(3)-t-GARCH and the St-AR (3,3,4) are summarized in Table 5.

A battery of misspecification tests based on the scaled residuals,[7] $\frac{\hat{u}_t}{\hat{h}_t}$, are also applied to investigate possible departures from the underlying model assumptions and are reported in Table 6. The estimation results in Table 5 indicate that the third lag of the Dow Jones returns is significant in the conditional mean specification for all models. These results suggest that the simple random walk explanation for the stock returns is clearly inappropriate for this data. Further, it calls into question the GARCH model with no conditional mean effects discussed in Engle and Patton (2001). The ARCH estimates for the conditional variance indicate the problem of long lags since there are significant ARCH effects at the fifth lag and even at the ninth lag. The Normal and St-GARCH(1,1) conditional variance estimates indicate the presence of unit roots. This can be interpreted as long memory and in general implies persistence of shocks in the conditional variance. Using the St-AR we find that the first quadratic lag is significant in the conditional variance specification.

It is also interesting to compare the ARCH-type models and the St-AR model using misspecification tests. These tests indicate significant departures from the Normality assumption as expected due primarily to excess kurtosis in the data. This is shown by the *p*-values on the Bera-Jarque (BJ) and

Table 6. Volatility Models with no Exogenous Variables.

P-Values for Misspecification Test Statistics				
	ARCH(5)	NGARCH	t-GARCH	St-*AR* (3,3,4)
BJ	(0.000)*	(0.000)*		
DAP	(0.000)*	(0.000)*		
DS	(0.000)*	(0.000)*	(0.000)*	(0.021)*
DK	(0.000)*	(0.000)*	(0.000)*	(0.000)*
WHITE	(0.000)*	(0.000)*	(0.000)*	(0.944)
LM (2)	(0.029)*	(0.178)	(0.124)	(0.668)
LM (4)	(0.039)*	(0.105)	(0.037)*	(0.298)
LB (16)	(0.053)	(0.209)	(0.046)*	(0.710)
ARCH (2)	(0.940)	(0.120)	(0.437)	(0.302)
ARCH (4)	(0.845)	(0.307)	(0.437)	(0.286)
ML (26)	(0.094)	(0.909)	(0.990)	(0.292)
ML (38)	(0.104)	(0.990)	(0.990)	(0.757)

*Rejection of the null hypothesis at 5% level of significance.

D'Agostino and Pearson (DAP) tests. The D'Agostino kurtosis (DK) and skewness (DS) tests also indicate problems. In addition, the ARCH and the t-GARCH models show departures from the assumptions of static heteroskedasticity and linear dependence as shown by the Langrange Multiplier (LM) test for autocorrelation and the White (W) test for heteroskedasticity. Although, the NGARCH(1,1) seems to perform better than the ARCH model, (shows no problems with linear dependence), it still exhibits problems with Normality and static heteroskedasticity. In contrast, misspecification testing on the St-*AR* model suggests no problems with static heteroskedasticity or any of the other assumptions. Also note that none of the models has problems with the assumption of dynamic heteroskedasticity as indicated by the ARCH test.[8]

Until now we have considered models of univariate volatility involving only the past history of the series itself to explain volatility. Economic theory, however, suggests that other variables such as the T-bill rate may contain information pertaining to the Dow Jones returns and its volatility. We now consider models that can incorporate such information.

4.2.2. Volatility Models with Exogenous Variables

The exogenous variable in this case will be the T-bill rate. Before considering the empirical specifications and the results we present the t-plot of the T-bill rate in Fig. 6. Observe that there has been a big decrease in the T-bill rate

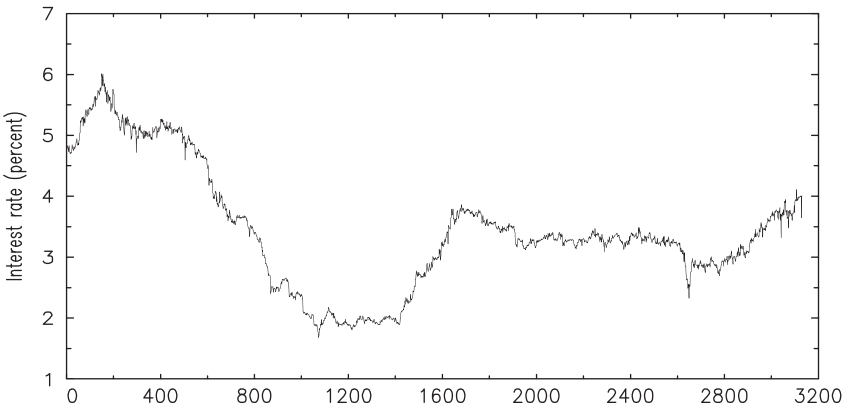


Fig. 6. Standardized *t*-plot (Three month T-bill rate).

from about 6% in late 1980s to around 4% in early 2000. We also provide a table with the descriptive statistics of the variables (Table 7). Following Engle and Patton, the T-bill rate is in percentage form while the Dow Jones returns (in percentage terms) are expressed in decimal form.

We now present the empirical specifications and report estimation results based on the NDLR, the St-DLR and the NGARCH-X models. We begin with the traditional homoskedastic NDLR model. For easier comparisons with the univariate specification for the Dow Jones returns provided above we use three lags and four degrees of freedom in the St-DLR specification. Recall that we had found that a St-AR(3,3,4) was statistically adequate for the Dow Jones.

[5] Normal Dynamic Linear Regression Model (NDLR)
The NDLR model takes the form shown

$$y_t = \beta_0^\top x_t + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i^\top x_{t-i} + u_t,$$
$$\left(u_t/\mathfrak{F}_{t-1}\right) \sim N(0, \sigma^2), \qquad t>p \tag{14}$$

where $\mathfrak{F}_{t-1} = \{\sigma(\mathbf{Y}_{t-1}^0), \mathbf{X}_t^0 = x_t^0\}$ forms the conditioning information set which contains exogenous variables, the past history of those variables as well as the past history of y_t. The important difference between the Normal and St-DLR model presented in Section 3 is in the form of the conditional variance. Observe that in this case the

Table 7. Descriptive Statistics: Dow Jones and T-bill Rate.

	DJ	T-bill
Mean	0.05504	5.38602
Std	0.01624	0.02801
Skewness	−0.52662	0.50425
Kurtosis	9.04738	2.66239

conditional variance, σ^2, is homoskedastic and thus cannot accommodate heteroskedasticity or non-linear dependence present in financial data. Also graphical evidence in this section suggests that the Student's t distribution may be more suitable for capturing the leptokurticity present in the Dow Jones returns data. For these reasons we change the distribution assumption from Normal to Student's t and estimate the St-DLR model. The Student's t specification allows for both static and dynamic heteroskedasticity as well as second-order dependence.

[6] Student's t Dynamic Linear Regression Model (St-*DLR*)

The detailed specification of the St-DLR model can be found in Section 3 (see Table 2). To better understand the interpretation of the parameters in the conditional variance specification consider the case of including one exogenous variable x and two lags in the empirical specification. This implies that $\mathbf{X}_t^* = (x_t, y_{t-1}, y_{t-2}, x_{t-1}, x_{t-2})^\top$. The matrix on the left-hand side of Eq. (15) represents the cross-product terms between the five conditioning variables which are present in the conditional variance equation. The elements of the matrix on the right-hand side represent the corresponding coefficient in front of the cross-product terms. For instance l_{11} is the coefficient of x_t^2 and l_{21} is the coefficient of $x_t y_{t-1}$.

$$\begin{bmatrix} x_t^2 & & & & \\ x_t y_{t-1} & y_{t-1}^2 & & & \\ x_t y_{t-2} & y_{t-1} y_{t-2} & y_{t-2}^2 & & \\ x_t x_{t-1} & y_{t-1} x_{t-1} & y_{t-2} x_{t-1} & x_{t-1}^2 & \\ x_t x_{t-2} & y_{t-1} x_{t-2} & y_{t-2} x_{t-2} & x_{t-1} x_{t-2} & x_{t-2}^2 \end{bmatrix} \rightarrow \begin{bmatrix} l_{11} & & & & \\ l_{21} & l_{22} & & & \\ l_{31} & l_{32} & l_{33} & & \\ l_{41} & l_{42} & l_{43} & l_{44} & \\ l_{51} & l_{52} & l_{53} & l_{54} & l_{55} \end{bmatrix} \tag{15}$$

Table 8 summarizes the results from the Normal DLR (NDLR) and the St-DLR (St-*DLR*) using three lags and four degrees of freedom. Note that we only report the significant coefficients for the conditional

Table 8. Normal DLR and Student's *t* DLR: Dow Jones and T-bill Rate.

Estimation Results				
	NDLR	*t*-stat	St-*DLR*	t-stat
$\hat{\beta}_0$	0.2787	(1.084)	−0.7690*	(−2.381)
$\hat{\alpha}_1$	−0.6544*	(−1.976)	0.0316	(1.639)
$\hat{\alpha}_2$	0.3368	(1.016)	−0.0259	(−1.339)
$\hat{\alpha}_3$	0.0414	(0.1612)	−0.0545*	(−2.841)
$\hat{\beta}_1$	0.0168	(0.9395)	0.6694	(1.506)
$\hat{\beta}_2$	−0.0185	−1.034	0.2224	(0.501)
$\hat{\beta}_3$	−0.0435*	(−2.437)	−0.1181	(−0.370)
μ_y			0.0716	(10.439)
μ_x			5.0910	(182.459)
σ			0.07065	(64.292)
l_{11}			19.5916*	(63.437)
l_{51}			−12.5391*	(−45.158)
l_{61}			−4.2620*	(−18.769)
l_{71}			−3.1263*	(−16.633)
l_{22}			1.4148*	(64.495)
l_{33}			1.4155*	(64.480)
l_{73}			−0.0494*	(−2.184)
l_{44}			1.4058*	(64.248)
l_{55}			27.9140*	(63.276)
l_{65}			−11.5175*	(−37.295)
l_{75}			−4.2104*	(−18.602)
l_{66}			28.0014*	(63.311)
l_{76}			−12.5838*	(−44.913)
l_{77}			19.5633*	(63.503)

Note: The numbers in the brackets refer to *t*-statistics.
*Rejection of the null hypothesis at 5% level of significance.

variance equation.[9] We observe some important differences between the two models. In the NDLR specification the contemporaneous variable (T-bill rate) is not significant, whereas the third lag of the T-bill rate is significant. Also, the first lag for the Dow Jones returns appears to be significant in the conditional mean. The results are very different for the St-DLR model. In this case, the contemporaneous T-bill rate and the third lag of the Dow Jones returns become significant in the conditional mean specification. Also, for the St-DLR a lot of the estimated conditional variance parameters, which involve of the cross-product terms of the Dow Jones (y) and the T-bill rate (x) as well as their lags, are significant. For example $x_t^2(l_{11}), y_{t-1}^2(l_{22})$, and

$y_{t-1}x_{t-3}(l_{73})$ appear to be significant. This indicates that the conditional variance captures a lot of the systematic information (non-linear dependence and heteroskedasticity) present in the data.

It should be pointed out, however, that before making any inferences and policy recommendations we need to ensure that the models are statistically adequate. Misspecification tests such as those discussed previously were applied on the above NDLR specification. They indicate some problems with heteroskedasticity and non-linear dependence as expected.

The St-DLR model is now compared to a conditional heteroskedastic model from the GARCH family, which allows for exogenous variables in the conditional variance specification.

[7] Normal GARCH(1,1) with Exogenous Variables

We present the GARCH model with exogenous variables estimated by Engle and Patton (2001) for comparison purposes. In particular, Engle and Patton (2001) estimate the following specification:

$$y_t = c_0 + u_t \quad (u_t/\mathfrak{F}_{t-1}) \sim N(0, h_t^2) \tag{16}$$

$$h_t^2 = \omega + \sum_{i=1}^{p} a_i u_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-j}^2 + \varphi X_{t-1} \tag{17}$$

where X_{t-1} represents the lagged level of the three-month T-bill rate.

We reproduce the results of Engle and Patton in Table 9; observe that the ARCH parameter is significant. Interestingly, while the coefficient of the

Table 9. Normal GARCH(1,1)-X: Dow Jones and T-bill Rate.

	GARCH (1,1)-X Model	
	GARCH	t-stat
c_0	0.060	(1.818)
ω	−0.001	(0.625)
$\hat{\alpha}_1$	0.044*	(5.5)
$\hat{\gamma}_1$	0.938	(0.1612)
φ	0.003*	(72.15)

Note: The numbers in the brackets refer to t-statistics.
*Rejection of the null hypothesis at 5% level of significance.

T- bill rate is small (0.003), it is quite significant. Engle and Patton claim that the "positive sign suggests that high interest rates are generally associated with higher levels of equity return volatility". This result was documented earlier by Glosten et al. (1993) who also find that the T-bill rate is positively related to equity return volatility.

It is interesting to note, however, that the above GARCH specification is restrictive in the sense that it does not allow for lags of the dependent variable (Dow Jones returns) or the exogenous variable (T-bill) rate in the conditional mean specification. Dynamics are only captured through the conditional volatility equation. In contrast, the St t-DLR results suggest that the current T-bill rate is significant in both the conditional mean and the conditional variance equation as shown in Table 8. Furthermore, the cross products of the T-bill rate with its lags are also significant in the conditional variance equation. These results raise some questions about the appropriateness of augmenting the GARCH model to allow for the effects of other exogenous variables on the volatility equation while ignoring them in the returns equation.

5. CONCLUSION

This paper followed the PR approach to propose the St-DLR model for capturing univariate volatility. The St-DLR model, although specified in terms of statistical information, can easily accommodate economic theory information. The statistical information takes the form of the three reduction assumptions (i) Student's t, (ii) Markov (p), and (iii) Second-Order Stationarity (SS), whereas economic theory information comes in the form of other exogenous variables that may contain information to account for the volatility of some series.

The PR gives rise to a statistical model that is defined in terms of observable random variables and their lags, and not the errors, as is the case with the GARCH-type formulations. This provides a richer framework since it allows us to examine the impact of relevant exogenous variables and their lags on the returns as well as the volatility of the returns. The use of the multivariate Student's t distribution leads to a specification for the conditional variance that is heteroskedastic. This allows us to model the conditional mean and variance parameters jointly leading to gains in efficiency. Moreover, it gives rise to an internally consistent model, which does not require any positivity restrictions on the coefficients or any additional memory assumptions for the stability of the conditional variance.

Empirical results of this paper suggest that the St-DLR model provides a promising way of modeling volatility. The fact that the contemporaneous T-bill rate is significant in the conditional mean and variance equations raises some questions regarding the appropriateness of the existing GARCH specifications, which model volatility through the error term and only allow for exogenous variable in the conditional variance equation. Moreover, the very different estimation results from the GARCH-type models and the St-DLR illustrate the importance of appropriate model choice and indicate the need for formal misspecification tests to check the validity of each specification. Finally, the St-DLR can be seen as an intermediate step to the multivariate framework which takes into account the effect of other variables.

NOTES

1. Volatility clustering refers to the tendency for large (small) price changes to be followed by other large (small) price changes and had been documented as early as Mandelbrot (1963) and Fama (1965).

2. Glosten et al. (1993) also find evidence that indicator variables can help to explain dynamic conditional volatility of equity returns.

3. Rob Engle (2002) is reported to have coined the acronym YAARCH, which stands for Yet Another ARCH model.

4. The inconsistency arises because, in this model, the linearity and heteroskedasticity assumptions contradict the conditional Normality.

5. The δ-method is a convenient technique for approximating the distribution of an arbitrary function of a random variable, when the distribution of that variable is known. This method is useful for finding estimates and their approximate standard errors when the estimable model cannot be expressed in term of the parameters of interest. For more details on this see Oehlert (1992).

6. A commutation matrix K consists of re-arranged rows of an identity matrix, such that $\mathrm{vec}\,X = K_{mn}\mathrm{vec}\,X'$, for any matrix X of dimensionality $(m \times n)$. For details see Magnus and Neudecker (1999).

7. Note that the misspecification tests for the St-AR model are based on the weighted residuals instead, proposed by Spanos (1990b). The weighted residuals are defined as $\hat{u}_t/\hat{\omega}_t - \hat{\omega}_t\varepsilon_t$, where $\hat{u}_t = y_t - \hat{\beta}_0 + \Sigma_{i=1}^{p}\hat{\beta}_i y_{t-i}$, and $\varepsilon_t \sim St(0, 1; v)$ is a simulated standard *i.i.d.* Student's t series. These residuals are a modification of the usual (Pearson) residuals, where the additional term purports to account for the non-linear effects of the conditional variance on y_t.

8. It should not be surprising that the Engle's (1982) ARCH test fails to find any significant effects in the GARCH-type models since these models incorporate dynamic heteroskedasticity. For the St-AR model, however, the ARCH test for dynamic heteroskedasticity represents a modification of Engle's ARCH test (1982)

which includes the squares and cross-products of the regressors under the null. This is consistent with the type of heteroskedasticity implied by the Student's t distribution.

9. The full table is available from the authors upon request.

ACKNOWLEDGMENTS

We would like to thank Anya McGuirk and Sudipta Sarangi for their invaluable advice and suggestions. We would also like to thank the editors, an anonymous referee and Dek Terrell for helpful comments.

REFERENCES

Andreou, E., & Spanos, A. (2003). Statistical adequacy and the testing of trend versus difference stationarity. *Econometric Reviews*, *22*(3), 217–237.

Aptech Systems. (2002). *GAUSS 4.0 Manual*, Maple Valley: Aptech Systems Inc.

Bauwens, L., Laurent, S., & Rombouts, J. V. K. (2003). *Multivariate GARCH models: A Survey*. CORE Discussion Paper 2003/31.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, *69*, 542–547.

Breen, W., Glosten, L. R., & Jagannathan, R. (1989). Economic significance of predictable variation in index returns. *Journal of Finance*, *44*, 1177–1189.

Campbell, J. Y. (1987). Stock returns and term structure. *Journal of Financial Economics*, *18*, 373–399.

Cox, D. R. (1990). Role of models in statistical analysis. *Statistical Science*, *5*, 169–174.

Doob, J. L. (1953). *Stochastic processes*. New York: John Wiley & Sons.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1008.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, *17*, 425–446.

Engle, R. F., & Bollerslev, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, *5*, 1–50 (with discussion).

Engle, R. F., Ito, T., & Lin, W. (1990). Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market. *Econometrica*, *58*, 525–542.

Engle, R. F., & Mezrich, J. (1996). GARCH for groups. *Risk*, *9*, 36–40.

Engle, R. F., & Patton, A. (2001). What good is a volatility model? *Quantitative Finance*, *1*, 237–245.

Fama, E. F. (1965). The behaviour of stock-market prices. *Journal of Business*, *38*, 34–105.

Giovannini, A., & Jorion, P. (1989). The time variation of risk and return in the foreign exchange and stock markets. *Journal of Finance*, *44*, 307–325.

Glosten, L. R., Jaganathan, R., & Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*, 1779–1801.

Granger, C. W. J. (2002). Some comments on risk. *Journal of Applied Econometrics*, *17*, 549–564.

Granger, C. W. J., Robins, R. P., & Engle, R. F. (1984). Wholesale and retail prices: Bivariate time series modeling with forecastable error variances. In: D. A. Belsley & E. Kuh (Eds), *Model Reliability* (pp. 1–17). Cambridge: MIT Press.

Hendry, D. F. (1995). *Dynamic econometrics*. Oxford: Oxford University Press.

Kolmogorov, A. N. (1933). *Foundations of the theory of probability* (2nd English edn.). New York: Chelsea Publishing Co.

Lehmann, E. L. (1990). Model specification: The views of fisher and neyman, and later developments. *Statistical Science*, *5*, 160–168.

Li, W. K., Ling, S., & McAleer, M. (2002). Recent theoretical results for time series models with GARCH errors. *Journal of Economic Surveys*, *16*, 245–269.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *65*, 297–303.

Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics. Wiley Series in Probability and Statistics*. Chichester, England: Wiley.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business*, *36*, 394–419.

McGuirk, A., Robertson, J., & Spanos, A. (1993). Modeling exchange rate dynamics: non-linear dependence and thick tails. *Econometric Reviews*, *12*, 33–63.

McGuirk, A., & Spanos, A. (2004). *Revisiting error autocorrelation correction: Common factor restrictions and granger non-causality*. Virginia Tech Working Paper.

McLeod, A. I., & Li, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, *4*, 269–273.

Oehlert, G. W. (1992). A note on the delta method. *American Statistician*, *46*, 27–29.

Paczkowski, R. (1997). *Monte Carlo examination of static and dynamic student's t regression models*. Ph.D. dissertation, Virginia Polytechnic Institute and State University, etd-0698-184217.

Pagan, A. R., & Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics*, *45*, 267–290.

Searle, S. R. (1982). *Matrix algebra useful for statistics*. New York: Wiley.

Spanos, A. (1986). *Statistical foundations of econometric modeling*. Cambridge: Cambridge University Press.

Spanos, A. (1987). Error autocorrelation revisited: The AR(1) case. *Econometric Reviews*, *6*, 285–294.

Spanos, A. (1989). On re-reading haavelmo: A retrospective view of econometric modeling. *Econometric Theory*, *5*, 405–429.

Spanos, A. (1990a). The simultaneous equation model revisited: Statistical adequacy and identification. *Journal of Econometrics*, *44*, 87–108.

Spanos, A. (1990b). *The student's t autoregeressive model with dynamic heteroskedasticity*. Working Paper VPI & SU.

Spanos, A. (1994). On modeling heteroskedasticity: The student's t and elliptical linear regression models. *Econometric Theory*, *10*, 286–315.

Spanos, A. (1995). On normality and the linear regression model. *Econometric Reviews*, *14*, 195–203.
Spanos, A. (1999). *Probability theory and statistical inference: Econometric modeling with observational data*. Cambridge: Cambridge University Press.
Spanos, A. (2000). Revisiting data mining: 'Hunting' with or without license. *Journal of Economic Methodology*, *7*, 231–264.
Spanos, A. (2001). Time series and dynamic models. In: B. Baltagi (Ed.), *A companion to theoretical econometrics,. (chap. 28)* (pp. 585–609). Oxford: Blackwell Publishers.
Spanos, A. (2005). Structural equation modeling, causal inference and statistical adequacy. In: *The Proceedings of the 12th International Congress of Logic, Methodology and Philosophy of Science.* (Forthcoming) The Netherlands: Kluwer.
Spanos, A., & McGuirk, A. (2002). The problem on near-multicollinearity revisited: erratic versus systematic volatility. *Journal of Econometrics*, *108*, 365–393.
Weiss, A. A. (1986). Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory*, *2*, 107–131.

APPENDIX: FIRST DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTION

The Log-Likelihood Function (LLF) is:

$$\begin{aligned} \ln L\big(\theta; \mathbf{Z}_{1,}\mathbf{Z}_{2,} \ldots \mathbf{Z}_T\big) \quad = \quad & T\ln\Gamma\left[\tfrac{1}{2}(v+m+1)\right] - T\ln(\pi v) \\ & -T\Gamma\left[\tfrac{1}{2}(v)\right] + \tfrac{1}{2}T\ln(\det\big(\mathbf{L}^\top\mathbf{L}\big) \\ & -\tfrac{1}{2}T\ln(\sigma^2) - \tfrac{1}{2}(v+m+1)\sum_{t=2}^{T} \ln\det(\gamma_t) \end{aligned} \tag{A1}$$

where

$$\gamma_t = \left[1 + \tfrac{1}{v}(\mathbf{X}_t^* - \mathbf{M}_t^*)^\top \Sigma_{22}^{-1}(\mathbf{X}_t^* - \mathbf{M}_t^*) + \tfrac{1}{v\sigma^2}(y_t - c_0 - \beta_0^\top \mathbf{x}_t - \alpha^\top \mathbf{y}_{t-p} - \beta^\top \mathbf{X}_{t-p}\right]$$

$$y_{t-p} = (y_{t-1}, \ldots y_{t-p})^\top, \mathbf{X}_{t-p} := (\mathbf{x}_{t-1}, \ldots \mathbf{x}_{t-p})^\top$$

and

$$\mathbf{L}^\top\mathbf{L} = \Sigma_{22}^{-1}$$

To obtain the derivatives we replace c_0 with $\mu_y - \beta_0^\top\mu_x - \alpha\mu_y - \beta^\top\mu_x$. Differentiating the LLF with respect to $\beta_0^\top, \alpha, \beta^\top, \mu_y, \mu_x, \sigma$, and $\mathbf{L}$ we derive the following first order conditions:

1. $\frac{\partial LLF}{\partial \beta_0} = \frac{v+m+1}{v}\sum_{t=1}^{T}\frac{1}{\gamma_t}\frac{u_t}{\sigma^2}\left(\mathbf{x}_t - \mu_x\right)^\top$

2. $\frac{\partial LLF}{\partial \beta} = \frac{v+m+1}{v}\sum_{t=1}^{T}\frac{1}{\gamma_t}\frac{u_t}{\sigma^2}\left(\mathbf{x}_{t-p} - \mu_x\right)^\top$

3. $\frac{\partial LLF}{\partial \alpha} = \frac{v+m+1}{v}\sum_{t=1}^{T}\frac{1}{\gamma_t}\frac{u_t}{\sigma^2}\left(\mathbf{y}_{t-p} - \mu_y\right)^\top$

4. $\frac{\partial LLF}{\partial \mu_y} = \frac{v+m+1}{v}\sum_{t=1}^{T}\frac{1}{\gamma_t}\left[(\mathbf{X}_t^* - \mathbf{M}_t^*)'\mathbf{L}^\top\mathbf{L}(iY) + \frac{u_t}{\sigma^2}(1-\alpha)^\top\right]$

5. $\frac{\partial LLF}{\partial \mu_x} = \frac{v+m+1}{v}\sum_{t=1}^{T}\frac{1}{\gamma_t}\left[\left(\mathbf{X}_t^* - \mathbf{M}_t^*\right)^\top\mathbf{L}^\top\mathbf{L}(iX) + \frac{u_t}{\sigma^2}\left(\beta_0^\top + \beta^\top\right)\right]$

6. $\frac{\partial LLF}{\partial \sigma} = -\frac{T}{\sigma} + \frac{v+m+1}{v}\frac{1}{\sigma^3}\sum_{t=1}^{T}\frac{1}{\gamma_t}u_t^2$

7. $$\frac{\partial LLF}{\partial \mathrm{vech}\mathbf{L}} = T\left[\mathrm{vec}\left(\mathbf{L}^\top\mathbf{L}\right)^{-1}\right]^\top\mathbf{GH}\left(\mathbf{L}^\top \otimes \mathbf{I}_m\right)\mathbf{G} - \frac{v+m+1}{v}\sum_{t=1}^{T}\frac{1}{\gamma_t}\left(\mathbf{X}_t^* - \mathbf{M}_t^*\right)^\top \otimes \left(\mathbf{X}_t^* - \mathbf{M}_t^*\right)^\top\mathbf{GH}\left(\mathbf{L}^\top \otimes \mathbf{I}_m\right)\mathbf{G}$$

where $\mathbf{H}$ is a selector matrix transforming vec($\mathbf{L}'\mathbf{L}$) into vech($\mathbf{L}^\top\mathbf{L}$). The reverse operation is performed by the selector matrix $\mathbf{G}$, and I_m a $(m \times m)$ identity matrix. The matrix iY is an indicator matrix that assigns the value of zero to all other variables in matrix M_t^* apart from μ_y. Similarly, iX is doing the same thing for μ_x.

This page intentionally left blank

ARCH MODELS FOR MULTI-PERIOD FORECAST UNCERTAINTY: A REALITY CHECK USING A PANEL OF DENSITY FORECASTS

Kajal Lahiri and Fushang Liu

ABSTRACT

We develop a theoretical model to compare forecast uncertainty estimated from time-series models to those available from survey density forecasts. The sum of the average variance of individual densities and the disagreement is shown to approximate the predictive uncertainty from well-specified time-series models when the variance of the aggregate shocks is relatively small compared to that of the idiosyncratic shocks. Due to grouping error problems and compositional heterogeneity in the panel, individual densities are used to estimate aggregate forecast uncertainty. During periods of regime change and structural break, ARCH estimates tend to diverge from survey measures.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 321–363

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20012-9

1. INTRODUCTION

Macroeconomic uncertainty as characterized by aggregate forecast density or intervals has long been recognized as an important factor in determining economic outcomes.[1] Friedman (1977), in his Nobel lecture, conjectured that high inflation causes high-inflation uncertainty, and that high uncertainty in turn causes inefficiencies in the economy. The 2003 economics Nobel laureate Engle (1983) wrote while introducing his celebrated autoregressive conditional heteroscedastic (ARCH) model,

> When inflation is unpredictable, risk averse economic agents will incur loss, even if prices and quantities are perfectly flexible in all markets. Inflation is a measure of the relative price of goods today and goods tomorrow; thus, uncertainty in tomorrow's price impairs the efficiency of today's allocation decisions.

Not surprisingly, these developments have triggered much research studying the effects of inflation uncertainty on savings, interest rates, investment, and other variables. More recently, the introduction of inflation targeting regimes in UK, Canada, Sweden, and many other countries during the 1990s has rekindled an enormous interest in generating credible inflation forecasts and the uncertainty associated with the forecasts.[2]

In order to deal with the causes and consequences of forecast uncertainty in a convincing way, it has to be measured correctly. This is particularly important for predictive uncertainty because, unlike forecasts, variance of predictive errors generated by macro time-series models is always unobserved, and hence, can never be evaluated against subsequently observed realizations. During the last 20 years, the most popular approach to estimate forecast uncertainty has been the univariate or multivariate ARCH models initiated by Engle (1982, 1983), who modeled the conditional variance of inflation forecast error as an ARCH process. Bollerslev (1986), Nelson (1991), Glosten, Jagannathan, and Runkle (1993) and others have generalized the basic ARCH model in various directions, and have found many applications in finance and economics.

There are, however, several possible reasons why ARCH-type models may not produce a reliable proxy for inflation uncertainty. First, the reliability of this measure depends on the reliability of the conditional mean and variance functions. This is why Engle (1982, 1983) had emphasized the need for various specification tests in these models. More generally, Rich and Tracy (2003) have questioned the use of conditional variance of a series, which is essentially related to its ex post predictability as a measure of forecast uncertainty, which is related to ex ante confidence of a prediction.

Second, most ARCH-type models assume that the regime characterizing inflation and inflation uncertainty is invariant over time. This assumption may not be true due to structural breaks. Although some researchers have tried to rectify this deficiency (see, for example, Evans, 1991; Evans & Wachtel, 1993), the success of these efforts relies on correctly specifying the structural break itself. Third, as noted by Bomberger (1996) and McNees (1989), forecasters use different models. Thus, uncertainty regarding the model that generates inflation may be a large part of inflation forecast uncertainty. Conditional variance generated by a single model cannot capture this important feature. Finally, Zarnowitz and Lambros (1987) observed that forecast uncertainty depends not only on past forecast errors but also on many future-oriented factors, such as policy changes, macro shocks, data revisions, and many others including purely subjective factors. ARCH-type models ignore this forward-looking behavior of forecasters and other relevant factors that are not included in the time-series specification.[3]

In the US, research on the measurement of forecast uncertainty was facilitated by the foresight of Victor Zarnowitz who under the aegis of American Statistical Association and the National Bureau of Economic Research pioneered a survey in 1968 in which, apart from other point forecasts, subjective density forecasts of real GDP and inflation were elicited from the survey respondents.[4] The average variance of the individual probability distributions can then be directly used as a time-series measure of forecast uncertainty. Studies that used this measure include, among others, Zarnowitz and Lambros (1987), Lahiri and Teigland (1987), Lahiri, Teigland, and Zaporowski (1988), and Batchelor and Dua (1993). This measure is attractive because the variance associated with the density forecast truly represents the uncertainty perceived by an individual forecaster. Batchelor and Dua (1995) and Giordani and Soderlind (2003) argued that this measure reflects the uncertainty a reader of survey forecasts faces if he randomly picks and trusts one of the point forecasts. One question here is whether we should also include forecast disagreement across individuals as a component of aggregate uncertainty. Since forecast disagreement reflects people's uncertainty about models, arguably it should be a part of the aggregate forecast uncertainty.

Due to its ready availability, the variance of the point forecasts (or disagreement) of the survey respondents has been widely used as a proxy for inflation uncertainty. Cukierman and Wachtel (1979, 1982) investigated the relationship between inflation uncertainty and inflation level as well as the variance of the change of relative prices. Although forecast disagreement is easy to compute, its disadvantages are obvious. First, individual biases may be part of forecast disagreement and not necessarily reflect forecast uncertainty. Second, as noted by Zarnowitz and Lambros (1987), it is possible that individual forecasters are extremely certain about their own forecasts, yet the forecasts themselves are substantially dispersed. Conversely, the uncertainty of individual forecasts may be high but point forecasts may be close. Disagreement would then overstate uncertainty in the former case, and understate it in the latter.

There is rather a large literature comparing these three measures of predictive confidence. Engle (1983) was the first to discuss the relationship between the forecast uncertainty derived from the ARCH model and from survey data. He showed that the conditional variance is approximately equal to the average individual forecast error variance. Zarnowitz and Lambros (1987) found that although the average forecast error variance and forecast disagreement were positively correlated, the latter tends to be smaller than the former. Batchelor and Dua (1993, 1996) compared the average individual forecast error variance with a number of proxies including forecast standard deviations from ARIMA, ARCH and structural models of inflation. They found that these proxies are not significantly correlated with the average individual forecast error variance. Other related works include Bomberger (1996), Evans (1991), and Giordani and Soderlind (2003). However, one drawback common to most of these studies is that when comparing different measures of inflation uncertainty the supporting information sets were often not exactly the same. For example, forecasters in a survey often have more information than historical data. They have partial information about current period when making forecasts. So a direct comparison of forecast disagreement or average individual forecast variance with a time-series measure may not be appropriate. Second, due to heterogeneity in forecasts, survey measures often have individual biases. When comparing survey measures with time-series estimates, we should first correct for these compositional effects in the panel. Finally, most existing studies are empirical. We need a theoretical framework to compare estimates of uncertainty based on aggregate time series and survey data on forecast densities.

In this paper, we propose a simple model to characterize the data generating process in Survey of Professional Forecasters (SPF). We argue that the estimated variance of the aggregate density as proposed by Diebold, Tay, and Wallis (1999) and Wallis (2004) gives a reasonably close lower bound on the forecast uncertainty imbedded in a well-specified time-series model. Since the former variance is simply the sum of average variance of the individual densities and the forecast disagreement (cf. Lahiri et al., 1988), our model presents a way of justifying the use of disagreement as a component of macroeconomic uncertainty.[5] Considering that time-series models often have misspecification problems, the sum of forecast disagreement and average individual forecast error variance from surveys may provide a dependable proxy for the aggregate forecast uncertainty over diverse periods. Our empirical analysis shows that the time series and survey measures are remarkably similar during periods of low and stable inflation. But during periods of high and unpredictable inflation, the two tend to diverge.

The remainder of the paper is organized as follows. In Section 2, we introduce our theoretical model to compare time series and survey measures of inflation uncertainty. In Section 3, we show how to extract the measure of forecast uncertainty from the SPF data as defined in Section 2. Sections 4 and 5 describe the SPF forecast density data, and present estimates of inflation forecast uncertainty defined as the sum of average individual forecast error variance and forecast disagreement. In Sections 6 and 7, we empirically compare the survey measures estimated in Section 5 with those estimated from some popular ARCH models. Section 8 concludes the paper.

2. THE MODEL

In this section, we propose a model to study the relationship between measures of forecast uncertainty based on survey data and time-series models. Giordani and Soderlind (2003) and Wallis (2004) have discussed how a panel of density forecasts should be used to generate forecast uncertainty representing the entire economy. However, they did not try to find the relationship between survey and time-series measures of forecast uncertainty.

Consider the prediction of the log of current general price level, p_t, the actual of which is available only in period $t+1$. Suppose the information set of forecaster z is $I_t(z)$. Then the prediction of p_t made by forecaster z can be expressed as $E(p_t|I_t(z))$. In period t, individual forecasters can obtain information from two sources. First, they have knowledge of the past history of the economy in terms of its macroeconomic aggregates. This aggregate information may, for instance, be collected from government publications. Often macro forecasts are generated and made publicly available by government agencies and also by private companies. We assume that this information is common to all forecasters as public information.[6] Although public information does not permit exact inference of p_t, it does determine a "prior" distribution of p_t, common to all forecasters. We assume the distribution of p_t conditional on the public information is $p_t \sim N(\bar{p}_t, \sigma^2)$. Second, since SPF forecasters usually make forecasts in the middle of each quarter for the current and the next year, they will invariably have some private information about p_t. Thus, while reporting the density forecasts, they have partial information about the first half of period t although official report of p_t would still not be available. This private information may take the form of data on relevant variables that are available at higher frequencies (e.g., Consumer Price Index (CPI), unemployment, Blue Chip monthly forecasts, etc.) that individual forecasters may use and interpret at their discretions. More importantly, it will be conditioned by individual forecasters' experience and expertise in specific markets. A key feature of this private information is that it is a mixture of useful information about p_t and some other idiosyncratic information that is orthogonal to p_t.[7] If we use $p_t(z)$ to denote the private information of forecaster z about p_t, then $p_t(z)$ will be a function of p_t and the idiosyncratic factor denoted also by z with $z \sim N(0, \tau_z^2)$. For simplicity, we assume

$$p_t(z) = p_t + z \tag{2.1}$$

By definition, z is distributed independent of p_t and individual-specific information of other forecasters. Although forecaster z knows $p_t(z)$, he is not sure about the relative size of p_t and z. The problem for the individual is how to infer about p_t from the public information and the private information, $p_t(z)$. This is a typical signal extraction problem.[8]

Using (2.1) it is easy to show that the distribution of p_t conditional on both the public and private information is normal with mean

$$E\big(p_t | I_t(z)\big) = E\big(p_t | p_t(z), \bar{p}_t\big) = (1 - \theta_z) p_t(z) + \theta_z \bar{p}_t \tag{2.2}$$

and variance $\theta_z \sigma^2$, where $\theta_z = \tau_z^2 / (\tau_z^2 + \sigma^2)$.

Consider an outside observer (or a central policy maker) who is to forecast the aggregate price level. The information available to him is only public information. He has no private information about period t that can help him to better infer about p_t. However, an unbiased point forecast of p_t can be made with available public information. Under mean squared error loss function, the optimal point forecast should be $\bar{p}_t$, with forecast uncertainty being σ^2. Note that this is just the forecast and forecast uncertainty from aggregate time-series models if the mean and the variance equations were correctly specified.

The situation facing forecaster z is different. He has both the public and private information. Given the information set $I_t(z)$, his point forecast of p_t

is given in (2.2) with the associated forecast uncertainty $\theta_z\sigma^2$. Following Wallis (2004), we can "mix" the individual distributions to obtain an aggregate forecast distribution and its variance. Or equivalently we can calculate the variance as the sum of average individual forecast error variance and forecast disagreement.[9] This can be done more easily in the current context. Using previous notations, we obtain the average of individual forecast uncertainties of p_t as

$$\overline{Var} = \frac{1}{N}\sum_z \mathrm{var}\big(p_t|I_t(z)\big) = \frac{1}{N}\sum_z \theta_z\sigma^2 = \bar{\theta}\sigma^2 \tag{2.3}$$

where N is the number of forecasters.

Since forecasters observe different $p_t(z)$, they will make different point forecasts. The forecast disagreement defined as the variance of point forecasts across individuals is [10]

$$\begin{aligned} Disag &= \tfrac{1}{N}\sum_z \left\{ E\big(p_t|I_t(z)\big) - \tfrac{1}{N}\sum_z E\big(p_t|I_t(z)\big) \right\}^2 \\ &\xrightarrow{p} \tfrac{1}{N}\sum_z E_z\big\{E\big(p_t|I_t(z)\big) - E_z\big[E\big(p_t|I_t(z)\big)\big]\big\}^2 = \tfrac{1}{N}\sum_z \mathrm{var}_z\big[E\big(p_t|I_t(z)\big)\big] \end{aligned} \tag{2.4}$$

By (2.2)

$$\mathrm{var}_z\big[E\big(p_t|I_t(z)\big)\big] = \mathrm{var}_z\big[(1-\theta_z)p_t + (1-\theta_z)z + \theta_z\bar{p}_t\big] = (1-\theta_z)^2\tau_z^2 \tag{2.5}$$

So, the forecast disagreement can be approximated by $\frac{1}{N}\Sigma_z(1-\theta_z)^2\tau_z^2$ and the sum of average individual forecast uncertainty and forecast disagreement is approximately equal to

$$\begin{aligned} \overline{Var} + Disag &= \frac{1}{N}\sum_z \big(\theta_z\sigma^2 + (1-\theta_z)^2\tau_z^2\big) \\ &= \frac{1}{N}\sum_z \left(\frac{\tau_z^2\sigma^2}{\sigma^2+\tau_z^2} + \frac{\sigma^4\tau_z^2}{\big(\sigma^2+\tau_z^2\big)^2} \right) \\ &= \frac{1}{N}\sum_z \frac{2\tau_z^2\sigma^4 + \tau_z^4\sigma^2}{\big(\sigma^2+\tau_z^2\big)^2} \\ &= \sigma^2\frac{1}{N}\sum_z \big[1-(1-\theta_z)^2\big] \le \sigma^2 \end{aligned} \tag{2.6}$$

since $0 \le \theta_z \le 1$.

Eq. (2.6) shows the relationship of the sum of the average individual forecast error variance and the forecast disagreement with the forecast uncertainty estimated from a well-specified time-series model. What we show is that typically the former is expected to be less than the latter.[11] However, we can establish an upper bound on the difference between the two measures empirically based on the findings of previous studies. Cukierman and Wachtel (1979) found that σ^2/τ_z^2, the ratio of the variance of the general price level to the variance of relative prices, was always less than unity during the sample period 1948–1974 and, except for five years, it was less than 0.15. Thus, we see that a value of $\sigma^2/\tau_z^2 < 1$ would imply $\overline{Var} + Disag$ to be at least $0.75\sigma^2$. In fact, $\overline{Var} + Disag$ will be larger than $0.98\sigma^2$ if $\sigma^2/\tau_z^2 < 0.15$. We also computed σ^2/τ_z^2 with the annual Producer Price Index (PPI) data during 1948–1989 from Ball and Mankiw (1995). These authors calculated PPI inflation and weighted standard deviation of relative price change for each year. We calculated σ as the standard error from an AR(1) regression of PPI inflation. We found that during this period σ^2/τ_z^2 is equal to 0.42 on the average, which implies that $\overline{Var} + Disag = 0.91\sigma^2$.

Fig. 1 shows the extent of underestimation of the time series uncertainty by $\overline{Var} + Disag$ as a function of σ^2/τ_z^2. As expected, the smaller this ratio, the closer will be $\overline{Var} + Disag$ to σ^2. For example, if σ^2/τ_z^2 is less than 0.1, $\overline{Var} + Disag > 0.99\sigma^2$. Intuitively, the higher the variance of individual-specific shocks relative to that of the general price level, the lower is the value of $p_t(z)$ in predicting p_t. This is because it is harder for individuals to extract information about p_t from the strong background noise of individual-specific shocks. Eq. (2.2) shows that, under this situation,

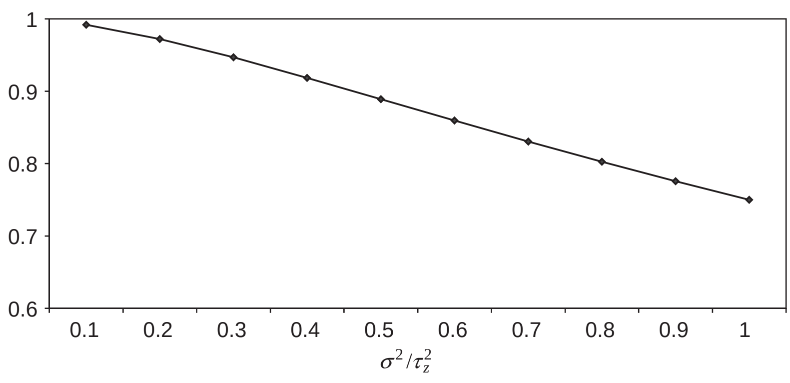


Fig. 1. The Lower Bound for the Sum of Average Individual Uncertainty and Forecast Disagreement.

individuals will rely more heavily on $\bar{p}_t$ to form their forecasts. If $\theta_z = 1$ in (2.2), $p_t(z)$ provides no information about p_t, and individual forecasters will just report $\bar{p}_t$ as their point forecasts. Then $\overline{Var} + Disag$ will be exactly equal to σ^2.

Under the assumption that the variances of individual-specific shocks are the same, i.e., $\tau_1^2 = \tau_2^2 = \cdots = \tau_N^2 = \tau^2$, we can establish the lower bound of $\overline{Var} + Disag$ in another way. First note that if the variances of individual-specific shocks are identical, (2.3), (2.4), and (2.6) can be simplified as $\overline{Var} = \theta\sigma^2$, $Disag = (1-\theta)^2\tau^2$, $\overline{Var} + Disag = (1-(1-\theta)^2)\sigma^2$, where $\theta = \tau^2/(\tau^2 + \sigma^2)$. After litter algebra, we can get

$$\overline{Var} + Disag = (1 - (Disag/\overline{Var})^2)\sigma^2 \tag{2.7}$$

In Section 5, we have estimated the average $Disag/\overline{Var}$ to be 0.27 for the SPF data during the sample period 1968Q4–2003Q4. Substituting this number into (2.7), we get $\overline{Var} + Disag$ to be approximately $0.92\sigma^2$. From all these analysis using both time series and survey data evidence, we may safely conclude that $\overline{Var} + Disag$ is usually between 90% and 100% of the time-series measures of forecast uncertainty. Therefore, when comparing survey and time-series measures of forecast uncertainty, we should include both the average individual forecast uncertainty and forecast disagreement as a measure of collective forecast uncertainty for the economy. This amounts to the use of variance of the aggregate forecast density as advocated by Diebold et al. (1999) and Wallis (2004), see also Giordani and Söderlind (2005).

The model we have just discussed mimics the SPF data set. The survey is mailed four times a year, on the day after the first (preliminary) release of the National Income and Product Accounts (NIPA) data for previous quarter. Forecasters are asked to return the survey before the middle of each forecasting quarter. Therefore, even though forecasters share common information about previous quarters, they also have some private information gathered during the forecasting quarter. During the 45–60 days from the end of previous quarter to when they actually report their forecasts, respondents can obtain partial information about current quarter from many objective sources. Forecasts may also reflect individual experiences in specific markets and forecasters' beliefs regarding the effects of current "news" on future inflation.[12] Thus, the data generating process of SPF is consistent with the simple model we have suggested. This framework provides a guide to compare survey measures from SPF data and time-series measures of forecast uncertainty.

We have assumed that the specification of the time-series model is correct, and that forecast failure does not happen due to structural breaks and policy changes. In reality, however, time-series data-generating processes are subject to breaks and model uncertainty. These possibilities will complicate the simple correspondence between survey and time-series measures of forecast uncertainty. However, our analysis suggests that individually neither the average variance of the individual forecast densities nor the variance of the point forecasts, but their sum, should be used to approximate the forecast uncertainty generated from a correctly specified time-series model.

3. ECONOMETRIC FRAMEWORK

In the previous section, we showed that the forecast uncertainty of an outside observer might be approximated by the sum of average forecast error variance and forecast disagreement, $\overline{Var} + Disag$. In this section, we outline how to calculate $\overline{Var} + Disag$ from SPF data. This issue is not as simple as one might think since the survey respondents are often asked to assign a probability of outcome to various intervals rather than to produce a continuous density function. Thus, how to extract the correct information from the density forecasts data is a problem. The standard approach to calculate the mean and variance from individual density forecasts is as follows (see, for instance, Lahiri & Teigland, 1987; Lahiri et al., 1988).

$$E(F) = \sum_{j=1}^{J} F_j \Pr(j) \quad \text{and} \quad Var(F) = \sum_{j=1}^{J} [F_j - E(F)]^2 \Pr(j)$$

where F_j and $\Pr(j)$ are the midpoint and probability of interval j, respectively. The lowest and highest intervals, which are open, are typically taken to be closed intervals of the same width as the interior intervals.

This approach implicitly assumes that all probability mass is concentrated at the interval midpoints. However, it will lead to the so-called "grouping data error". The standard approach to correcting for grouping data error is "Sheppard's correction" (Stuart & Ord, 1994), which gives the corrected mean the same as the uncorrected mean, but the corrected variance as the uncorrected variance minus 1/12 of the squared bin width. Though popular, there are problems with the Sheppard's correction when applied to SPF data.[13] An alternative proposed by Giordani and Soderlind (2003) is to fit normal distributions to each histogram, and the mean and variance are estimated by minimizing the sum of the squared difference between the survey probabilities and the probabilities for the same intervals implied by the normal distribution. We will follow their approach in this paper.[14]

To obtain an appropriate measure of inflation forecast uncertainty, we need to correct for not only the grouping error, but also the errors due to systematic individual biases in forecast densities. In recent years, the individual heterogeneity in economic forecasts has been increasingly emphasized. For example, Lahiri and Ivanova (1998) and Souleles (2004) use data from the Michigan Index of Consumer Sentiment, and document differences across demographic and other groups in their expectations. Mankiw, Reis, and Wolfers (2003) also document substantial disagreement among economic agents about expected future inflation using survey data from different sources. Mankiw and Reis (2002) propose a "sticky-information" model to explain the variation of disagreement over time. A similar model by Carroll (2003) emphasizes the differential effect of macroeconomic news on household expectations. Disagreement results from the differences across demographic groups in their propensity to pay attention to news reports.

Although the literature has focused mostly on the heterogeneity in point forecasts, some authors have raised the issue of heterogeneity in forecast uncertainty also. For example, Davies and Lahiri (1995, 1999) decompose the variance of forecast errors into variances of individual-specific forecast errors and aggregate shocks. They found significant heterogeneity in the former. Rich and Tracy (2003) also found evidence of statistically significant forecaster fixed effects in SPF density forecasts data. They took this as evidence that forecasters who have access to superior information, or possess a superior ability to process information are more confident in their point forecasts. Ericsson (2003) studies the determinants of forecast uncertainty systematically. He points out that forecast uncertainty depends upon the variable being forecast, the type of model used for forecasting, the economic process actually determining the variable being forecast, the information available, and the forecast horizon. If different forecasters have different information sets and use different forecast models, the anticipated forecast uncertainties will be different across forecasters even if agents are forecasting the same variable at the same forecast horizon. In the following part of this section, we extend the framework of Davies and Lahiri (1995, 1999) to illustrate how to correct for heterogeneity in forecasts.

Let t denote the target period of forecast, h the forecast horizon, or the time left between the time the forecast was made and t, and i the forecaster. Let A_t be the realized value, or the actual, of the forecasted variable, and let A_{th} be the latest realization known to forecasters at the time of forecast. Let

γ_{th} denote the expected change in the actual over the forecast period, and F_{ith} denote the point forecast made by forecaster i at time $t-h$ about the inflation rate in period t. Because we can expect errors in information collection, judgment, calculation, transcription, etc. as well as private information, not all forecasts will be identical. Let us call these differences as "idiosyncratic" error and let μ_{ith} be individual i's idiosyncratic error associated with his forecast for target t made at horizon h. Finally, let ϕ_i be forecaster i's overall average bias. Using these notations, the individual point forecast can be expressed as:

$$F_{ith} = A_{th} + \gamma_{th} + \mu_{ith} + \phi_i \tag{3.1}$$

Note that the presence of systematic bias ϕ_i does not necessarily imply that forecasters are irrational. The reasons of systematic forecast bias may include asymmetric loss function used by forecasters (see Zellner, 1986; Christofferson & Diebold, 1997), or the propensity of forecasters to achieve publicity for extreme opinions (Laster, Bennett, & Geoum, 1999).

Using (3.1), the disagreement among forecasters at time $t-h$ is then

$$\frac{1}{N}\sum_{i=1}^{N}\left(F_{ith} - \bar{F}_{th}\right)^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\phi_i - \bar{\phi} + \mu_{ith} - \bar{\mu}_{th}\right)^2 = \sigma_{\phi}^2 + \sigma_{\mu_{th}}^2 \tag{3.2}$$

where $\bar{F}_{th} = \frac{1}{N}\Sigma_{i=1}^{N} F_{ith}$, $\bar{\phi} = \frac{1}{N}\Sigma_{i=1}^{N}\phi_i$, $\bar{\mu}_{th} = \frac{1}{N}\Sigma_{i=1}^{N}\mu_{ith}$. In (3.2), σ_{ϕ}^2 reflects the dispersion of systematic forecast biases across forecasters and is unrelated to the forecast target and forecast horizon. So it does not reflect the differences in forecasters' views about how price level will change in h periods from now. Only $\sigma_{\mu_{th}}^2$ should be included in the calculation of forecast disagreement for it reflects forecasters' disagreement because of differential information sets (or different forecasting models).

Assuming rationality, and in the absence of aggregate shocks, the actual at the end of period t will be the actual at the end of period $t-h$ (A_{th}) plus the full information anticipated change in the actual from the end of period $t-h$ to the end of period t (γ_{th}). Let the cumulative aggregate shocks occurring from the end of period $t-h$ to the end of period t be represented by λ_{th}. By definition, λ_{th} is the component of the actual that is not anticipated by any forecaster. Then, the actual inflation of period t can be expressed as

$$A_t = A_{th} + \gamma_{th} + \lambda_{th} \tag{3.3}$$

Note that the aggregate shocks from the end of $t-h$ to the end of t (λ_{th}) are comprised of two components: changes in the actual that occurred but were not anticipated, and changes in the actual that were anticipated but did not occur.

Subtracting (3.1) from (3.3) yields an expression for forecast error where forecasts differ from actuals due to individual biases, cumulative aggregate shocks, and idiosyncratic errors.

$$e_{ith} = A_t - F_{ith} = \phi_i + \lambda_{th} + \mu_{ith} \tag{3.4}$$

Then the individual forecast error variance (V_{ith}) can be expressed as

$$V_{ith} = Var_i(e_{ith}) = Var_i\left(\phi_i + \lambda_{th} + \mu_{ith}\right) = Var_i(\lambda_{th}) + Var_i\left(\mu_{ith}\right) \tag{3.5}$$

The first term $Var_i(\lambda_{th})$ measures the perceived uncertainty of aggregate shocks. We allow them to be different across individuals. This is consistent with the model in the previous section in which individuals have different probability forecasts for the aggregate price level due to heterogeneity in information and other reasons. As for the variance of idiosyncratic forecast errors, we assume it is constant over t and h but varies across i. More specifically, we assume $\mu_{ith} \sim N(0, \sigma_i^2)$. This term captures the fixed effects in the individual forecast error variances. From (3.5), only $Var_i(\lambda_{th})$ should be included in the measure of inflation forecast uncertainty. σ_i^2 is unrelated to the target itself and should be excluded from the aggregate measure of inflation uncertainty in order to control for compositional effects in the variances, (see Rich & Tracy, 2003).

As showed in the previous section, the average individual forecast error variance should be included in the measure of aggregate inflation forecast uncertainty. Based on (3.5), it can be calculated as

$$\frac{1}{N}\sum_i V_{ith} = \frac{1}{N}\sum_i Var_i(\lambda_{th}) + \frac{1}{N}\sum_i \sigma_i^2 \tag{3.6}$$

As discussed above, an accurate measure of uncertainty should include only $\underline{\sigma^2_{\lambda_{th}}} = \frac{1}{N}\Sigma_i Var_i(\lambda_{th})$. Thus the measure of forecast uncertainty calculated as $Var + Disag$ is

$$U_{th} = \sigma^2_{\lambda_{th}} + \sigma^2_{\mu_{th}} \tag{3.7}$$

4. DATA

We apply the model developed in the previous section to SPF data set. As noted before, SPF was started in the fourth quarter of 1968 by American Statistical Association and National Bureau of Economic Research and taken over by the Federal Reserve Bank of Philadelphia in June 1990. The respondents are professional forecasters from academia, government, and business. The survey is mailed four times a year, the day after the first release of the NIPA data for the preceding quarter. Most of the questions ask for the point forecasts on a large number of variables for different forecast horizons. A unique feature of SPF data set is that forecasters are also asked to provide density forecasts for aggregate output and inflation. In this study, we will focus on the latter. Before we use this data, we need first to consider several issues, including:

(1) The number of respondents has changed over time. It was about 60 at first and decreased in mid-1970s and mid-1980s. In recent years, the number of forecasters was around 30. So, we have an incomplete panel data.
(2) The number of intervals or bins and their length has changed over time. During 1968Q4–1981Q2 there were 15 intervals, during 1981Q3–1991Q4 there were six intervals, and from 1992Q1 onward there are 10 intervals. The length of each interval was 1 percentage point prior to 1981Q3, then 2 percentage points from 1981Q3 to 1991Q4, and subsequently 1 percentage point again.
(3) The definition of inflation in the survey has changed over time. It was defined as annual growth rate in GNP implicit price deflator (IPD) from 1968Q4 to1991Q4. From 1992Q1 to 1995Q4, it was defined as annual growth rate in GDP IPD. Presently it is defined as annual growth rate of chain-type GDP price index.
(4) Following NIPA, the base year for price index has changed over our sample period. It was 1958 during 1968Q4–1975Q4, 1972 during 1976Q1–1985Q4, 1982 during 1986Q1–1991Q4, 1987 during 1992Q1–1995Q4, 1992 during 1996Q1–1999Q3, 1996 during 1999Q4–2003Q4, and finally 2000 from 2004Q1 onward.
(5) The forecast horizon in SPF has changed over time. Prior to 1981Q3, the SPF asked about the annual growth rate of IPD only in the current year. Subsequently, it asked the annual growth rate of IPD in both the current and following year. However, there are some exceptions. In certain surveys before 1981Q3, the density forecasts referred to the annual growth rate of IPD in the following year, rather than the current year.[15] Moreover, the Federal Reserve Bank of Philadelphia is uncertain about the target years in the surveys of 1985Q1 and 1986Q1. Therefore, even though for most target years, we have eight forecasts with horizons varying from approximately $^1/2$ to $7^1/2$ quarters,[16] for some target years, the number of forecasts is less than eight.

Problems (2)–(4) can be handled by using appropriate actual values and intervals although they may cause the estimation procedure a little more complicated. Following Zarnowitz and Lambros (1987), we focus on the density forecasts for the change from year t-1 to year t that were issued in the four consecutive surveys from the last quarter of year t-1 through the third quarter of year t. The actual horizons for these four forecasts are approximately $4^1/2$, $3^1/2$, $2^1/2$, and $1^1/2$ quarters but we shall refer to them simply as horizons 4, ... , 1. Problems (1) and (5) imply that we will have a lot missing values.

After eliminating observations with missing data, we obtained a total of 4,942 observations over the sample period from 1968Q4 to 2004Q3. For purpose of estimation, we need to eliminate observations for infrequent respondents. Following Zarnowitz and Lambros (1987), we focus on the "regular" respondents who participated in at least 12 surveys during the sample period. This subsample has 4,215 observations in total.

To estimate the model, we also need data on the actual, or realized values of IPD inflation (A_t as in previous section). Since the NIPA data often goes through serious revisions, we need to select the appropriate data for the actual. Obviously, the most recent revision is not a good choice because forecasters cannot forecast revisions occurring many years later. Especially, the benchmark revision often involved adjustment of definitions and classifications, which is beyond the expectation of forecasters. Thus, we choose the first July revisions of the annual IPD data to compute A_t. For example, we compute inflation rate from 1968 to 1969 as

$$A_t = 100 \times \left(\frac{IPD^J_{1969,1} + IPD^J_{1969,2} + IPD^J_{1969,3} + IPD^J_{1969,4}}{IPD^J_{1968,1} + IPD^J_{1968,2} + IPD^J_{1968,3} + IPD^J_{1968,4}} - 1 \right)$$

where $IPD^J_{1969,q}$ is the IPD level in the qth quarter of year 1969 released in July 1970 and $IPD^J_{1968,q}$ is the IPD level in the qth quarter of year 1968 released in July 1969. These are the real-time data available from the Federal Reserve Bank of Philadelphia. See Section 6 for more detailed description of this data set.

5. ESTIMATION

In this section, we describe how to extract inflation forecast uncertainty as defined in Section 3 from SPF data set. By (3.4), forecast error can be decomposed into three parts:

$$e_{ith} = A_t - F_{ith} = \lambda_{th} + \phi_i + \mu_{ith}$$

Note that, μ_{ith} is uncorrelated over i, t, h, and λ_{th} is uncorrelated over t. In addition, they both have zero mean and are uncorrelated with each other under the assumption of rational expectation. Following Davies and Lahiri (1999), an estimate of systematic forecast bias can then be derived as follows:

$$\hat{\phi}_i = \frac{1}{TH}\sum_t \sum_h (A_t - F_{ith}) \tag{5.1}$$

The mean bias across all forecasters is then

$$\hat{\bar{\phi}} = \frac{1}{N}\sum_i \hat{\phi}_i \tag{5.2}$$

and the variance of individual bias across forecasters is

$$\hat{\sigma}^2_{\phi} = \frac{1}{N-1}\sum_i \left(\hat{\phi}_i - \hat{\bar{\phi}}\right)^2 \tag{5.3}$$

Note that, the composition of forecasters varies over both t and h, which implies that σ^2_{ϕ} also changes over t and h.

Using (3.2), we can obtain estimates of inflation forecast disagreement as

$$\hat{\sigma}^2_{\mu_{th}} = \frac{1}{N}\sum\nolimits_{i=1}^{N} \left(F_{ith} - \bar{F}_{th}\right)^2 - \hat{\sigma}^2_{\phi} \tag{5.4}$$

Note that (5.4) is computed over the subsample of "regular" respondents. So, we implicitly assume that including infrequent respondents will not change the estimate of forecast disagreement appreciably.

Next we consider the estimation of the average individual forecast error variance as defined in Section 3. From (3.5), we have

$$V_{ith} = Var_i(\lambda_{th}) + \sigma^2_i \tag{5.5}$$

As argued in previous section, σ^2_i should not be included in the aggregate measure of forecast uncertainty. The distribution of forecasters over time is not random. Some forecasters participated only in the early period of the survey, others participated only in the later period. Following Rich and Tracy (2003), we regress the variances of individual densities on a set of year dummy variables and a set of individual dummy variables for each forecast horizon. The estimated respondent fixed effects reflect the extent to which a particular respondent's inflation uncertainty systematically differs from the average adjusting for the years that the respondent participated in the survey. By subtracting out these fixed-effect estimates from the respondent's inflation uncertainty estimates, we can control for the changes in the composition of the survey. Then by applying (3.6), we obtain the average individual forecast error variance corrected for the "composition" effect, i.e. $\hat{\sigma}^2_{\lambda_{th}}$.

Given the estimates of forecast disagreement and average individual forecast error variance, we could compute the inflation forecast uncertainty based on (3.7) as

$$\hat{U}_{th} = \hat{\sigma}^2_{\lambda_{th}} + \hat{\sigma}^2_{\mu_{th}} \tag{5.6}$$

However, when plotting inflation forecast uncertainty and its two components with different horizons pooled together, one more adjustment seems reasonable. There is a possibility that the same shocks occurring at two different horizons relative to the same target will have different effects on forecaster's confidence about his forecast. Forecasters may be more uncertain about the effect of a shock when the horizon is long. This type of "horizon effects" should be removed when we want to examine how forecast uncertainty varies over time continuously over quarters. To remove the horizon effects, we regressed $Var_i(\lambda_{th})$ on horizon dummies. More specifically,

$$Var_i(\lambda_{th}) = \alpha_i + D_1\delta_{i1} + D_2\delta_{i2} + D_3\delta_{i3} + v_{ith} \tag{5.7}$$

where $D_h = \begin{cases} 1 \text{ if observation is for horizon } h,\ h = 1,\ 2,\ 3. \\ 0 \text{ otherwise} \end{cases}$

Note that δ_{ih} measures the difference between forecast error variance with horizon of four quarters and those with others. This implies that forecasts with horizon of four quarters are the benchmark for comparison. We converted forecasts with horizons less than four quarters to forecasts with 4-quarter horizon. Since we are only interested in the average horizon effects, we formulate (5.7) as a random coefficients model. Specifically, we assume that

$$(\alpha_i, \delta_{i2}, \delta_{i3}, \delta_{i4}) = (\alpha, \delta_2, \delta_3, \delta_4) + v_i \quad \text{with} \quad E(v_i) = 0 \quad E\left(v_i v_i'\right) = \Gamma \tag{5.8}$$

Models (5.7) and (5.8) can be estimated[17] using the method formulated in Greene (2000). Then the estimated average 4-quarter equivalent individual forecast error variance and inflation forecast uncertainty are $\hat{\sigma}^2_{\lambda_{th}} - \hat{\delta}_h$ and $\hat{\sigma}^2_{\lambda_{th}} - \hat{\delta}_h + \hat{\sigma}^2_{\mu_{th}}$ respectively.[18]

Figs. 2a and b present the estimated uncertainty ($\hat{U}_{th} - \hat{\delta}_h$) and its two components, the average variance ($\hat{\sigma}^2_{\lambda_{th}} - \hat{\delta}_h$) and the disagreement ($\hat{\sigma}^2_{\mu_{th}}$),

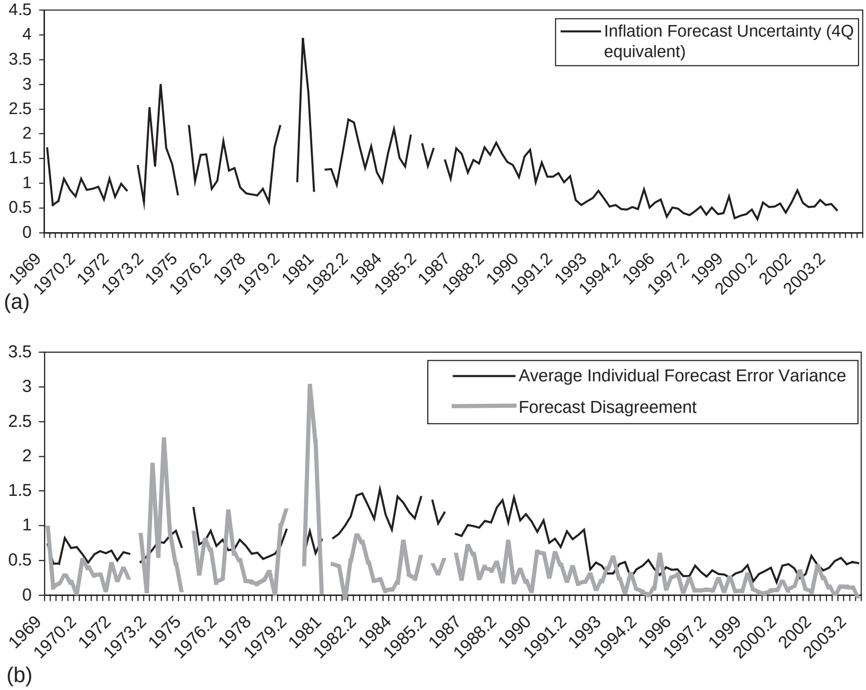


Fig. 2. (a) Inflation Forecast Uncertainty Based on Survey Data (Horizon Adjusted); (b) Two Components of Inflation Forecast Uncertainty.

from 1968Q4 to 2003Q4.[19] First, we see that the inflation forecast uncertainty as measured by $\hat{U}_{th} - \hat{\delta}_h$ was low before 1973 and after 1992. This is consistent with previous studies. A similar pattern persists for its two components also. This roughly confirms Friedman's (1977) conjecture that greater inflation uncertainty is associated with higher level of inflation. Second, the forecast disagreement on the average over the whole sample period is lower (0.38) than the average of individual forecast uncertainty (0.70) over the sample period. This is consistent with the empirical findings in previous studies such as Zarnowitz and Lambros (1987). Third, forecast disagreement increased significantly in mid-1970s and early 1980s. In mid-1970s, the U.S. economy was hit by the first oil shock. Such an event raises people's uncertainty about future change in the inflation regime. But it had interestingly only a slight effect on the average variance of inflation, $\hat{\sigma}^2_{\lambda_{th}} - \hat{\delta}_h$. The late 1970s and early 1980s witnessed similar episodes when the economy

Table 1. Estimates of Average Horizon Effects in Forecast Uncertainty.

Variable	Coefficient	Std. Error	*p*-Value
Intercept	0.6676	0.029	0.000
3Q ahead dummy	−0.0006	0.023	0.980
2Q ahead dummy	−0.0929	0.019	0.000
1Q ahead dummy	−0.1959	0.018	0.000

was hit by the second oil shock and a sudden change in monetary policy regime from the interest rate targeting to the money supply.[20] Fourth, the sample variance of forecast disagreement is 0.19 over the sample period while that for the average individual forecast uncertainty is 0.11. This is consistent with Lahiri et al. (1988)'s finding that forecast disagreement tend to be more volatile than the average individual forecast uncertainty. Fifth, the correlation coefficient of the overall measure of uncertainty ($\hat{U}_{th} - \hat{\delta}_h$) with average variance ($\hat{\sigma}^2_{\lambda_{th}} - \hat{\delta}_h$) and forecast disagreement ($\hat{\sigma}^2_{\mu_{th}}$) are 0.72 and 0.85 respectively. This result implies that forecast disagreement is a good proxy for uncertainty. Almost all studies in this area have reached this conclusion. Finally, note from Table 1 that there are significant horizon effects. Interestingly, we find that forecast uncertainty, on the average, falls significantly from two-to one-quarter-ahead forecasts.

6. GARCH MODEL OF UNCERTAINTY

In Section 2, we argued that the sum of the forecast disagreement and the average individual forecast error variance approximates the time-series measure of uncertainty fairly well. In this section, we test how well this argument is valid by comparing these two measures empirically. The most popular time-series models for estimating forecast uncertainty are ARCH and its various extensions. In these models, the forecast uncertainty is measured as the time-varying conditional variance of innovations.[21]

Following Engle and Kraft (1983), we model quarterly inflation π_t as an AR(4) process

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2\pi_{t-2} + \beta_3\pi_{t-3} + \beta_4\pi_{t-4} + \varepsilon_t \tag{6.1}$$

To test for the presence of conditional heteroscedasticity in the form of ARCH/GARCH process, we estimated (6.1) by OLS on quarterly inflation data from 1955Q2 to 2004Q2 released in 2004Q3. The squared residuals are then regressed on its own lags up to 20. The χ^2 test statistic was significant at the significance level of 5%, suggesting that ARCH/GARCH effect is presented.[22]

To capture the ARCH/GARCH (generalized autoregressive conditional heteroscedastic) effect, we tried three formulations for the conditional variance of ε_t: the popular GARCH(1,1) model in which the conditional variance of inflation is formulated as

GARCH(1,1):

$$h_t = E(\varepsilon_t^2|\psi_{t-1}) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2 h_{t-1} \tag{6.2}$$

where $\alpha_0 > 0, \alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_1 + \alpha_2 < 1$,[23] and $\psi_{t-1} = \{\pi_{t-1}, h_{t-1}, \varepsilon_{t-1} \ \ldots\}$ is the information set at date $t-1$.

ARCH(4):

$$h_t = E(\varepsilon_t^2|\psi_{t-1}) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \alpha_3\varepsilon_{t-3}^2 + \alpha_4\varepsilon_{t-4}^2 \tag{6.3}$$

where $\alpha_0 > 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0,\ \sum_{i=1}^{4}\alpha_i < 1$.

GJR–GARCH(1,1):

$$h_t = E(\varepsilon_t^2|\psi_{t-1}) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2 h_{t-1} + \alpha_3 D_{t-1}\varepsilon_{t-1}^2 \tag{6.4}$$

where $\alpha_0 > 0,\ \alpha_1 \geq 0,\ \alpha_2 \geq 0,\ \alpha_1 + \alpha_2 + 0.5\alpha_3 < 1$, and $D_t = \begin{cases} 1 \text{ if } \varepsilon_t < 0 \\ 0 \text{ otherwise} \end{cases}$

Table 2 shows the estimates of the three models over the same sample. Values of log likelihood, AIC and SC reveal that GJR–GARCH(1,1) model provides the best fit among the three models. This result is meaningful since it is well known that an unpredicted fall in inflation produces less uncertainty than an unpredicted rise in inflation (Giordani & Soderlind, 2003). One explanation may be that during a period of unexpected high inflation, people are uncertain about whether or not the monetary authorities will adopt a disinflationary policy at the potential cost of higher unemployment or lower growth rate of output. However, if there is unexpected low inflation, people believe that the monetary authorities will seek to maintain the low inflation, so inflation forecast uncertainty would be low. Ball (1992) used this argument to explain why higher inflation leads to higher inflation uncertainty.

To make the time-series results comparable to the survey measure, we estimated the model with real-time macro data available from the Federal Reserve Bank of Philadelphia.[24] This data set includes data as they existed in the middle of each quarter, from November 1965 to the present. For each

Table 2. Estimates of Univariate ARCH models.

Variable	GARCH(1,1)	ARCH(4)	GJR–GARCH(1,1)
Mean equation			
Constant	0.0510*	0.0567	0.0567**
	(0.0288)	(0.0374)	(0.0273)
π_{t-1}	0.4708***	0.4372***	0.5018***
	(0.0767)	(0.0746)	(0.0820)
π_{t-2}	0.1780**	0.2252*	0.1621*
	(0.0811)	(0.1171)	(0.0854)
π_{t-3}	0.0393	0.0325	0.0195
	(0.0826)	(0.1058)	(0.0817)
π_{t-4}	0.2301***	0.2147***	0.2454***
	(0.0712)	(0.0762)	(0.069)
Variance equation			
Constant	0.0018	0.0185***	0.0020
	(0.0013)	(0.0067)	(0.0013)
ε^2_{t-1}	0.1278***	0.2038	0.2127**
	(0.0482)	(0.1393)	(0.0869)
ε^2_{t-2}		0.1482	
		(0.1341)	
ε^2_{t-3}		0.1547*	
		(0.0852)	
ε^2_{t-4}		0.2677**	
		(0.1242)	
$D_{t-1}\varepsilon^2_{t-1}$			−0.2127***
			(0.0831)
h_{t-1}	0.8439***		0.8546***
	(0.0535)		(0.0746)
R^2	0.807	0.795	0.810
Log-L	5.005	2.341	10.614
AIC	0.000	0.027	−0.058
SC	0.084	0.111	0.026

Note: Standard errors are given in parentheses.
***Significant at 1% level.
**Significant at 5% level.
*Significant at 10% level.

vintage date, the observations are identical to those one would have observed at that time. To estimate our models, we make use of only the data for Output Price Index.[25] (It was first GNP IPD, then GDP IPD and finally Chain-weighted price index for real GDP since 1996, see Section 4 for detailed discussion). The quarterly inflation rate is defined as log difference of quarterly price index. Specifically, $\pi_t = 100 \times (\ln(p_t) - \ln(p_{t-1}))$, where p_t denotes the price level at date t.

The estimation and forecast procedure is as follows. First, the above models are estimated using the quarterly inflation data available at a particular point of time starting from 1955Q1 and ending at the last quarter before the forecasting date. This construction is intended to reproduce the information sets of forecasters in real time. Then the estimated model is used to forecast the inflation forecast uncertainty. Since the survey measure reports uncertainty associated with the forecast of annual inflation rate in the target year made in different quarters before the end of that year, we cannot just compare the conditional variance of quarterly inflation forecasts with the survey measure directly. Previous studies have not been sufficiently clear about this important issue. Based on the models for quarterly inflation, however, we could derive a measure of forecast uncertainty comparable to the survey measure. For that purpose, we should first make some changes to the previous notations.

Let π_t denote annual inflation rate for year t, $\pi_{t,i}$ denote quarterly inflation rate in the ith quarter of year t, and $p_{t,i}$ denote the price index for the ith quarter of year t. Consider the forecast for annual inflation rate in year t made in the first quarter of that year.[26] By definition, the annual inflation rate in year t can be expressed as the sum of quarterly inflation rate in each quarter of that year.

$$\begin{aligned}\pi_t &= 100 \times (\ln(p_{t,4}) - \ln(p_{t-1,4})) \\ &= 100 \times (\ln(p_{t,4}) - \ln(p_{t,3}) + \ln(p_{t,3}) \\ &\quad - \ln(p_{t,2}) + \ln(p_{t,2}) - \ln(p_{t,1}) \\ &\quad + \ln(p_{t,1}) - \ln(p_{t-1,4})) \\ &= \pi_{t,4} + \pi_{t,3} + \pi_{t,2} + \pi_{t,1} \end{aligned} \tag{6.5}$$

The forecast for π_t in the first quarter of that year is as follows:

$$\begin{aligned}\pi_t^f &= 100 \times (\ln(p_{t,4}^f) - \ln(p_{t-1,4})) \\ &= 100 \times (\ln(p_{t,4}^f) - \ln(p_{t,3}^f) + \ln(p_{t,3}^f) \\ &\quad - \ln(p_{t,2}^f) + \ln(p_{t,2}^f) - \ln(p_{t,1}^f) \\ &\quad + \ln(p_{t,1}^f) - \ln(p_{t-1,4})) \\ &= \pi_{t,4}^f + \pi_{t,3}^f + \pi_{t,2}^f + \pi_{t,1}^f \end{aligned} \tag{6.6}$$

where X^f denotes the forecast for the variable X made in the first quarter[27] of the year considered. The error with annual inflation forecast made in the first quarter of year t is the difference between π_t and π_t^f.

$$\begin{aligned} e_t &= \pi_t - \pi_t^f \\ &= \left(\pi_{t,4} + \pi_{t,3} + \pi_{t,2} + \pi_{t,1}\right) - \left(\pi_{t,4}^f + \pi_{t,3}^f + \pi_{t,2}^f + \pi_{t,1}^f\right) \\ &= e_{t,4} + e_{t,3} + e_{t,2} + e_{t,1} \end{aligned} \tag{6.7}$$

where $e_{t,i}$ is the error in forecasting the ith quarterly inflation of year t made in the first quarter of the same year. As before, the quarterly inflation is modeled as AR(4). So, the forecast for quarterly inflation rate of the fourth quarter of year t is based on the regression

$$\pi_{t,4} = \beta_0 + \beta_1 \pi_{t,3} + \beta_2 \pi_{t,2} + \beta_3 \pi_{t,1} + \beta_4 \pi_{t-1,4} + \varepsilon_{t,4} \tag{6.8}$$

where $\varepsilon_{t,4}$ is the innovation in the fourth quarter of year t which is assumed to be normally distributed with zero mean and a time-varying conditional variance. Note that

$$\pi_{t,4}^f = \beta_0 + \beta_1 \pi_{t,3}^f + \beta_2 \pi_{t,2}^f + \beta_3 \pi_{t,1}^f + \beta_4 \pi_{t-1,4} \tag{6.9}$$

where all forecasts are made in the first quarter of year t. Note also that $\pi_{t-1,4}$ is known to forecasters at that time. Based on above assumptions, we have

$$e_{t,4} = \pi_{t,4} - \pi_{t,4}^f = \beta_1 e_{t,3} + \beta_2 e_{t,2} + \beta_3 e_{t,1} + \varepsilon_{t,4} \tag{6.10}$$

Similarly, we have

$$e_{t,3} = \pi_{t,3} - \pi_{t,3}^f = \beta_1 e_{t,2} + \beta_2 e_{t,1} + \varepsilon_{t,3} \tag{6.11}$$

$$e_{t,2} = \pi_{t,2} - \pi_{t,2}^f = \beta_1 e_{t,1} + \varepsilon_{t,2} \tag{6.12}$$

$$e_{t,1} = \pi_{t,1} - \pi_{t,1}^f = \varepsilon_{t,1} \tag{6.13}$$

By successive substitution, we have

$$e_{t,i} = \sum_{j=1}^{i} b_{j-1} \varepsilon_{t,i-j+1} \tag{6.14}$$

where $b_j = \beta_1 b_{j-1} + \ldots + \beta_4 b_{j-4}$, $b_0 = 1$, and $b_j = 0$ for $j < 0$.

So, the forecast uncertainty conditional on information set in the first quarter of year t is

$$\begin{aligned} W_{t3} &= \operatorname{var}\left(e_t | \psi_{t,1}\right) = \operatorname{var}\left(e_{t,4} + e_{t,3} + e_{t,2} + e_{t,1} | \psi_{t,1}\right) \\ &= \sum_{k=1}^{4} \left(\sum_{j=0}^{4-k} b_j\right)^2 E\left(\varepsilon_{t,k}^2 | \psi_{t,1}\right) \end{aligned}$$

More generally, we have

$$W_{th} = \begin{cases} \sum_{k=1}^{h+1}\left(\sum_{j=0}^{h+1-k} b_j\right)^2 E\left(\varepsilon_{t,3-h+k}^2 | \psi_{t,4-h}\right) & \text{if } h = 1,\ 2,\ 3. \\ \sum_{k=1}^{4}\left(\sum_{j=0}^{4-k} b_j\right)^2 E\left(\varepsilon_{t,k}^2 | \psi_{t-1,4}\right) & \\ +\left(\sum_{j=1}^{4} b_j\right)^2 E\left(\varepsilon_{t-1,4}^2 | \psi_{t-1,4}\right) & \text{if } h = 4. \end{cases} \quad (6.15)$$

where $\psi_{t,i}$ denotes the information set in the i^{th} quarter of year t.[28] W_{th} denotes the forecast uncertainty for annual inflation rate in year t with horizon of h quarters. It is comparable to $\hat{U}_{th}$ in (5.6), the survey forecast uncertainty for annual inflation in year t with horizon of h quarters.

Figs. 3a–d compare the survey forecast uncertainty $\hat{U}_{th}$[29] with time-series uncertainty W_{th} for different forecast horizons.[30] The time profiles and the levels are quite similar after late 1980s. The correlations between these two series are very high after 1984 (for all forecast horizons, the correlation coefficients are above 0.8). During this period, the inflation rate is relatively low and stable. However, the correlation coefficients are quite low before 1984, (see Table 4). This period is notable for high and volatile inflation. It seems that the two measures give similar results when the inflation process is stable. On the other hand, they diverge when there are structural breaks in the inflation process. This is understandable since ARCH-type models assume that the regime for inflation and inflation uncertainty is invariant over time while survey forecasters surely try to anticipate structural breaks when they make forecasts.[31] Although these two measures have very low correlation in the earlier period, both of them were sensitive to the effects of the first oil shock in 1973–1974, and the second oil shock and the change in the monetary rule in 1979–1980 and 1981–1982, respectively. During these periods, both measures reported big surges. For the survey measure, by looking at Fig. 2b, we find that the increase is mostly due to the increase in the forecast disagreement. This again justifies the inclusion of the forecast disagreement into inflation uncertainty. As Kurz (2002) has argued, when important policies are enacted, heterogeneity in beliefs creates disagreement in forecasts that in turn contributes to the overall economic uncertainty.

To better understand the divergence between the survey measure and time-series measures, we examined the point forecasts produced by the SPF data[32] and GARCH(1,1). We found that before 1984, the point forecasts estimated by the two approaches were much more different than those after 1984. On average, the sum of absolute value of the difference between these two

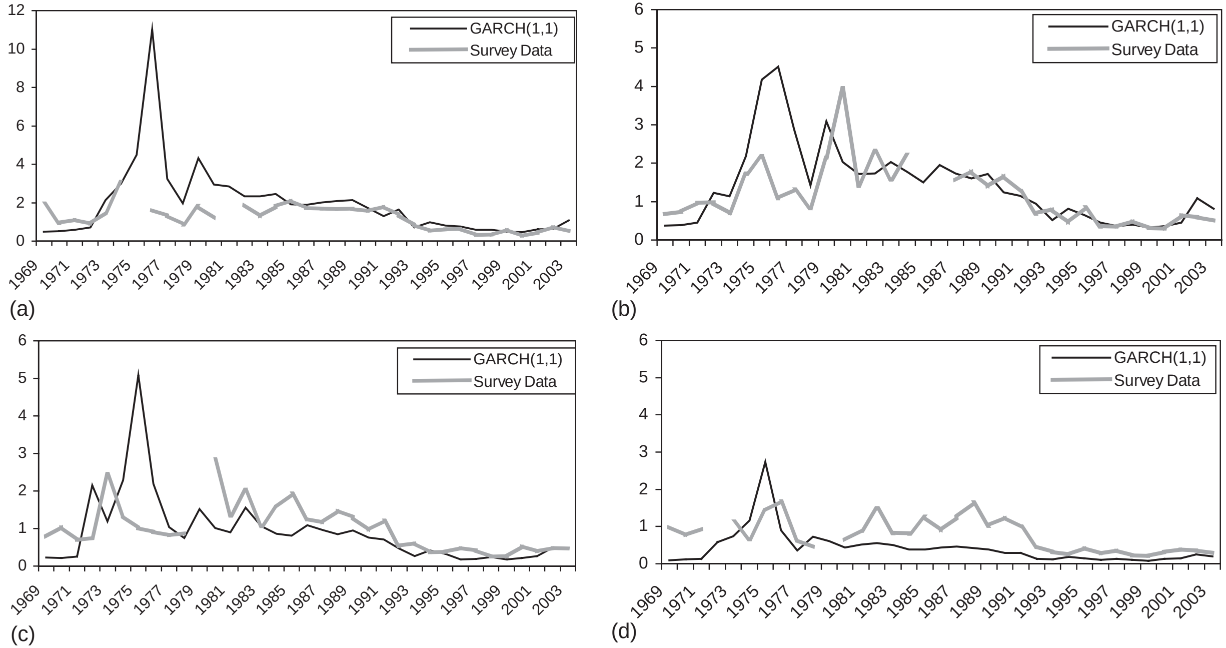


Fig. 3. (a) Four-Quarter-Ahead Forecast Uncertainty; (b) Three-Quarter-Ahead Forecast Uncertainty; (c) Two-Quarter-Ahead Forecast Uncertainty; (d) One-Quarter-Ahead Forecast Uncertainty.

estimates is 100% larger before 1984 compared to the post-1984 period. During mid-1970s to early 1980s, the GARCH point forecasts missed the actual inflation rates substantially. The average absolute forecast errors for the AR (4) model was calculated to be 1.21 during 1968-1983 (1.41 during 1973–1983) but only 0.39 during 1984–2003.[33] The same for the survey measure is 1.0 during 1968–1983 (1.11 during 1973–1983) and 0.51 during 1984–2003. Thus, it seems that when there are structural breaks, the survey measure does better than the time-series models in forecasting inflation while during a period of low and stable inflation, the latter does better. Since uncertainties based on ARCH-type models are functions of forecast errors, it is not surprising to find that ARCH-type uncertainties are bigger than the survey measure during mid-1970s and early 1980s in Figs. 3a–d. This may also explain why the survey measure of uncertainty is bigger than ARCH-type uncertainties from mid-1980s to early 1990s for two- and one-quarter-ahead forecasts.

Another undesirable feature of Figs. 3a–d is that the GARCH(1,1) measure has big spikes in 1975 and 1976. This problem is especially serious for four-quarter-ahead forecasts. This may be due to the interaction between forecast errors and model parameters. Notice that due to the effect of the first oil shock, GARCH(1,1) reports big forecast errors in 1974 and 1975. Actually the two largest forecast errors for the whole sample period occurred in these two years.[34] Consequently, as a result of the GARCH specification big forecast errors will show up as big forecast uncertainty in the following years. That is why we have big forecast uncertainty in 1975 and 1976. One simple way to correct this problem is just to dummy out the several quarters that have large forecast errors. Actually, just dummying out the first and second quarters of 1975 will reduce the spike in 1976 significantly. But this is only an ex post solution. Another explanation for the spikes in 1975 and 1976 is that GARCH(1,1) exaggerates people's responses to past forecast errors. Considering this, we modified the standard GARCH(1,1) model by formulating the conditional variance as a function of lagged conditional variance and absolute forecast error (instead of squared forecast error).[35] This method worked very well to reduce the spikes in 1975 and 1976, especially for long horizon forecast uncertainty. Figs. 4a–d report forecast uncertainty estimated with this modified GARCH(1,1) model and it is obvious that the big spike in 1976 is reduced by more than 50% for four-quarter-ahead forecasts. We also examined the correlation between the survey measure and the measure based on this modified GARCH(1,1). Although it is still much smaller during 1968–1983 than during 1984–2003, there is significant improvement over the standard GARCH(1,1), especially for the earlier period. Actually, as showed in

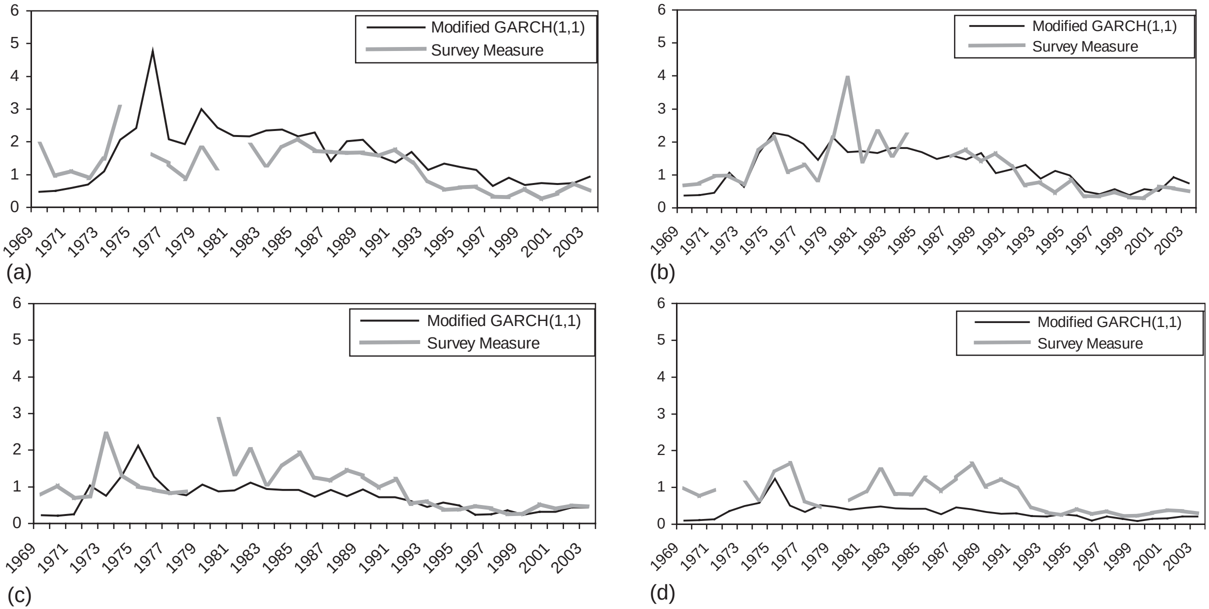


Fig. 4. (a) Four-Quarter-Ahead Forecast Uncertainty; (b) Three-Quarter-Ahead Forecast Uncertainty; (c) Two-Quarter-Ahead Forecast Uncertainty; (d) One-Quarter-Ahead Forecast Uncertainty.

Table 4, this model performs better than other models in the sense of reproducing the survey measure.

Above analysis implies that a time-series model that takes account of structural breaks may match the survey measure better if the structural breaks are correctly specified. One possible candidate is a model proposed by Evans (1991),[36] in which the parameters in the mean equation of the ARCH models are allowed to vary over time. Evans proposed three measures of inflation forecast uncertainty, one of which is the sum of conditional variance of innovation to inflation process and the parameter uncertainty. To see if this formulation helps to match the survey and times series measures of inflation forecast uncertainty, we estimated the model and obtained forecasts using real time data recursively. Specifically, the model can be written as:

$$\pi_{t+1} = x_t\beta_{t+1} + \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, h_{t+1}) \tag{6.16}$$

and

$$x_t = [1, \pi_{t-1}, \pi_{t-2}, \pi_{t-3}, \pi_{t-4}],\ \beta_{t+1} = \beta_t + V_{t+1},\ V_{t+1} \sim N(0, Q)$$

where V_{t+1} is a vector of normally distributed shocks to the parameter vector β_{t+1} with a homoskedastic diagonal covariance matrix Q. For the theoretical and empirical arguments for the random walk formulation, see Evans (1991) and Engle and Watson (1985). This model can be written in state space form and estimated with the Kalman Filter.

Observation equation:

$$\pi_{t+1} = \begin{pmatrix} x_t & 1 \end{pmatrix} \begin{pmatrix} \beta_{t+1} \\ \varepsilon_{t+1} \end{pmatrix} + \omega_{t+1}$$

where $\omega_{t+1} = 0$

State equation:

$$\begin{pmatrix} \beta_{t+1} \\ \varepsilon_{t+1} \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \beta_t \\ \varepsilon_t \end{pmatrix} + \begin{pmatrix} V_{t+1} \\ \varepsilon_{t+1} \end{pmatrix}$$

The above model is actually a special case of the unobserved component time-series model with ARCH disturbances discussed in Harvey, Ruiz, and Sentana (1992). As pointed out by these authors, with time-varying parameters, past forecast errors, and conditional variances are no longer in the information sets of forecasters and must be inferred from the estimation of the state. They suggested a way to deal with this problem. Following their

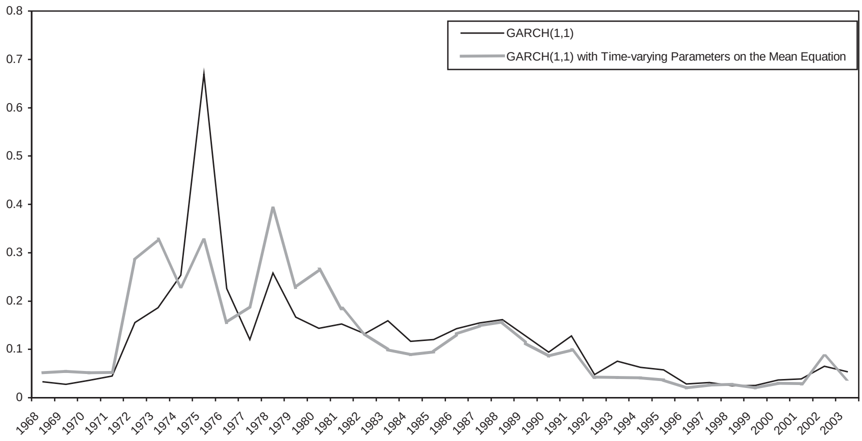


Fig. 5. Current Quarter Forecast Uncertainty.

method, we estimated the above model using real-time quarterly inflation data recursively.[37] The result is shown in Fig. 5 for current quarter forecasts. We find that the estimated inflation forecast uncertainty from a time-varying parameter GARCH model is smaller than that estimated from the ordinary GARCH model during the mid-1970s. This result suggests that the divergence between time-series measures and survey measures may be resolved to a great extent by generalizing the time-series models that take into account explicitly the structural breaks and parameter drift.

7. INFLATION UNCERTAINTY USING VAR-ARCH MODELS

Univariate ARCH model and its extensions formulate inflation as a function only of its own past values. This assumption ignores the interactions between inflation and other variables that can be used to forecast inflation. However, as Stock and Watson (1999) have demonstrated, some macroeconomic variables can help improve inflation forecast. Among these variables, they find that inflation forecasts produced by the Phillips curve generally have been more accurate than forecasts based on other macroeconomic variables, such as interest rates, money supply, and commodity prices. In this section, we investigate if other macroeconomic variables help to match the survey measure and time-series measures of inflation forecast uncertainty. Considering the availability of real-time data, we will use the traditional Phillips curve based on unemployment.[38] The forecasting model of inflation then becomes

$$\pi_t = \beta_0 + \beta_1\pi_{t-1} + \beta_2\pi_{t-2} + \beta_3\pi_{t-3} + \beta_4\pi_{t-4} + \delta_1 u_{t-1} + \delta_2 u_{t-2} + \delta_3 u_{t-3} + \delta_4 u_{t-4} + \varepsilon_t \tag{7.1}$$

where, π_t, u_t, and their lags are quarterly inflation and unemployment rates. To test for the presence of conditional heteroscedasticity in the form of ARCH/GARCH process, we estimated (7.1) by OLS over the sample period from 1955Q2 to 2004Q2 released in 2004Q3. The squared residuals were then regressed on it's own lags up to 20. The χ^2 test found a significant ARCH/GARCH effect at the 1% level.[39] One problem with forecasting inflation by (7.1) is that we need to first forecast unemployment rate when the forecast horizon is more than one period. This problem can be solved by modeling unemployment also as an AR(4) process:

$$u_t = \gamma_0 + \gamma_1 u_{t-1} + \gamma_2 u_{t-2} + \gamma_3 u_{t-3} + \gamma_4 u_{t-4} + \upsilon_t \tag{7.2}$$

Putting (7.1) and (7.2) together, we model inflation and unemployment as VAR-(G)ARCH processes.

$$\begin{pmatrix} \pi_t \\ u_t \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \gamma_0 \end{pmatrix} + \begin{pmatrix} \beta_1 & \delta_1 \\ 0 & \gamma_1 \end{pmatrix}\begin{pmatrix} \pi_{t-1} \\ u_{t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_4 & \delta_4 \\ 0 & \gamma_4 \end{pmatrix}\begin{pmatrix} \pi_{t-4} \\ u_{t-4} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \upsilon_t \end{pmatrix} \tag{7.3}$$

or compactly

$$Z_t = A_0 + A_1 Z_{t-1} + \ldots + A_4 Z_{t-4} + \mu_t$$

where

$$Z_t = \begin{pmatrix} \pi_t \\ u_t \end{pmatrix}, \quad \mu_t = \begin{pmatrix} \varepsilon_t \\ \upsilon_t \end{pmatrix}$$

Innovations in the inflation equation are assumed to have a time varying conditional variance. Specifically, we assume

$$E(\mu_t|\psi_{t-1}) = 0 \text{ and } \mathrm{var}(\mu_t|\psi_{t-1}) = H_t = \begin{pmatrix} \sigma_t^2 & \rho_{\varepsilon\upsilon}\sigma_\upsilon\sigma_t \\ \rho_{\varepsilon\upsilon}\sigma_\upsilon\sigma_t & \sigma_\upsilon^2 \end{pmatrix}$$

where

$$\sigma_t^2 = E(\varepsilon_t^2|\psi_{t-1}) = h(\varepsilon_{t-1}, \ldots, \sigma_{t-1}^2, \ldots),\ \sigma_\upsilon^2 = E(\upsilon_t^2|\psi_{t-1}) = E(\upsilon_t^2)$$

and

$$E(\varepsilon_t \upsilon_t | \psi_{t-1}) = \sigma_{\varepsilon\upsilon t} = \rho_{\varepsilon\upsilon}\sigma_\upsilon\sigma_t$$

Following Bollerslev (1990), we assume constant conditional correlation coefficient to simplify Maximum Likelihood estimation. Assuming conditional normality, the model can be estimated as described in that paper. Table 3 shows the estimates of VAR–ARCH(1) and VAR–GARCH(1,1) over the same sample as before.

As in the previous section, if the target variable is annual inflation rate forecasted at different horizons, the forecast error is the sum of corresponding forecast errors for the quarterly variables. Consider the forecast made in the first quarter of year t. The forecast errors for the vector of quarterly inflation and quarterly unemployment rate in the four quarters of year t are $e_{t,1} = \mu_{t,1}$, $e_{t,2} = A_1 e_{t,1} + \mu_{t,2}$, $e_{t,3} = A_1 e_{t,2} + A_2 e_{t,1} + \mu_{t,3}$, and $e_{t,4} = A_1 e_{t,3} + A_2 e_{t,2} + A_3 e_{t,1} + \mu_{t,4}$, where $X_{t,i}$ denotes the value of X in the ith quarter of year t.

By successive substitution, we have

$$e_{t,i} = \sum_{j=1}^{i} B_{j-1}\mu_{t,i-j+1} \tag{7.4}$$

where $B_j = A_1 B_{j-1} + \ldots + A_4 B_{j-4}, B_0 = I, B_j = 0$ for $j<0$.

After some tedious algebra, it can be shown that the forecast error for quarterly inflation in the ith quarter of year t is

$$e^{\pi}_{t,i} = \sum_{j=1}^{i} b_{j-1}\varepsilon_{t,i-j+1} + \sum_{j=1}^{i} \lambda_{j-1}\upsilon_{t,i-j+1} \tag{7.5}$$

where b_j is defined as in (6.14) and $\lambda_j = \Sigma_{k=0}^{3} d_k b_{j-k}$, $d_k = \delta_1 c_{k-1} + \ldots + \delta_4 c_{k-4}$ with $d_0 = 0$, and $c_j = \gamma_1 c_{j-1} + \ldots + \gamma_4 c_{j-4}$ with $c_0 = 1$ and $c_j = 0$ for $j<0$.

So, the forecast uncertainty of annual inflation conditional on the information set in the first quarter of year t is[40]

$$\begin{aligned}
W_{t3} &= \mathrm{var}(e^{\pi}_{t} | \psi_{t,1}) \\
&= \mathrm{var}(e^{\pi}_{t,4} + e^{\pi}_{t,3} + e^{\pi}_{t,2} + e^{\pi}_{t,1} | \psi_{t,1}) \\
&= \sum_{k=1}^{4}\left(\sum_{j=0}^{4-k} b_j\right)^2 E(\varepsilon^2_{t,k}|\psi_{t,1}) + \sum_{k=1}^{4}\left(\sum_{j=0}^{4-k} \lambda_j\right)^2 E(\upsilon^2_{t,k}|\psi_{t,1}) \\
&\quad + 2\rho_{\varepsilon\upsilon}\sum_{k=1}^{4}\left(\sum_{j=0}^{4-k}\lambda_j\right)\left(\sum_{j=0}^{4-k} b_j\right)\sqrt{E(\varepsilon^2_{t,k}|\psi_{t,1})E(\upsilon^2_{t,k}|\psi_{t,1})}
\end{aligned}$$

Table 3. Estimates of Multivariate Models.

Variable	VAR–ARCH(1)	VAR–GARCH(1,1)
Inflation equation π_t		
Constant	0.2422***	0.2288***
	(0.0705)	(0.066)
π_{t-1}	0.4430***	0.3888***
	(0.1027)	(0.0769)
π_{t-2}	0.1580**	0.2044***
	(0.0758)	(0.0783)
π_{t-3}	0.0991	0.0774
	(0.0785)	(0.0781)
π_{t-4}	0.2933***	0.3067***
	(0.0805)	(0.0698)
u_{t-1}	−0.2967***	−0.2568***
	(0.0591)	(0.0599)
u_{t-2}	0.2922**	0.2245**
	(0.1325)	(0.1105)
u_{t-3}	−0.1066*	0.0222
	(0.1535)	(0.1093)
u_{t-4}	0.0691	−0.0291
	(0.0719)	(0.0569)
Unemployment equation u_t		
Constant	0.2588***	0.2585***
	(0.0863)	(0.0863)
u_{t-1}	1.6232***	1.6219***
	(0.0713)	(0.0703)
u_{t-2}	−0.7117***	−0.71***
	(0.1352)	(0.1264)
u_{t-3}	−0.0285	−0.0299
	(0.1359)	(0.1163)
u_{t-4}	0.0737	0.0747
	(0.0719)	(0.0633)
Variance equation		
Constant	0.0368***	0.0021
	(0.0056)	(0.0017)
ε_{t-1}^2	0.4271***	0.1627**
	(0.1376)	(0.0723)
σ_{t-1}^2		0.8008***
		(0.0852)
$\rho_{\varepsilon\upsilon}$	−0.0361	−0.026
	(0.0757)	(0.0751)
σ_υ	0.273***	0.273***
	(0.0139)	(0.0139)

Note: Standard errors are given in parentheses.
*Significant at 10% level.
**Significant at 5% level.
***Significant at 1% level.

More generally, we have

$$
\begin{aligned}
W_{th} = {} & \sum_{k=1}^{h+1}\left(\sum_{j=0}^{h+1-k} b_j\right)^2 E(\varepsilon^2_{t,3-h+k}|\psi_{t,4-h}) \\
& + \sum_{k=1}^{h+1}\left(\sum_{j=0}^{h+1-k} \lambda_j\right)^2 E(\upsilon^2_{t,3-h+k}|\psi_{t,4-h}) \\
& + 2\rho_{\varepsilon\upsilon}\sum_{k=1}^{h+1}\left(\sum_{j=0}^{h+1-k} \lambda_j\right)\left(\sum_{j=0}^{h+1-k} b_j\right) \\
& \times\sqrt{E(\varepsilon^2_{t,3-h+k}|\psi_{t,4-h})E(\upsilon^2_{t,3-h+k}|\psi_{t,4-h})} \quad \text{if } h = 1, 2, 3 \\
W_{th} = {} & \sum_{k=1}^{4}\left(\sum_{j=0}^{4-k} b_j\right)^2 E(\varepsilon^2_{t,k}|\psi_{t-1,4}) + \sum_{k=1}^{4}\left(\sum_{j=0}^{4-k} \lambda_j\right)^2 E(\upsilon^2_{t,k}|\psi_{t-1,4}) \\
& + 2\rho_{\varepsilon\upsilon}\sum_{k=1}^{4}\left(\sum_{j=0}^{4-k} b_j\right)\left(\sum_{j=0}^{4-k} \lambda_j\right) \\
& \times\sqrt{E(\varepsilon^2_{t,k}|\psi_{t-1,4})E(\upsilon^2_{t,k}|\psi_{t-1,4})} \\
& + \left(\sum_{j=1}^{4} b_j\right)^2 E(\varepsilon^2_{t-1,4}|\psi_{t-1,4}) + \left(\sum_{j=1}^{4} \lambda_j\right)^2 E(\upsilon^2_{t-1,4}|\psi_{t-1,4}) \\
& + 2\rho_{\varepsilon\upsilon}\left(\sum_{j=1}^{4} b_j\right)\left(\sum_{j-1}^{4} \lambda_j\right) \\
& \times\sqrt{E(\varepsilon^2_{t-1,4}|\psi_{t-1,4})E(\upsilon^2_{t-1,4}|\psi_{t-1,4})} \quad \text{if } h = 4
\end{aligned}
\tag{7.6}
$$

where $\psi_{t,i}$ denotes the information set in the ith quarter of year *t*. W_{th} denotes the forecast uncertainty for annual inflation rate in year *t* with horizon of *h* quarters.

Table 4 shows the simple correlation of the survey measure with different time-series measures. Some interesting conclusions may be drawn from this table. First, on average GARCH models simulate the survey measure better than ARCH models. The average correlations over different forecast horizons are 0.354 and 0.322 for VAR–ARCH(1) and ARCH(4), respectively, much lower than the correlations for corresponding VAR–GARCH(1,1) and GARCH(1,1) models, which were 0.465 and 0.447, respectively. It confirms the finding in the literature that changes in the

Table 4. Correlations of Time Series and Survey Measures of Uncertainty.

		VAR–GARCH(1,1)	VAR–ARCH(1)	GJR–GARCH(1,1)	GARCH(1,1)	Modified GARCH(1,1)	ARCH(4)
4Q-Ahead Forecast	Whole sample	0.642*	0.297	0.384	0.427	0.530	0.317
	1968–1983	0.474*	0.137	0.118	0.224	0.244	0.166
	1984–2001	0.875	0.895	0.918*	0.910	0.873	0.886
3Q-Ahead Forecast	Whole sample	0.528	0.566	0.690	0.584	0.701*	0.429
	1968–1983	0.385	0.569*	0.569*	0.369	0.542	0.187
	1984–2001	0.905*	0.802	0.794	0.886	0.843	0.749
2Q-Ahead Forecast	Whole sample	0.294	0.255	0.338	0.294	0.459*	0.176
	1968–1983	0.039	0.139*	0.065	−0.037	0.087	−0.081
	1984-2001	0.858	0.744	0.874	0.861	0.877*	0.747
1Q-Ahead Forecast	Whole sample	0.396	0.296	0.464	0.481	0.565*	0.366
	1968-1983	0.087	−0.016	0.265	0.401*	0.376	0.318
	1984-2001	0.887	0.801	0.922*	0.870	0.840	0.750
Average across horzions	Whole sample	0.465	0.354	0.469	0.447	0.564*	0.322
	1968-1983	0.246	0.207	0.254	0.240	0.312*	0.148
	1984-2001	0.881	0.811	0.877	0.882*	0.858	0.783

*Values are the highest correlation coefficients among all the six time-series models considered.

conditional variance of inflation persist over time. Actually, one motivation of developing GARCH model is to provide a better formulation to capture this feature of inflation data (Bollerslev, 1986). Second, univariate models perform as well as bivariate models in simulating the survey measure of uncertainty. This may be quite surprising at first glance since it is well documented that the Phillips curve provides a better point forecast of inflation than autoregressive models of inflation. Ericsson (2003) discusses various determinants of forecast uncertainty. He points out that forecast uncertainty depends upon the variable being forecast, the type of model used for forecasting, the economic process actually determining the variable being forecast, the information available, and the forecast horizon. Ericsson also differentiates between actual forecast uncertainty and anticipated forecast uncertainty, which is model dependent. In our case, the difference between forecast uncertainty from univariate models and that from multivariate models lies only in the difference in model specification since they have the same target variable, same data generating process, information set, and forecast horizon. Thus, it is not surprising that different models will produce different anticipated forecast uncertainty. Third, models that allow for asymmetric effects of positive and negative innovations came closer to the survey measure compared to models that do not. Actually, among the six models we report, GJR–GARCH(1,1) has the second highest average correlation with the survey measure of uncertainty. Fourth, our modified GARCH(1,1) model is the best in terms of correlation with the survey measure. As discussed in Section 6, this is because it modifies the GARCH specification by replacing squared-past errors with absolute errors. Finally, for all models, the correlation of uncertainty of the two approaches is quite high for the period 1984–2003, a period with low and stable inflation, but low for the period 1969–1983, a period notable for high and unstable inflation. Actually, the relative performance of different models in simulating the survey measure depends on how well it can simulate the survey measure during 1968–1983. For example, the correlation with the survey measure is relatively high for GJR–GARCH(1,1) and modified GARCH(1,1) in this period. As discussed in Section 5, our empirical estimate of the sum of the forecast disagreement and the average individual forecast uncertainty indirectly captures model uncertainty by incorporating the forecast disagreement. But the time series models discussed in this paper are based on an invariant model to produce forecast uncertainty for all periods. This may explain the big divergence of these two types of measures during periods with high model uncertainty.

8. CONCLUSIONS

In this paper, we develop a theoretical model to study the relationship between different measures of inflation forecast uncertainty based on survey data and aggregate time-series models. We found that the sum of the average variance of the individual densities and the disagreement slightly underestimates forecast uncertainty based on aggregate time-series data, and this underestimation is a function of the ratio of the variance of aggregate shocks to that of the idiosyncratic shocks. Given the existing empirical estimates of the ratio, we expect the underestimation to be minimal.

Even though the sum of the average variance and the disagreement is simply the variance of the average density, we cannot directly use the variance of the aggregate distribution to measure the aggregate forecast uncertainty due to grouping data problems, and compositional heterogeneity in the panel. In this paper we lay out a tedious procedure to extract the conceptually correct measure of aggregate uncertainty from the panel of density forecasts.

The SPF density forecast data have an accordion structure where each respondent forecasts year-over-year inflation rate every quarter in the current and the last year. In order to compare the forecast uncertainty from the surveys with those from the quarterly time-series data at various horizons, we developed the appropriate multi-period forecast variance formulas for multi-variate ARCH-type models. During the relatively stable years in our sample (viz., 1984–2002), the survey uncertainty and the ARCH-based time-series uncertainty were remarkably similar in their average values, correlations and temporal movements. Another finding of our analysis is that the univariate asymmetric GARCH model did as well as the bivariate GARCH model with Phillips curve.

During periods of rapid structural change (like the oil crises or Fed's new operating policies of early 1980s), we found that the two approaches could yield quite different estimates of uncertainty. Since these periods are characterized by larger forecast errors in time-series models than in survey data, the time-series uncertainty shoots up much more than the survey measure. During periods of rapid change, it is the disagreement component of uncertainty and not the average variance of the densities that responds, but not by as much as the time-series uncertainty. We found a simple way to make the time-series estimates robust to exceptionally large forecast errors, the trick is to model the conditional variance not as a function of squared forecast errors, but as a function of its absolute value.

Another fruitful approach is to model the time-series ARCH model as a time-varying parameter model. We found that this approach also helps in robustifying ARCH uncertainty estimates to abrupt structural breaks. Even then, the survey uncertainty tends to respond more prospectively than time-series models. Thus, during periods of structural breaks and regime changes which often elude forecasters, survey measure of uncertainty, when available, can be a very dependable 'reality-check' against which standard time-series measures can be evaluated. This reminds us of the sentiment Manski (2004) expressed when he wrote, "Economists have long been hostile to subjective data. Caution is prudent but hostility is not warranted." Our analysis shows that by carefully utilizing the SPF density forecasts data, we can garner substantial knowledge about the temporal pattern of aggregate uncertainty of forecasts, particularly when time series models could be failing. There remains much unexploited potential in the use of this subjective database.

NOTES

1. Interval and probability forecasts have a longer track record in weather forecasts and sports picks.
2. See, for instance, Sims (2002), Šmidková (2003), Cogley, Morozov, and Sargent (2003), and Garratt, Lee, Pesaran, and Shin (2003).
3. Sims (2002) discusses the current state and limitations of models generating macro forecasts and their uncertainty. Lahiri and Liu (2005) have examined the validity of the ARCH-type specifications using data on density forecasts.
4. Since 1990 the Federal Reserve Bank of Philadelphia manages the survey, now called the SPF.
5. In recent research, forecast heterogeneity has appeared prominently as a component of overall macroeconomic uncertainty. See, for instance, Kurz (2002), Carroll (2003), Mankiw, Reis, and Wolfers (2003), Souleles (2004), and Giordani and Söderlind (2005).
6. For example, it is commonly believed that the USDA forecasts for agricultural prices are used as 'benchmark' by many private and public forecasters, see Irwin, Gerlow, and Liu (1994) and Kastens, Schroeder, and Plain (1998).
7. Our formulation is also consistent with the evidence presented by Fuhrer (1988) that survey data contain useful information not present in the standard macroeconomic database.
8. In Lucas' (1972, 1973), signal extraction model, this private information is assumed to be the price in market z. Kurz and Motolese (2001) use a similar decomposition.
9. The decomposition of the variance of the aggregate distribution as the sum of average individual forecast error variance and forecast disagreement is discussed in

Lahiri et al. (1988) and Wallis (2004). Giordani and Söderlind (2003) provided an alternative justification for the decomposition, which is based on integrating out private information to get the aggregate distribution of inflation. As pointed out by Wallis (2004), this is an inappropriate approach because "aggregating forecasts is not the same as aggregating information sets".

10. Sample variance will converge in probability to the average of population variance. See Greene (2000, p. 504).

11. It is interesting to note that the so-called fan charts depicting multi-period forecast uncertainty as generated by Bank of England and Riksbank are very similar in spirit to the variance of the aggregate density from SPF. Starting from a base line model predictions that all share, bank economists and forecasting experts subjectively generate these charts incorporating their individual beliefs and specialized knowledge. In a recent paper, Cogley et al. (2003) show that the predictive intervals estimated from a Bayesian VAR that incorporate diverse statistical sources of uncertainty including model uncertainty, policy drift, structural shifts, and other shocks are more diffuse than the Bank of England's fan charts. This result is consistent with our Eq. (2.6). See Garratt et al. (2003) for a similar forecasting model generating uncertainty.

12. In addition to informational difference, use of different forecasting models, different beliefs, and subjective factors may be other reasons for the diversity of forecasts, (see Kurz & Motolese, 2001). Bomberger (1996) has emphasized the role of different models in generating disagreement.

13. For example, in the first quarter of 1985, many forecasters put most of the probability mass in the open lower interval.

14. We are grateful to Paolo Giordani and Paul Söderlind for kindly providing their programs.

15. The surveys for which this is true are 1968Q4, 1969Q4, 1970Q4, 1971Q4, 1972Q3 and Q4, 1973Q4, 1975Q4, 1976Q4, 1977Q4, 1978Q4, and 1979Q2–Q4.

16. Forecasts are made around the middle of each quarter.

17. (5.7) and (5.8) are estimated for forecasters who participated in the survey for 28 or more times to ensure that each forecaster reports at least one observation for each forecast horizon.

18. We should point out that these adjustments for horizons were done solely for presenting the uncertainty as a quarterly series in Figs. 2a and b. All other analysis was conducted with horizon effects left in.

19. It is necessary to explain the meaning of numbers in Figs 2a and b. Taking uncertainty in 1972 as an example, it is about 1, which means that the standard deviation is also 1. So, a 90% confidence band constructed from a normal distribution would have been $\pm 1.6\%$ around the point forecast in 1972. It would be (2.4%, 5.6%) if the point forecast is 4%.

20. Evans and Wachtel (1993) found that forecast disagreement is more closely associated with regime uncertainty than the uncertainty with a given inflation structure. Since everyone knew about the change in monetary policy, the abrupt shift in disagreement in the early 1980s had to be due to different beliefs and models rather than due to different information sets, see Fulford (2002) and Kurz (2002).

21. See Engle (1982, 1983), Bollerslev (1986), Nelson (1991), and Glosten et al. (1993) and surveys by Bollerslev, Chou, and Kroner (1992) and Bera and Higgins (1993).

22. The value of the test statistic TR^2 was equal to 36.81. The critical value for χ^2 test with 20 degrees of freedom is 31.4 at significance level of 5%.

23. These conditions are sufficient but not necessary to ensure stationarity and nonnegativity of h_t.

24. A description of this data set can be found in Croushore and Stark (2001).

25. The data is seasonally adjusted. For the vintage of 1996Q1, the observation for 1995Q4 is missing because of a delay in the release of statistical data caused by the Federal government shutdown. For most vintages, the data start from 1947Q1. For some vintages, data may start at a different date. So, the number of observations for estimation varies across vintages not only because more observations are included over time, but also because the changes of starting date. But this only occurs for 1992Q1–1992Q4, 1999Q4–2000Q1 (starting date is 1959Q1) and 1996Q1–1997Q2 (starting date is 1959Q3) with a total of 12 among 143 vintages.

26. This forecast actually has a horizon of $3^1/2$ quarters because forecasts are made at the middle of each quarter. As pointed out before, we refer to this forecast horizon as three quarters.

27. With real-time data, it is around February 15.

28. Baillie and Bollerslev (1992) derived the formula for calculating $E(\varepsilon_{t,k}^2|\psi_{t,i})$ for GARCH models. Similar formula for ARCH(4) and GJR–GARCH(1,1) can be found in Engle and Kraft (1983) and Blair, Poon, and Taylor (2001), respectively.

29. As noted before, we keep horizon effects when we compare survey measure with time series measure.

30. Time series measures of uncertainty in Figs. 3a–d is estimated and forecasted with GARCH(1,1) model. Although GJR–GARCH(1,1) model has a slightly better fit to the most recent data as found in Table 4, the profile of uncertainty over time are quite similar for these two models.

31. With a Markov-switching model, Evans and Wachtel (1993) estimated the regime uncertainty. According to their finding, regime uncertainty was low and stable after 1984, but high and volatile during 1968–1983. Although they use in-sample forecast based on revised data, their finding shows that the divergence between survey measures and ARCH-type measures is due to the omission of regime uncertainty from the latter.

32. This is the consensus forecast of our sample of regular forecasters who participated at least 12 surveys.

33. The quarters for which the survey measure is missing are not included when calculating the average of absolute forecast errors for GARCH(1,1).

34. Four-quarter-ahead forecast for annual inflation in 1974 and three-quarter-ahead forecast for annual inflation in 1975.

35. Taylor (1986) and Schwert (1989a, b) modeled conditional standard deviation as a distributed lag of absolute residuals. However, we found that their models could not reduce the spikes in 1975 and 1976.

36. As pointed out by Evans and Wachtel (1993), this model still fails to account for the effect of anticipated future shifts in inflation regime. They suggest a Markov-switching model to explain the structural break in inflation process caused by changing regimes. One direction for future research is to generalize their model to multi-period real-time forecasts.

37. Kim and Nelson (1998) provide a GAUSS program for estimating this model. We adapt their program to get one-period-ahead forecasts.

38. This is the only variable available in real-time data set at the Federal Reserve bank of Philadelphia web site as a measure of output gap.

39. The value of test statistic TR^2 was 38.52. The critical value for χ^2 test with 20 degrees of freedom is 31.4 at significance level 1%.

40. Hlouskova, Schmidheiny, and Wagner (2004) derived the general formula for the multi-step minimum mean squared error (MSE) prediction of the conditional means, variances and covariances for multivariate GARCH models with an application in portfolio management.

ACKNOWLEDGMENTS

We thank Rob Engle, Thomas Fomby, R. Carter Hill, Dek Terrell and other participants of the conference for many helpful comments and suggestions on a previous version of the paper. John Jones, Terrence Kinal, Paul Söderlind and Kenneth Wallis also made valuable comments. We, however, are responsible for any remaining errors and shortcomings.

REFERENCES

Baillie, R. T., & Bollerslev, T. (1992). Prediction in dynamic models with time-dependent conditional variance. *Journal of Econometrics*, *52*, 91–113.

Ball, L. (1992). Why does high inflation raise inflation uncertainty? *Journal of Monetary Economics*, *29*, 371–388.

Ball, L., & Mankiw, N. G. (1995). Relative-price changes as aggregate supply shocks. *Quarterly Journal of Economics*, *110*, 161–193.

Batchelor, R., & Dua, P. (1993). Survey vs. ARCH measures of uncertainty. *Oxford Bulletin of Economics and Statistics*, *55*, 341–353.

Batchelor, R., & Dua, P. (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science*, *41*, 68–75.

Batchelor, R., & Dua, P. (1996). Empirical measures of inflation uncertainty: A cautionary note. *Applied Economics*, *28*, 333–341.

Bera, A. K., & Higgins, M. L. (1993). ARCH models: Properties, estimation and testing. *Journal of Economic Surveys*, *7*, 305–362.

Blair, B. J., Poon, S. H., & Taylor, S. J. (2001). Forecasting S&P 100 volatility: the incremental information content of implied volatilities and high-frequency index return. *Journal of econometrics*, *105*, 5–26.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1990). Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics*, *72*, 498–505.

Bollerslev, T., Chou, R. Y., & Kroner, K. P. (1992). ARCH modeling in finance: a review of theory and empirical evidence. *Journal of Econometrics*, *52*, 5–59.

Bomberger, W. A. (1996). Disagreement as a measure of uncertainty. *Journal of Money, Credit and Banking*, *28*, 381–392.

Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *Quarterly Journal of Economics*, *118*, 269–298.

Christofferson, P. F., & Diebold, F. X. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, *13*, 808–817.

Cogley, T., Morozov, S. & Sargent, T. J. (2003). Bayesian fan charts for UK inflation: forecasting and sources of uncertainty in an evolving monetary system. Draft, July.

Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of econometrics*, *105*, 111–130.

Cukierman, A., & Wachtel, P. (1979). Differential inflationary expectations and the variability of the rate of inflation: theory and evidence. *American Economic Review*, *69*, 595–609.

Cukierman, A., & Wachtel, P. (1982). Relative price variability and nonuniform inflationary expectations. *Journal of Political Economy*, *90*, 146–157.

Davies, A., & Lahiri, K. (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics*, *68*, 205–227.

Davies, A., & Lahiri, K. (1999). Re-examining the rational expectations hypothesis using panel data on multi-period forecasts. In: C. Hsiao, K. Lahiri, L. F. Lee & H. Pesaran (Eds), *Analysis of Panels and Limited Dependent Variable Models* (pp. 226–254). Cambridge: Cambridge University Press.

Diebold, F., Tay, A., & Wallis, K. (1999). Evaluating density forecasts of inflation: The Survey of Professional Forecasters. In: R. Engle & H. White (Eds), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger* (pp. 76–90). Oxford: Oxford University Press.

Engle, R. F. (1982). Autoregressive conditiocnal heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*, 987–1008.

Engle, R. F. (1983). Estimates of the variance of US inflation based upon the ARCH model. *Journal of Money, Credit and Banking*, *15*, 286–301.

Engle, R. F., & Kraft, D. (1983). Multiperiod forecast error variances of inflation estimated from ARCH models. In: A. Zellner (Ed.), *Applied Time Series Analysis of Economic Data* (pp. 293–302). Washington, DC: Bureau of the Census.

Engle, R. F., & Watson, M. W. (1985). The Kalman filter: applications to forecasting and rational-expectations models. In: T. F. Bewley (Ed.), *Advances in Econometrics Fifth World Congress*, (Vol. 1, pp. 245–283). Cambridge: Cambridge University Press.

Ericsson, N. R. (2003). Forecast uncertainty in economic modeling. In: D. F. Hendry & N. R. Ericsson (Eds), *Understanding Economic Forecasts* (pp. 68–92). Cambridge: MIT Press.

Evans, M. (1991). Discovering the link between inflation rates and inflation uncertainty. *Journal of Money, Credit and Banking*, *23*, 169–184.

Evans, M., & Wachtel, P. (1993). Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit and Banking*, *25*, 475–511.

Friedman, M. (1977). Nobel Lecture: inflation and unemployment. *The Journal of Political Economy*, *85*, 451–472.

Fuhrer, J. C. (1988). On the information content of consumer survey expectations. *Review of Economics and Statistics*, *70*, 140–144.

Fulford, S.L. (2002). The impact of monetary policy on the cross-sectional variance of expectations of inflation. Mimeo, Stanford University.

Garratt, A., Lee, K., Pesaran, M. H., & Shin, Y. (2003). Forecast uncertainty in macroeconomic modelling: an application to the UK ecomomy. *Journal of the American Statistical Association*, *98*, 829–838.

Giordani, P., & Söderlind, P. (2003). Inflation forecast uncertainty. *European Economic Review*, *47*, 1037–1059.

Giordani, P. & Söderlind, P. (2005). Is there evidence of pessimism and doubt in subjective distributions? Implications for the equity premium puzzle. *Forthcoming in Journal of Economic Dynamics and Control*.

Glosten, L., Jagannathan, R., & Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*, 1779–1801.

Greene, W. H. (2000). *Econometric analysis 4th Ed.* Upper Saddle River, N J: Prentice Hall.

Harvey, A., Ruiz, E., & Sentana, E. (1992). Unobserved component time series models with ARCH disturbances. *Journal of Econometrics*, *52*, 129–157.

Hlouskova, J., Schmidheiny, K. & Wagner, M. (2004). *Multistep predictions for multivariate GARCH models: closed form solution and the value for portfolio management*. Université de Lausanne, Ecole des HEC, DEEP, EconPapers.

Irwin, S. H., Gerlow, M. E., & Liu, T. R. (1994). The forecasting performance of livestock futures prices: a comparison of USDA expert predictions. *Journal of Futures Markets*, *14*, 861–875.

Kastens, T. L., Schroeder, T. C., & Plain, R. (1998). Evaluation of extension and USDA price and production forecasts. *Journal of Agricultural and Resource Economics*, *23*, 244–261.

Kim, C.-J., & Nelson, C. R. (1998). *State-Space Model with Regime-Switching: Classical and Gibbs-sampling Approaches with Applications*. Cambridge, MA: MIT Press.

Kurz, M. (2002). Heterogeneous forecasting and Federal Reserve information. Mimeo, Stanford University.

Kurz, M., & Motolese, M. (2001). Endogenous uncertainty and market volatility. *Economic Theory*, *17*, 497–544.

Lahiri, K., & Ivanova, D. (1998). A time series and cross sectional analysis of consumer sentiment and its components. In: K. H. Oppenländer & G. Poser (Eds), *Social structural change–consequences for business cycle surveys* (pp. 337–362). England: Ashgate Publishing, Aldershot.

Lahiri, K. & Liu, F. (2005). Modeling multi-period inflation uncertainty using a panel of density forecasts. *Forthcoming in Journal of Applied Econometrics*.

Lahiri, K., & Teigland, C. (1987). On the normality of probability distribution and GNP forecasts. *International Journal of Forecasting*, *3*, 269–279.

Lahiri, K., Teigland, C., & Zaporowski, M. (1988). Interest rates and the subjective probability distribution of inflation forecasts. *Journal of Money, Credit and Banking*, *20*, 233–248.

Laster, D., Bennett, P., & Geoum, I. S. (1999). Rational bias in macroeconomic forecast. *The Quarterly Journal of Economics*, *114*, 293–318.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, *4*, 103–124.

Lucas, R. E. (1973). Some international evidence on output-inflation tradeoffs. *American Economic Review*, *63*, 326–334.

Mankiw, N. G., & Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new Keynesian Phillips curve. *Quarterly Journal of Economics*, *117*, 1295–1328.

Mankiw, N. G., Reis, R., & Wolfers, J. (2003). Disagreement about inflation expectation. *NBER Macroeconomics Annual*, *18*, 209–248.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*, 1329–1376.

McNees, S.(1989). Why do forecasts differ? *New England Economic Review*, (January/February), pp. 42–54

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370.

Rich, R., & Tracy, J. (2003). Modeling uncertainty: *Predictive accuracy as a proxy for predictive confidence*. Federal Reserve Bank of New York Staff Reports no.161.

Schwert, G. W. (1989a). Why does stock market volatility change over time. *Journal of Finance*, *44*, 1115–1153.

Schwert, G. W. (1989b). Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy*, *39*, 83–126.

Sims, C. (2002). The role of models and probabilities in the monetary policy process. *Brookings Papers on Economic Activity*, *2002*(2), 1–62.

Šmidková, K. (2003). Methods available to monetary policy makers to deal with uncertainty. Paper presented at the conference Forecasting in a Central Bank, Bank of England, August, London.

Souleles, N. (2004). Expectations, heterogeneous forecast errors, and consumption: Micro evidence from the Michigan Consumer Sentiment Surveys. *Journal of Money, Credit and Banking*, *36*, 39–72.

Stock, J., & Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, *44*, 293–335.

Stuart, A., & Ord, J.K. (1994). *Kendall's advanced theory of statistics* (Vol. 1). New York: Halsted Press.

Taylor, S. (1986). *Modeling financial time Series*. New York: Wiley and Sons.

Wallis, K.F. (2004). Combining density and interval forecasts: A modest proposal. Mimeo, University of Warwick, UK.

Zarnowitz, V., & Lambros, L. A. (1987). Consensus and uncertainty in economic prediction. *Journal of Political Economy*, *95*, 591–620.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, *81*, 446–451.

This page intentionally left blank

NECESSARY AND SUFFICIENT RESTRICTIONS FOR EXISTENCE OF A UNIQUE FOURTH MOMENT OF A UNIVARIATE GARCH(*P*,*Q*) PROCESS☆

Peter A. Zadrozny

ABSTRACT

A univariate GARCH(p,q) process is quickly transformed to a univariate autoregressive moving-average process in squares of an underlying variable. For positive integer m, eigenvalue restrictions have been proposed as necessary and sufficient restrictions for existence of a unique mth moment of the output of a univariate GARCH process or, equivalently, the 2mth moment of the underlying variable. However, proofs in the literature that an eigenvalue restriction is necessary and sufficient for existence of unique 4th or higher even moments of the underlying variable, are either incorrect, incomplete, or unnecessarily long. Thus, the paper contains a short and general proof that an eigenvalue restriction is

☆The paper represents the author's views and does not necessarily represent any official positions of the Bureau of Labor Statistics. Thanks to an anonymous referee, L. Giraitis, D. Koulikov, R. Leipus, H. Neudecker, and D. Surgailis for comments.

Econometric Analysis of Financial and Economic Time Series/Part A
Advances in Econometrics, Volume 20, 365–379

ISSN: 0731-9053/doi:10.1016/S0731-9053(05)20013-0

necessary and sufficient for existence of a unique 4th moment of the underlying variable of a univariate GARCH process. The paper also derives an expression for computing the 4th moment in terms of the GARCH parameters, which immediately implies a necessary and sufficient inequality restriction for existence of the 4th moment. Because the inequality restriction is easily computed in a finite number of basic arithmetic operations on the GARCH parameters and does not require computing eigenvalues, it provides an easy means for computing "by hand" the 4th moment and for checking its existence for low-dimensional GARCH processes. Finally, the paper illustrates the computations with some GARCH(1,1) processes reported in the literature.

1. INTRODUCTION

A univariate generalized autoregressive conditional heteroskedasticity or GARCH(p,q) process (Engle, 1982; Bollerslev, 1986) is quickly transformed to a univariate autoregressive moving-average (ARMA) process in squares of an underlying variable. Henceforth, for brevity, unless otherwise qualified, we indicate "GARCH(p,q)" by "GARCH." GARCH processes and their generalizations (Mittnik, Paolella, & Rachev, 2002) have been used to model volatilities or time-varying variances of underlying variables, usually financial-asset returns or residuals from estimated time-series models. GARCH processes are usually considered to have Gaussian or normally distributed disturbances, however, because asset returns can have large volatilities, such as the unexpected large drop in the U.S. stock market in October 1987, their distributions are often ascribed "heavier" tails than implied by the Gaussian distribution. Thus, Mandelbrot (1963a,b) studied asset-return distributions using stable-Paretian distributions. More recently, McCulloch (1997), Rachev and Mittnik (2000), and others studied time series of asset returns using GARCH and other processes driven by stable-Paretian and other heavy-tailed disturbances. Thus, it is useful to have a method for easily computing the 4th moment of an underlying variable of a univariate GARCH process, to check whether it exists and, if so, whether it indicates heavier than Gaussian tails. All the moments considered here are unconditional.

For positive integer m, eigenvalue restrictions have been proposed as necessary and sufficient conditions for existence of a unique 2mth moment of the underlying variable of a univariate GARCH process. Proofs in the literature that an eigenvalue restriction is necessary and sufficient for existence of unique 4th or higher even moments of the underlying variable of a GARCH process are either incorrect, incomplete, or unnecessarily long. Before detailing the present paper's contribution to this literature, we clarify our use of the term "4th moment." A GARCH process is a type of ARMA process that linearly transforms squared disturbances, ε_t^2, to squared variables, y_t^2, so that the process nonlinearly transforms unsquared or underlying disturbances, ε_t, to unsquared or underlying variables, y_t. Thus, the "2mth moment of the underlying variable of a GARCH process" is the 2mth moment of y_t and is equivalent to the mth moment of y_t^2. In this regard, a GARCH process is (covariance) stationary if and only if the underlying variable, y_t, has a 2nd moment. Thus, unless otherwise qualified, "4th moment" means the 4th moment of y_t.

The paper contains a short and general proof that an eigenvalue restriction is necessary and sufficient for existence of the unique 4th moment of the underlying variable of a univariate GARCH process. The paper derives an expression for computing the 4th moment in terms of the GARCH parameters, which immediately implies a necessary and sufficient inequality restriction for checking the moment's existence. Because the eigenvalue and inequality restrictions are separately necessary and sufficient for the moment's existence, they are equivalent. Because the inequality restriction is easily computed in a finite number of basic arithmetic operations on the GARCH parameters and does not require computing eigenvalues, it provides an easy means for computing "by hand" the 4th moment and for checking its existence for low-dimensional GARCH processes.

The recent literature contains the following related results. In the following, all statements of "necessity" and "sufficiency" refer to the existence of the 4th moment of the underlying variable of a general univariate GARCH process.

He and Terasvirta (1999) stated an inequality restriction (Theorem 1, p. 827) on a univariate GARCH process, which they claim (pp. 833–840) is necessary and sufficient. Ling and McAleer (2002) question (pp. 724–728, note 1) whether He and Terasvirta's necessity proof is complete. In any case, because they do not use a state-space representation, He and Terasvirta's discussion is unnecessarily long. Karanasos (1999) stated (Theorem 3.1, p. 66) an inequality restriction and proved (pp. 73–74), its necessity, but, as Ling and McAleer noted (pp. 723–724), he did not prove its sufficiency. Ling (1999) stated and proved (Theorem 6.2, p. 702) an eigenvalue restriction's sufficiency.

Ling and McAleer (2002) stated (Theorem 2.1, pp. 723–724) an eigenvalue restriction and purportedly proved (pp. 726–727) its necessity. However, they actually proved necessity only for the special case $\alpha_1 > 0$. Their proof works if and only if (iff) $\alpha_1 > 0$ or $\beta_1 > 0$, but not if $\alpha_1 = \beta_1 = 0$. Ling and McAleer claim (p. 727) their proof still works if $\alpha_1 = \beta_1 = 0$, but as written it does not because it requires a vector R to have all positive elements after a certain number of repeated steps and this is not the case if $\alpha_1 = \beta_1 = 0$. By contrast, the proof given here holds for any univariate GARCH process.

A stationary GARCH process has an equivalent infinite Volterra-type representation called an ARCH(∞) representation. Giraitis and Surgailis (2002, theorem 3.1) and Kazakevicius, Leipus, and Viano (2004, theorem 2.1) derived equivalent inequality restrictions on an ARCH(∞) representation of a univariate GARCH process which are necessary and sufficient for existence of a unique 4th moment of the underlying variable of the process and which may be compared with the present inequality restriction. See also Giraitis, Leipus, and Surgailis (2005, section 2).

Eigenvalue restrictions require computing eigenvalues, which can generally be done analytically "by hand" only for matrices no larger than three dimensional, hence, for GARCH processes with no more than three lags. Thus, generally, eigenvalue restrictions can be checked only numerically on a computer. By contrast, as illustrated in Section 4, the present necessary and sufficient inequality restriction for the 4th moment's existence is easily checked "by hand" in a finite number of basic arithmetic operations.

The remainder of the paper is organized as follows. Section 2 states a univariate GARCH process in state-space form in order to derive the 4th-moment inequality restriction. Section 3 proves that the inequality restriction is necessary and sufficient for existence of a unique 4th moment of the underlying variable of a univariate GARCH process of any degree. Section 3 also proves that the 4th-moment inequality restriction is equivalent to an eigenvalue restriction. Section 4 numerically illustrates the 4th-moment inequality restriction with six GARCH(1,1) processes from the literature. Section 5 contains concluding remarks.

2. STATE-SPACE FORM OF A GARCH(*P*,*Q*) PROCESS

For discrete-time periods t, let $y_t^2 = \varepsilon_t^2 h_t$ denote the square of the underlying variable in the process, where h_t is generated by the univariate GARCH(p,q) process

$$h_t = \alpha_0 + \sum_{i=1}^{n} \alpha_i \varepsilon_{t-i}^2 h_{t-i} + \sum_{i=1}^{n} \beta_i h_{t-i} \tag{2.1}$$

α_i and β_i are constant parameters, $n = \max(p, q)$, $\alpha_i = 0$ for $i > p$ if $p < n$, $\beta_i = 0$ for $i > q$ if $q < n$, and ε_t is a disturbance. Although the aim is to verify necessary and sufficient restrictions on the GARCH(p,q) parameters for existence of the 4th moment of the underlying variable, $\mathrm{E}y_t^4 = \mathrm{E}\left(\varepsilon_t^4 h_t^2\right)$, because we assume the disturbance's 4th moment, $\mathrm{E}\varepsilon_t^4$, exists, we concentrate on restrictions for existence of $\mathrm{E}h_t^2$.

Notation is simplified but no generality is lost when process (2.1) is written as a GARCH(n,n) process. The proofs in Section 3 do not depend on the presence of the second summation in (2.1), $\sum_{i=1}^{n} \beta_i h_{t-i}$, hence, on whether $q \geq 1$ or $q \geq 0$. The following GARCH assumptions allow $q \geq 0$, but require $p \geq 1$.

We assume the following for GARCH process (2.1): (i) $n = \max(p,q)$, for $p \geq 1$ and $q \geq 0$; (ii) $\alpha_0 > 0$, (iii) $\alpha_i \geq 0$ and $\beta_i \geq 0$, for $i = 1, \ldots, n$; (iv) $\alpha_i > 0$, for one or more $i = 1, \ldots, n$; (v) $\alpha_n > 0$ or $\beta_n > 0$, or both; and (vi) ε_t is distributed identically, independently, with zero mean, $\mathrm{E}\varepsilon_t = 0$, finite positive variance, $\sigma_\varepsilon^2 = \mathrm{E}\varepsilon_t^2 > 0$, and finite positive 4th moment, $\sigma_\varepsilon^4 = \mathrm{E}\varepsilon_t^4 > 0$ which is the central and noncentral 4th moment of ε_t because $\mathrm{E}\varepsilon_t = 0$. As usual, without loss of generality, we set $\sigma_\varepsilon^2 = 1$ and, thus, effectively merge σ_ε^2 into α_i.

Assumption (v) is convenient but unnecessary, because if $\alpha_n = \beta_n = 0$, then, assumption (iv) guarantees that we can reduce n until $\alpha_i > 0$ for some $i = 1, \ldots, n-1$. We take "ε_t is distributed independently" to mean that ε_t is distributed independently not just of past values of itself but also of past values of h_t. We do not need any particular distributional assumption such as Gaussianity.

Finally, we assume GARCH process (2.1) is stationary. Throughout, by "stationarity" we mean weak or covariance stationarity. By contrast, for example, Nelson (1990) considers strong stationarity of GARCH(1,1) processes. Following Milhoj (1985), Bollerslev (1986) proved that, for $\alpha_i \geq 0$ and $\beta_i \geq 0$, process (2.1) is stationary iff (vii) $\sum_{i=1}^{n} f_i < 1$ where $f_i = \alpha_i + \beta_i$. Although the final Theorem 4 here proves that GARCH process (2.1) is stationary if $\mathrm{E}h_t^2$ exists, we assume stationarity in order to simplify the discussion.

In sum, we assume (i)–(vii) for GARCH process (2.1) and call these "the GARCH assumptions."

If GARCH process (2.1) is stationary, then, its mean, $\mathrm{E}h_t = \mu$, exists, is positive, and is given by

$$\mu = \alpha_0 \Big/ \left(1 - \sum_{i=1}^{n} f_i\right) \tag{2.2}$$

Let a tilde denote a mean-adjusted variable, so that $\tilde{h}_t = h_t - \mu$ and $\tilde{\varepsilon}_t^2 = \varepsilon_t^2 - \sigma_\varepsilon^2 = \varepsilon_t^2 - 1$. Then, we write the mean-adjusted form of GARCH process (2.1) in the ARMA form

$$\tilde{h}_t = \sum_{i=1}^{n} f_i \tilde{h}_{t-i} + \sum_{i=1}^{n} g_i \xi_{t-i} \tag{2.3}$$

where $\xi_t = \tilde{\varepsilon}_t^2 h_t$, $f_i = \alpha_i + \beta_i$, and $g_i = \alpha_i$.

A state-space form of an ARMA process comprises an observation equation and a state equation in terms of a state vector. Let $x_t = (x_{1,t}, \ \ldots \ , x_{n,t})^{\mathrm{T}}$ denote an $n \times 1$ state vector, where superscript T denotes transposition. Then, following Ansley and Kohn (1983), we can write ARMA Eq. (2.3) in state-space form, with observation equation $\tilde{h}_t = e_1^{\mathrm{T}} x_t$, where $e_1 = (1, 0, \ \ldots \ 0)^{\mathrm{T}}$ is the $n \times 1$ vector with first element 1 and all other elements 0, and state equation

$$x_t = F x_{t-1} + g \xi_{t-1}, \ F = \begin{bmatrix} f_1 & 1 & 0 & . & . & . & 0 \\ f_2 & 0 & 1 & . & & & . \\ . & . & . & . & . & & . \\ . & . & & . & . & . & . \\ . & . & & & . & . & 0 \\ . & . & & & & . & 1 \\ f_n & 0 & . & . & . & . & 0 \end{bmatrix}, \ g = \begin{bmatrix} g_1 \\ . \\ . \\ . \\ . \\ . \\ g_n \end{bmatrix} \tag{2.4}$$

Observation equation $\tilde{h}_t = e_1^T x_t$ implies $\tilde{h}_t$ is the first element of x_t. Because F is a companion matrix, its eigenvalues are identical to the roots of the characteristic equation $\lambda^n - f_1 \lambda^{n-1} - \ \ldots \ - f_{n-1}\lambda - f_n = 0$ (Anderson, 1958, p. 177), so that stationarity condition $\sum_{i=1}^{n} f_i < 1$ is equivalent to $\rho(F) < 1$, where $\rho(F)$ denotes the spectral radius or maximal absolute eigenvalue of F. We say that the state Eq. (2.4) is stationary and the state-transition matrix F is stable iff $\rho(F) < 1$.

3. 4TH-MOMENT INEQUALITY AND EIGENVALUE RESTRICTIONS

We are interested in proving that derived restrictions for existence of $\mathrm{E}y_t^4 = \mathrm{E}(\varepsilon_t^4 h_t^2)$ are necessary and sufficient. However, we assume that $\mathrm{E}\varepsilon_t^4$ exists; $\mathrm{E}h_t^2$ and $\mathrm{E}\tilde{h}_t^2$ are linked by $\mathrm{E}h_t^2 = \mathrm{E}\tilde{h}_t^2 + \mu^2$; $\mathrm{E}\tilde{h}_t^2$ and V are linked by $\mathrm{E}\tilde{h}_t^2 = v_{11} = e_1^T V e_1$; and, Theorem 2 proves that v_{11} exists iff V exists, where v_{11} and

V are defined two paragraphs below. Thus, proving that restrictions for existence of $\mathrm{E}y_t^4$ are necessary and sufficient reduces to proving such restrictions for V.

We shall generally maintain two terminological conventions. First, when we write an inequality such as "$x \geq 0$" we shall mean not only that the inequality holds but also that the variable in it exists and, hence, is finite. Second, as noted before, we refer to assumptions (i)–(vii) on GARCH process (2.1) as "the GARCH assumptions."

Post-multiplying Eq. (2.4) by its transpose, taking unconditional expectation, E, of the result, and apply independence assumption (vi), so that $\mathrm{E}x_t\xi_t^{\mathrm{T}} = 0_{n\times 1}$, the $n \times 1$ zero vector, implies the Lyapunov equation $V = FVF^{\mathrm{T}} + \left(\mathrm{E}\xi_t^2\right)gg^{\mathrm{T}}$, where $V = \mathrm{E}x_t x_t^{\mathrm{T}}$. Then,

$$\mathrm{E}\xi_t^2 = \mathrm{E}\left(\tilde{\varepsilon}_t^2\right)^2\left(\mathrm{E}\tilde{h}_t^2 + 2\mathrm{E}\tilde{h}_t\mu + \mu^2\right) = \mathrm{E}\left(\tilde{\varepsilon}_t^2\right)^2\left(\mathrm{E}\tilde{h}_t^2\right) + \mu^2$$

because $\tilde{\varepsilon}_t$ and $\tilde{h}_t$ are stochastically independent and $E\tilde{h}_t = 0$. In particular, $\tilde{h}_t^2$ depends only on constant parameters and variables dated before period t. Thus, the Lyapunov equation is equivalent to

$$V = FVF^{\mathrm{T}} + \theta g e_1^{\mathrm{T}} V e_1 g^{\mathrm{T}} + \mu^2\theta g g^{\mathrm{T}} \tag{3.1}$$

where $\theta = \mathrm{E}(\tilde{\varepsilon}_t^2)^2 = \mathrm{E}\varepsilon_t^4 - (\mathrm{E}\varepsilon_t^2)^2 = \sigma_\varepsilon^4 - (\sigma_\varepsilon^2)^2 = \sigma_\varepsilon^4 - 1 > 0$ (e.g., by Jensen's inequality). When ε_t is Gaussian, $\sigma_\varepsilon^4 = 3$, so that $\theta = 2$.

Using $\mathrm{vec(ABC)} = [C^{\mathrm{T}} \otimes A]\mathrm{vec}(B)$, for matrices A, B, and C conformable to the product ABC (Magnus & Neudecker, 1988, p. 30), where $\mathrm{vec}(\cdot)$ denotes the column vectorization of a matrix (column 1 on top of column 2, etc.) and $\otimes$ denotes the Kronecker product. We state Eq. (3.1) equivalently as

$$w = Aw + \mu^2\theta(g \otimes g) \tag{3.2}$$

where $w = \mathrm{vec}(V)$ and $A = F \otimes F + \theta(g \otimes g)(e_1 \otimes e_1)^{\mathrm{T}}$.

Gantmacher (1959, pp. 50–57) discusses the following implications of irreducibility. A matrix or vector, M, is nonnegative ($M \geq 0$) or positive ($M > 0$) iff all of its elements are nonnegative or positive. A real, $n \times n$, and nonnegative matrix, M, is irreducible iff it has no invariant coordinate subspace with dimension less than n. A theorem by Frobenius says that if M is a real, $n \times n$, nonnegative, and irreducible matrix, then, M has an eigenvalue, λ, and associated left or right eigenvector, z, such that λ is real, positive, and equal to the maximal absolute eigenvalue of M or $\lambda = \rho(M)$, and z is real and positive. We need the following lemma.

Lemma 1. Assume that GARCH assumptions (i)–(vii) hold. Then, matrices $F \geq 0$ and $A = F \otimes F + \theta(g \otimes g)(e_1 \otimes e_1)^T \geq 0$ are irreducible.

Proof of Lemma 1. Let e_i denote the $n \times 1$ elementary vector i, with one in position i and zeroes elsewhere. The n-dimensional real vector space is spanned by $e_1, \ldots, e_n$. Consider $f_1 = 0, \ldots, f_{n-1} = 0$ and n successive mappings with F starting from e_n: $Fe_n = e_{n-1}$, $F^2 e_n = e_{n-2}$, $\ldots$, $F^{n-1} e_n = e_1$, and $F^n e_n = f_n e_n$. The first $n-1$ mapped vectors are equal to the first $n-1$ elementary vectors, $e_1, \ldots, e_{n-1}$; because $f_n > 0$, the last mapped vector lies in the space spanned by the last elementary vector, e_n. Thus, starting from e_n and successively mapping n times with F, the mapped vectors span the n-dimensional real vector space. The same conclusion holds if we start the mappings from any other elementary vector and if $f_1 \geq 0, \ldots, f_{n-1} \geq 0$. Thus, F is irreducible. $F \otimes F \geq 0$ and, by a similar argument in $n \times n$-block form, is irreducible. Because $F \otimes F \geq 0$ and $\neq 0$, $\theta > 0$, and $(g \otimes g)\ (e_1 \otimes e_1)^{\mathrm{T}} \geq 0$, adding $\theta(g \otimes g)\ (e_1 \otimes e_1)^{\mathrm{T}} \geq 0$ to $F \otimes F$ cannot decrease the nonnegative elements of the first block-column of $F \otimes F$ and does not change its irreducibility. Thus, $A = F \otimes F + \theta(g \otimes g)(e_1 \otimes e_1)^{\mathrm{T}}$ is irreducible and Lemma 1 is proved.

The proof of Lemma 1 depends on $f_n > 0$, so that if $f_n = 0$ because $\alpha_n = \beta_n = 0$, maintaining the lemma requires reducing n until $f_n > 0$.

3.1. Necessary and Sufficient Inequality Restriction

Stationarity or $\rho(F) < 1$ implies that $B = \sum_{i=0}^{\infty} F^i g g^{\mathrm{T}} (F^{\mathrm{T}})^i \geq 0$ exists (Wilkinson, 1965, p. 59), so that Eq (3.1) can be stated equivalently as

$$V = \left(v_{11} + \mu^2\right)\theta B \tag{3.3}$$

where $v_{11} = e_1^{\mathrm{T}} V e_1$, the (1,1) element of V. Pre-multiplying Eq. (3.3) by e_1^{T}, post-multiplying the result by e_1, and rearranging, leads to Eq. (3.4), where $b_{11} = e_1^{\mathrm{T}} B e_1$, the (1,1) element of B. Thus, we have Theorem 1, which states the necessary and sufficient restriction $0 < \theta b_{11} < 1$ for existence of $\mathrm{E}\tilde{h}_t^2 = v_{11} > 0$.

Theorem 1. Assume that GARCH assumptions (i)–(vii) hold. Then,

$$v_{11} = \mu^2 \left(\frac{\theta b_{11}}{1 - \theta b_{11}} \right) > 0 \tag{3.4}$$

exists iff, in addition, $0 < \theta b_{11} < 1$.

Proof of Theorem 1. The GARCH assumptions imply that $\mu > 0$ and $\theta > 0$ exist. In particular, stationarity implies that $B \geq 0$ exists. Because $F \geq 0$ and is irreducible (Lemma 1) and $g \geq 0$ and $\neq 0$, $b_{11} = e_1^{\mathrm{T}} B e_1 \geq e_1^{\mathrm{T}} \mathrm{F}^j g g^{\mathrm{T}} \left(F^{\mathrm{T}}\right)^j e_1 > 0$, for some positive integer j. Thus, $\mu^2 \theta b_{11} > 0$ and Eq. (3.4) implies that $v_{11} > 0$ exists iff, in addition, $\theta b_{11} < 1$ and Theorem 1 is proved.

Next, Eq. (3.3) implies Theorem 2, which links the existence of V with the existence of v_{11}.

Theorem 2. Assume that GARCH assumptions (i)–(vii) hold. Then, $V \geq 0$ exists iff $v_{11} > 0$ exists.

Proof of Theorem 2. Suppose that $V \geq 0$ exists. Then, $v_{11} = e_1^{\mathrm{T}} V \mathrm{e}_1 \geq 0$ exists. The GARCH assumptions imply that $\mu > 0$ and $\theta > 0$ exist, that F is stable, nonnegative, and irreducible, and that $g \geq 0$ and $\neq 0$. These properties imply that $B \geq 0$ and $b_{11} > 0$ exist. Thus, Eq. (3.3) implies that $v_{11} = (v_{11} + \mu^2)\theta b_{11} \geq \mu^2 \theta b_{11} > 0$ and necessity is proved. Suppose that $v_{11} > 0$ exists. The GARCH assumptions imply that $\mu > 0$, $\theta > 0$, and $B \geq 0$ exist. Thus, Eq. (3.3) implies that $V \geq 0$ exists and sufficiency and Theorem 2 are proved.

3.2. Necessary and Sufficient Eigenvalue Restriction

Eq. (3.2) has the unique solution $w = \mu^2 \theta (I_{n^2} - A)^{-1} (g \otimes g)$ iff $(I_{n^2} - A)$ is nonsingular, where I_{n^2} denotes the $n^2 \times n^2$ identity matrix, and this occurs iff A has no eigenvalue equal to one, because each eigenvalue of $(I_{n^2} - A)$ is one minus an eigenvalue of A. Theorem 3 tells us that $v_{11} > 0$ only if $\rho(A) < 1$, because otherwise w might have unacceptable negative values.

Theorem 3. Assume that GARCH assumptions (i)–(vii) hold. Then, $w = \mathrm{vec}(V) \geq 0$ and $v_{11} = w_1 > 0$ exist, where w_1 denotes the first element of w, iff, in addition, $\rho(A) < 1$.

Proof of Theorem 3. Suppose that $\rho(A) < 1$. Then, $\rho(I_{n^2} - A) < 1$, so that $w = \mu^2 \theta (I_{n^2} - A)^{-1} (g \otimes g)$ uniquely solves Eq. (3.2). The GARCH assumptions imply that $\theta > 0$ and $\mu > 0$ exist. The solution $w = \mu^2 \theta (I_{n^2} - A)^{-1} (g \otimes g)$ implies that $w = \mu^2 \theta (I_{n^2} + A + A^2 + \ldots)(g \otimes g)$. The irreducibility of A implies that $(e_1 \otimes e_1)^{\mathrm{T}} A^j (g \otimes g) > 0$, for some positive integer j. See the proof of Lemma 1. Thus, $v_{11} = w_1 > 0$ and sufficiency is proved. Suppose that $\delta x^{\mathrm{T}} = x^{\mathrm{T}} A$, where δ is a maximal eigenvalue of A, so that $|\delta| = \rho(A)$, and x is an associated left eigenvector of A. Because A

is real, nonnegative, and irreducible, Frobenius's theorem (Gantmacher, 1959, pp. 50–57) implies that δ and x are real and positive. Pre-multiplying Eq. (3.2) by x^{T} implies that

$$(1-\delta)x^{\mathrm{T}}w = \mu^2\theta x^{\mathrm{T}}(g\otimes g) \tag{3.5}$$

However, $x>0$, $w\geq 0$, $w_1>0$, and $g\otimes g\geq 0$ and $\neq 0$ imply that $x^{\mathrm{T}}w>0$ and $x^{\mathrm{T}}(g\otimes g)>0$. Thus, $\mu^2\theta>0$ implies that $(1-\delta)=\mu^2\theta x^{\mathrm{T}}(g\otimes g)/x^{\mathrm{T}}w>0$, so that Eq. (3.5) implies that $\delta=\rho(A)<1$ and necessity and Theorem 3 are proved.

Putting together Theorems 1–3 implies that $\theta b_{11}<1$ and $\rho(A)<1$ are equivalent necessary and sufficient restrictions for existence of the 4th moment of the underlying variable of GARCH process (2.1). We state this conclusion formally as Corollary 1.

Corollary 1. Assume that GARCH assumptions (i)–(viii) hold. Then, $\theta b_{11}<1$ iff $\rho(A)<1$.

Statistics tells us that a 4th moment exists only if the corresponding 2nd moment exists. Here, this means that $\rho(A)<1$ only if $\rho(F)<1$. Theorem 4 specializes this result and shows that 4th-moment existence implies stronger restrictions than stationarity or 2nd-moment existence. For example, under the GARCH assumptions, stationarity occurs iff $B\geq 0$ and $b_{11}>0$ exist, but 4th-moment existence occurs iff, in addition, $\theta b_{11}<1$.

Theorem 4. $\rho(F)^2<\rho(A)$, where $A=F\otimes F+\theta(g\otimes g)(e_1\otimes e_1)^{\mathrm{T}}$

Proof of Theorem 4. Let $\lambda^2(z\otimes z)^{\mathrm{T}}=(z\otimes z)^{\mathrm{T}}(F\otimes F)$ and $\delta x=Ax$, where λ and δ are maximal eigenvalues of F and A, so that $|\lambda|=\rho(F)$ and $|\delta|=\rho(A)$, and z and x are associated left and right eigenvalues of F and A. The GARCH assumptions and Frobenius's theorem imply that λ, δ, z, and x are real and positive. Pre-multiplying $\delta x=Ax$ by $(z\otimes z)^{\mathrm{T}}$ and rearranging implies that

$$\left(\delta-\lambda^2\right)(z\otimes z)^{\mathrm{T}}x=\theta\left(z^{\mathrm{T}}g\right)^2x_1 \tag{3.6}$$

where x_1 denotes the first element of x. However, $\theta>0$, $z>0$, $x>0$, $g\geq 0$, and $g\neq 0$, imply that $(z\otimes z)^{\mathrm{T}}x>0, z^{\mathrm{T}}g>0$, and $x_1>0$, so that Eq. (3.6) implies that $(\delta-\lambda^2)=\theta(z^{\mathrm{T}}g)^2x_1/(z\otimes z)^{\mathrm{T}}x>0$ and Theorem 4 is proved.

4. ILLUSTRATION OF THE 4TH-MOMENT INEQUALITY RESTRICTION

We now illustrate Eq. (3.4) in Table 1 with six GARCH(1,1) processes from the literature. To do this, we first write b_{11} in terms of a finite number of basic arithmetic operations on f_i and g_i. Pre-multiplying

$$B = \sum\nolimits_{i=0}^{\infty} F^i g g^{\mathrm{T}} (F^{\mathrm{T}})^i$$

by F, post-multiplying the result by F^{T}, subtracting the result from the initial equation for B, and rearranging, implies that

$$B = FBF^{\mathrm{T}} + gg^{\mathrm{T}} \tag{3.7}$$

which is a Lyapunov equation, linear in B.

The companion form of F suggests that Eq. (3.7) can be solved for b_{11} by eliminating elements of B from "back to front" until only b_{11} remains. Although such an approach might generally be overly complex and impractical, it is easily applied when $n = 2$. We consider $n = 2$ in part because it includes nearly all the GARCH processes we have seen in the empirical literature. For $n = 2$, Eq. (3.7) implies that

$$b_{11} = \frac{(1 - f_2)(g_1^2 + g_1^2) + 2f_1 g_1 g_2}{(1 - f_2)(1 - f_1^2 - f_2^2) - 2f_1^2 f_2} \tag{3.8}$$

The GARCH nonegativity assumptions $f_i \geq 0$ and $g_i \geq 0$ and the stationarity assumption $\sum_{i=1}^{n} f_i < 1$ imply that the numerator and denominator in Eq. (3.8) are both positive.

Although a GARCH process linearly transforms squared disturbances, ε_t^2, into squared variables, y_t^2, it nonlinearly transforms ε_t to y_t, so that y_t is non-Gaussian even when ε_t is Gaussian. We may consider this nonpreservation

Table 1. Examples of θb_{11} and k_y/k_ε for GARCH(1,1) Processes and $\theta = 2$.

Case	Authors	α_1	β_1	f_1	θb_{11}	k_y/k_ε
1	Engle (1982)	0.955	0.000	0.955	20.6	∞
2	Bollerslev (1986)	0.135	0.829	0.964	0.516	2.07
3	Baillie–Bollerslev (1989)	0.061	0.910	0.971	0.130	1.15
4	Bollerslev (1987)	0.057	0.921	0.978	0.149	1.18
5	Drost–Klaassen (1997)	0.052	0.932	0.984	0.170	1.21
6	Hsieh (1989)	0.191	0.806	0.997	12.2	∞

of Gaussianity in terms of kurtosis, by considering Eq. (3.4) as

$$\frac{k_y}{k_e} = \frac{\theta b_{11}}{1 - \theta b_{11}} + 1 \tag{3.9}$$

where $k_y = \mathrm{E}y_t^4/(\mathrm{E}y_t^2)^2$ and $k_\varepsilon = \mathrm{E}\varepsilon_t^4/(\mathrm{E}\varepsilon_t^2)^2$ are the kurtoses of y_t and ε_t. Eq. (3.9) implies that a GARCH transformation always increases kurtosis.

Table 1 considers only the case of $n = 1$ or GARCH(1,1) processes, which covers most of the empirical literature. Some exceptions are Engle (1982) and Geweke (1988) who consider nonstationary GARCH(0,4) and GARCH(0,2) processes which we exclude because they do not have 4th moments. In Table 1, we assume that $\varepsilon_t \sim N(0,1)$ (Gaussian, zero mean, unit variance) even though Bollerslev, Drost, and Klaassen, and Hsieh assume that ε_t is non-Gaussian. The cases in Table 1 are ordered by increasing $f_1 = \alpha_1 + \beta_1$ and, for $n = 1$, $\sigma_\varepsilon^2 = 1$, and $\theta = 2$, Eqs. (3.8) and (3.9) reduce to

$$b_{11} = \frac{g_1^2}{1 - f_1^2} = \frac{\alpha_1^2}{1 - (\alpha_1 + \beta_1)^2} \tag{3.10}$$

$$\frac{k_y}{k_\varepsilon} = \frac{1 - f_1^2}{1 - f_1^2 - 2g_1^2} = \frac{1 - (\alpha_1 + \beta_1)^2}{1 - (\alpha_1 + \beta_1)^2 - 2\alpha_1^2} \tag{3.11}$$

Table 1 contains values of θb_{11} and k_y/k_ε, for $\theta = 2$, according to Eqs. (3.10) and (3.11) for six GARCH(1,1) processes from the literature.

Table 1 depicts small α_1 (cases 2–5) associated with small $\theta b_{11} < 1$, for $\theta = 2$, and finite k_y/k_ε, larger and large α_1 (cases 6 and 1) associated with large $\theta b_{11} > 1$ and infinite k_y/k_ε, and, this pattern appears to be independent of the value of β_1 and to depend more on the value of α_1 (compare cases 1 and 6 with cases 2–5). The pattern is confirmed in Fig. 1, which depicts the feasible area of α_1 and β_1 in which the underlying variable in a scalar GARCH(1,1) process with $\varepsilon_t \sim N(0,1)$ has 2nd and 4th moments. The feasible area for stationarity or 2nd-moment existence is between the horizontal-α_1 axis, the vertical-β_1 axis, and the straight line $\beta_1 = 1 - \alpha_1$; the feasible subarea for 4th-moment existence is between the axes and the curved line $\beta_1 = -\alpha_1 + \sqrt{1 - 2\alpha_1^2}$. Evidently, 4th-moment existence restricts both α_1 and β_1, but restricts α_1 more: if a univariate GARCH(1,1) process is stationary and its underlying variable has a 4th moment, then, a minimum $\alpha_1 = 0$ implies a maximum $\beta_1 = 1$, but a minimum $\beta_1 = 0$ implies a maximum $\alpha_1 = 1/\sqrt{3} \doteq 0.5773$, because β_1 is either real and negative or complex with a negative real part when $1/\sqrt{3} < \alpha_1 < 1$. Adding 4th-moment existence

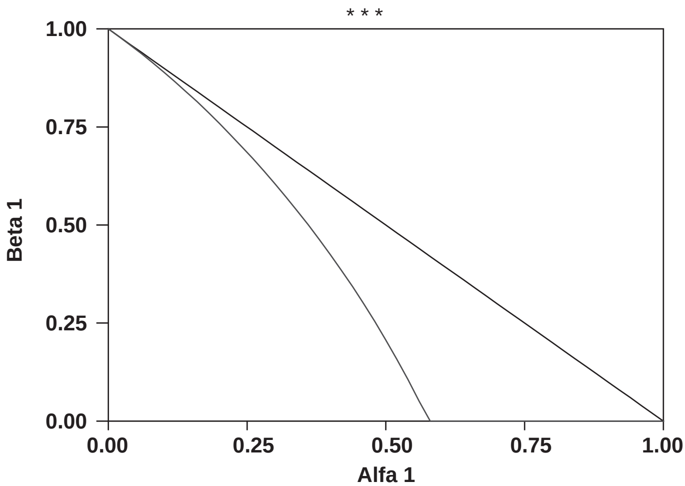


Fig. 1. Parameter Regions for 2nd- and 4th-Moment Existence.

to 2nd-moment existence, reduces the feasible area of α_1 and β_1 by about one third. Adding more GARCH lags ($n > 1$) or higher even-moment existence further reduces the feasible area of α_1 and β_1.

Eqs. (3.8)–(3.11), Table 1, and Fig. 1 illustrate an important advantage of inequality restriction $\theta b_{11} < 1$ over eigenvalue restriction $\rho(A) < 1$. For $n = 2$, Eq. (3.8) states $\theta b_{11} < 1$ in terms of a finite number of basic arithmetic operations on the GARCH parameters, and, for $n = 1$, Table 1 evaluates and Fig. 1 depicts this inequality. In principle, this could be done for any n, but, when n is large enough, algebraic detail becomes overwhelming. By contrast, generally, an eigenvalue restriction can be written out explicitly in terms of the GARCH parameters only when $n \leq 3$. When $n > 3$, this can be done only in unlikely cases in which the characteristic polynomial of $A = F \otimes F + \theta(g \otimes g)(e_1 \otimes e_1)^{\mathrm{T}}$ factors into polynomials of degree ≤ 3.

5. CONCLUSION

Either Eq. (3.4) or (3.9) indicate that the underlying variable, y_t, of univariate GARCH process (2.1) has a 4th moment iff $0 < \theta b_{11} < 1$ exists. Non-Gaussian disturbances without 4th moments, in particular, stable Paretian disturbances, have been considered (McCulloch, 1997; Rachev & Mittnik, 2000). Whether or not a disturbance has a 4th moment, in practice we would like to know whether an estimated GARCH process preserves existence of a disturbance's 4th moment. This would involve developing a statistical test of $\theta b_{11} < 1$ which accounts for the sampling variability of an estimated θb_{11}, based on the estimated GARCH parameters. Ideally, this would be an easily formed test, based on standard parameter estimates, along the lines of Dickey and Fuller's (1979) test of unit-root nonstationarity of an autoregressive linear time-series process.

Various papers in the literature claim to prove necessity and sufficiency of restrictions for 4th-moment existence in univariate GARCH processes, but most do not discuss any sort of nonnegativity (e.g., Hafner, 2003). Here, we see the crucial role of nonnegativity in the necessity proofs. Nonnegativity also figures crucially in existence proofs of stationarity of GARCH processes, but there nonnegativity occurs more simply, as is seen in Eq. (2.2). Bougerol and Picard (1992) and Ling and McAleer (2002) are exceptions in the literature for discussing nonnegativity in their respective existence proofs for 2nd and 4th moments of GARCH processes.

The state-space form of a GARCH process used here is simpler than the forms used by Bougerol and Picard (1992), Ling and McAleer (2002), and Mittnik et al. (2002), because the transition matrix F is nonstochastic. Thus, the proofs here are based on more elementary concepts and are simpler. For example, there is no need to consider a stochastic Lyapunov exponent and to verify its convergence, as in Bougerol and Picard (1992).

Finally, for any positive integer m, it would be interesting to generalize present restrictions to necessary and sufficient restrictions for existence of unique 2mth-moments of the underlying variable of a univariate GARCH process (cf., Ling & McAleer, 2002).

REFERENCES

Anderson, T. W. (1958). *The statistical analysis of time series*. New York, NY: Wiley.

Ansley, C. F., & Kohn, R. (1983). Exact likelihood of vector autoregressive moving-average process with missing or aggregated data. *Biometrika*, *70*, 275–278.

Baillie, R., & Bollerslev, T. (1989). The message in daily exchange rates: A conditional variance tale. *Journal of Business and Economic Statistics*, *7*, 297–305.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, *69*, 542–547.

Bougerol, P., & Picard, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics*, *52*, 115–127.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association*, *74*, 427–431.

Drost, F. C., & Klaassen, C. A. J. (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics*, *81*, 193–221.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*, 987–1008.

Gantmacher, F. R. (1959). *The theory of matrices*, Vol.2. Providence, RI: Chelsea Publishing.

Geweke, J. (1988). Exact inference in models with autoregressive conditional heteroskedasticity. In: W. A. Barnett, E. R. Berndt & H. White (Eds), *Dynamic Econometric Modelling* (pp. 73–104). Cambridge, U K: Cambridge University Press.

Giraitis, L., Leipus, R., & Surgailis, D. (2005). Recent advances in ARCH modeling. In: G. Teyssiere & A. P. Kirman (Eds), *Long memory in economics* (pp. 3–38). New York, NY: Springer-Verlag.

Giraitis, L., & Surgailis, D. (2002). ARCH-type Bilinear models with double-long memory. *Stochastic Processes and their Applications*, *100*, 275–300.

Hafner, C. (2003). Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics*, *1*, 26–54.

He, C., & Terasvirta, T. (1999). Fourth moment structure of the GARCH(p,q) process. *Econometric Theory*, *15*, 824–846.

Hsieh, D. (1989). Modelling heteroskedasticity in daily foreign-exchange rates. *Journal of Business and Economic Statistics*, *7*, 307–317.

Karanasos, M. (1999). The second moment and the autocovariance function of the squared errors of the GARCH model. *Journal of Econometrics*, *90*, 63–76.

Kazakevicius, V., Leipus, R., & Viano, M. C. (2004). Stability of random coefficient ARCH models and aggregation schemes. *Journal of Econometrics*, *120*, 139–158.

Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedasticity model. *Journal of Applied Probability*, *36*, 688–705.

Ling, S., & McAleer, M. (2002). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Econometric Theory*, *18*, 722–729.

Magnus, J. R., & Neudecker, H. (1988). *Matrix differential calculus with applications in statistics and econometrics*. New York, NY: Wiley.

Mandelbrot, B. (1963a). New methods in statistical economics. *Journal of Political Economy*, *71*, 421–440.

Mandelbrot, B. (1963b). The variation of certain speculative prices. *Journal of Business*, *36*, 394–419.

McCulloch, J. H. (1997). Measuring tail thickness in order to estimate the stable index α: A critique. *Journal of Business and Economic Statistics*, *15*, 74–81.

Milhoj, A. (1985). The moment structure of ARCH processes. *Scandinavian Journal of Statistics*, *12*, 281–292.

Mittnik, S., Paolella, M. S., & Rachev, S. (2002). Stationarity of stable power-GARCH processes. *Journal of Econometrics*, *106*, 97–107.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, *6*, 318–334.

Rachev, S., & Mittnik, S. (2000). *Stable paretian models in finance*. New York, NY: Wiley.

Wilkinson, J. H. (1965). *The algebraic eigenvalue problem*. Oxford, UK: Clarendon Press.

This page intentionally left blank