Essays on Autoregressive Conditional Heteroskedasticity

STOCKHOLM SCHOOL OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

EFI Mission
EFI, the Economic Research Institute at the Stockholm School of Economics, is a scientific institution that works independently of economic, political and sectional interests. It conducts theoretical and empirical research in the management and economic sciences, including selected related disciplines. The Institute encourages and assists in the publication and distribution of its research findings and is also involved in the doctoral education at the Stockholm School of Economics. At EFI, the researchers select their projects based on the need for theoretical or practical development of a research domain, on their methodological interests, and on the generality of a problem.

Research Organization
The research activities at the Institute are organized into 22 Research Centres.
Centre Directors are professors at the Stockholm School of Economics.

EFI Research Centre:	Centre Director:
Management and Organisation (A)	Sven-Erik Sjöstrand
Centre for Entrepreneurship and Business Creation (E)	Carin Holmquist
Public Management (F)	Nils Brunsson
Information Management (I)	Mats Lundeberg
Centre for People and Organization (PMO)	Andreas Werr (acting)
Centre for Innovation and Operations Management (T)	Christer Karlsson
Centre for Risk Research (CFR)	Lennart Sjöberg
Economic Psychology (P)	Guje Sevón
Centre for Consumer Marketing (CCM)	Magnus Söderlund
Centre for Information and Communication Research (CIC)	Per Andersson (acting)
Marketing, Distribution and Industrial Dynamics (D)	Björn Axelsson
Centre for Strategy and Competitiveness (CSC)	Örjan Sölvell
Centre for Business and Economic History (BEH)	Håkan Lindgren
Accounting and Managerial Finance (B)	Johnny Lind
Centre for Financial Analysis and Managerial Economics in Accounting (BFAC)	Kenth Skogsvik
Finance (FI)	Clas Bergström
Centre for Health Economics (CHE)	Bengt Jönsson
International Economics and Geography (IEG)	Mats Lundahl
Economics (S)	Paul Segerstrom
Economic Statistics (ES)	Anders Westlund
Law (RV)	Erik Nerep
Centre for Tax Law (SR)	Bertil Wiman

Chair of the Board: Professor Carin Holmquist
Director: Associate Professor Filip Wijkström

Address
EFI, Box 6501, SE–113 83 Stockholm, Sweden • Website: www.hhs.se/efi/
Telephone: +46(0)8-736 90 00 • Fax: +46(0)8-31 62 70 • E-mail efi@hhs.se

Essays on Autoregressive Conditional Heteroskedasticity

Annastiina Silvennoinen

STOCKHOLM SCHOOL
OF ECONOMICS
HANDELSHÖGSKOLAN I STOCKHOLM

EFI, The Economic Research Institute

Dissertation for the Degree of Doctor of Philosophy, Ph.D.
Stockholm School of Economics 2006

ISBN NR 91–7258–711–3

Keywords:
Financial econometrics; Time series; Asymmetry; Estimation; Multivariate GARCH; Nonlinearity; Programming; Unconditional skewness; Variable correlation GARCH model; Volatility model evaluation

Distributed by:
EFI, Stockholm School of Economics
Box 6501, SE–113 83 Stockholm, Sweden
www.hhs.se/efi/

Printed by:
Elanders Gotab, Stockholm 2006

Sometimes I thought to myself, 'Why?' and sometimes I thought, 'Wherefore?' and sometimes I thought, 'Inasmuch as which?' and sometimes I didn't quite know what I was thinking about.

And sometimes I stopped to think, and forgot to start again.

inspired by A. A. Milne

Contents

Acknowledgements

Over the past five years I have come to realize that, even though I do my best work during the dark hours of the night, there is something other than the brightly shining stars needed to guide the way for a dedicated researcher. Without a captain with enormous experience on the rough seas of econometric research, my journey through these PhD studies could often have steered off course. For his superb skill at the helm, I would like to express my deepest gratitude to my supervisor, Timo Teräsvirta.

I would also like to thank Changli He for his excellent guidance through the jungle of theoretical econometrics. His encouragement, insight, and help will always be remembered.

There has always been a source of support for me that so rarely received any recognition, never asked for it, and was all too often forgotten or taken for granted. You, my family, have always been there for me and have been more than anyone could ever hope for. Thank you for handing me the axe and preparing me for the fight. The trees will grow back.

Special thanks go to Mika Meitz, a colleague and a friend for all these years. I thank you for always being there, for sharing the moments, staying up late nights discussing the fundamentals, for agreeing and even more for disagreeing. I admire your ability to always find new ways and will wait for you to grow to appreciate the 'black gold'.

Thanks to Rickard Sandberg for his unquestioning support. I thank you for easing my way through the troubled times and for convincing me to continue, even in those moments when all else was failing.

Thanks also to Sophie von Pöffendahl for her unique way of pointing out the essentials and showing what I should be focusing on.

I am very grateful to have had the opportunity to visit the School of Finance and Economics at University of Technology Sydney. My very special thanks go to Tony Hall for his kind invitation to be acquainted with the academic circles down under. The visit was financed by Carl Silfvén's Foundation.

I also had the good fortune to visit the Center for Operations Research and Econometrics at Université catholique de Louvain. I would like to thank Luc Bauwens and Marie Curie Fellowship for making that visit possible.

This journey has taken me all around the world and has provided me with fresh new perspectives. I have seen places I never expected to see. I have learned so much from the work of other people and even more from those people themselves. There have been many fine academics on this road and many more wonderful people who have helped me on my way. I trust you know who you are and I thank you all for your help and kindness.

Last but not least I would like to acknowledge the financial support from Jan Wallander's and Tom Hedelius' Foundation.

Sydney, September 2006

Annastiina Silvennoinen

Introduction

This thesis consists of five research papers in the area of financial econometrics, and the focus is in topics related to discrete time modelling of financial volatility. The purpose of this Introduction is to give brief background to the research areas considered in the papers as well as short, intuitive descriptions of the specific topics in them. For more detailed and technical accounts of the contents of the papers, as well as for further literature references, the reader is referred to introductions in the individual papers.

Financial decision making is generally based on return and risk as well as on the tradeoff between them. Econometric analysis of risk, or volatility, is therefore of great importance in areas such as asset pricing, portfolio optimization, derivative pricing, hedging, and risk management. Volatility analysis aims at explaining the underlying causes of volatility. This can be done by building models that accurately describe the dynamics of the volatility and capture its typical features. In financial markets volatility models are used for forecasting the volatility of, say, stock prices or investment returns. Volatility can be seen as a response to news in the marketplace, some of which may be known to the agents, such as announcements of economic indicators or events, and have a relatively predictable effect on volatility. However, other factors are less well defined in both amplitude and frequency, and other means are called for in analysing or forecasting volatility caused by them.

There are several features present in financial data that have been recorded, conjectured, and confirmed as being properties of such data in a number of studies over the past decades. One of the earliest observations was that large movements in prices tend to be followed by movements of same magnitude, positive or negative. Similar behaviour is seen for small price changes. This implies that the volatility exhibits clustering and thereby persistent behaviour. For early observations, see Mandelbrot (1963) and Fama (1965), and for later studies for instance Chou (1988) and Schwert (1989). The volatility clustering is a phenomenon that appears periodically, that is, a period of high volatility is followed by a period of low volatility, and vice versa. Therefore, even though persistent, the present level of volatility should not affect a long-run forecast of volatility. Black (1976) and Christie (1982), among others, have noted that returns have a negative relationship to volatility. This so-called leverage effect means that negative news have a greater impact on the following day's volatility than positive news do. Finally, the marginal distribution of returns has heavier

tails than the normal distribution would imply, a fact generally referred to as excess kurtosis. These stylized facts mentioned above are aspects that a volatility model should accommodate.

The Autoregressive Conditional Heteroskedasticity (ARCH) model put forth by Engle (1982) provides an elegant way of parameterizing time-varying volatility and hence allowing for time-varying risk. Developed further by Bollerslev (1986) and Taylor (1986), the Generalized ARCH (GARCH) model parameterizes current volatility as a function of past squared returns and past volatilities. The GARCH model provides a way of forecasting volatility which is a central part of financial applications. This has prompted researchers to propose numerous specifications and extensions to GARCH models. A GARCH model is able to capture volatility clustering and asymmetric responses to news. It can even explain some of the excess kurtosis by being a mixture of normal distributions with varying volatilities. See Bollerslev, Engle, and Nelson (1994), Palm (1996), Shephard (1996), and Teräsvirta (2007), among others, for surveys of this literature.

In addition to the stylized facts mentioned above, there are also other documented features sometimes present in financial data. For instance, it is often found that the marginal distribution of returns is (negatively) skewed. However, these reports can be somewhat misleading because the usual measure for the skewness coefficient is seriously affected by outliers or extreme observations, see Kim and White (2004). Furthermore, the measurement and inference of skewness is based on the assumption of normality which usually is violated, see Peiró (2002, 2004). Nevertheless, because skewness is sometimes observed, there have been several attempts to model it through defining conditional skewness. This requires that one assumes a skewed conditional error distribution, an assumption that can be difficult to justify if the noise process should carry no information regarding the return process. The paper *'Parameterizing unconditional skewness in models for financial time series'* [1] tries to reveal the conditions under which the marginal distribution can be skewed while the noise process has symmetric distribution. The conditional mean process is often regarded as having no predictive power and therefore assumed to have no dynamic structure. Thus a common practice is merely to substract a constant mean from the return process. However, there is some evidence that the dynamics in the conditional mean, albeit weakly, may exhibit nonlinear behaviour, see Engle, Lilien, and Robins (1987) and Brännäs and de Gooijer (2004). Therefore we consider models where the shocks can have an effect both on the conditional mean and the conditional variance. We consider processes with a symmetric or asymmetric, a linear or nonlinear conditional mean to be able to isolate the different channels through which a skewed marginal distribution can be obtained. The contribution of the paper is to examine the effect of nonlinear or asymmetric structure in the conditional mean and variance on the third-moment structure, which, according to the best of our knowledge, has not yet been addressed in the existing literature. In general, we find that asymmetries or nonlinearities in the conditional mean are of greater importance than they are in the conditional standard deviation or variance when it comes to generating skewed marginal distributions. If

[1] This paper is joint work with Changli He and Timo Teräsvirta.

the conditional mean is symmetric and linear, then the unconditional skewness can only follow from the asymmetry of the conditional standard deviation or variance. However, in the case of no or a constant conditional mean, the marginal distribution is unavoidably symmetric.

It may be of interest to see how the past news affect not only the current volatility but also the magnitude of today's returns. We introduce a definition of the Shock Impact Curve which generalizes the News Impact Curve of Engle and Ng (1993). It describes the impact of a shock on the mean squared error of the return. It combines the effects of the conditional mean and the conditional standard deviation or variance on the squared returns.

The discussion so far has been focused on univariate properties of financial data. Many financial decisions, however, do not depend on the behaviour of a single asset only. Comovement between assets returns in, say, a portfolio is an important factor when it comes to practical finance. Asset allocation and hedging are examples in which a major role is played by the covariance of the assets in a portfolio, see for instance Bollerslev, Engle, and Wooldridge (1988) and Hansson and Hördahl (1998). Another area of finance that benefits from modelling volatility and correlation transmission and spillover effects are the studies of contagion, see for example Tse and Tsui (2002) and Bae, Karolyi, and Stulz (2003). Soon after the univariate ARCH model was introduced and its usefulness realized, the researchers started to investigate the possibility of extending the idea of conditional heteroskedasticity into multivariate setting. The earliest models, proposed by Engle, Granger, and Kraft (1984) and Bollerslev, Engle, and Wooldridge (1988), were merely attempts to imitate the univariate ARCH (GARCH) model. It soon became obvious that Multivariate GARCH (MGARCH) models posed problems that either were not present, or were easily tackled, in the univariate models. One problem arising is the so-called 'curse of dimensionality'. As the number of parameters increases rapidly with the dimension of the model, the parameters become difficult to interpret and the model quite rapidly infeasible to estimate. Opting for a more parsimonious model, however, can restrict the dynamics of the conditional covariances and thereby limit the usefulness of the model. Another issue is that the conditional covariance matrix of the returns has to be positive definite. Ensuring this requirement can be difficult, whether it is imposed through parameter restrictions or model structure. While trying to overcome these difficulties, there have been several proposals for MGARCH models that use parametric, semi- and non-parametric approaches to modelling conditional covariances. The paper *'Multivariate GARCH models'* [2] reviews development of multivariate modelling; see also Bauwens, Laurent, and Rombouts (2006).

An area that has recently received much attention in the MGARCH literature is the modelling of the dynamics in the conditional correlations between the assets. The dynamic correlation models are appealing because they are easy to interpret. In these models, the conditional covariance matrix is decomposed into the standard deviations of each of the univariate return series, and to the correlations among them.

[2] This paper is joint work with Timo Teräsvirta.

The conditional variances of each of the series are modelled using univariate GARCH specifications. By now there exists a vast literature on univariate GARCH models, their properties, misspecification tests, and so on, that will make choosing a suitable model for each of the return series an easier task. What is left is to model the correlations between the series. A further advantage, from a computational point of view, is that the correlation dynamics are often easily separable from the volatility dynamics, which makes the estimation relatively easy.

The first conditional correlation model is the one of Bollerslev (1990), in which the correlations are constant over time. This assumption may, however, be too simplistic in practical applications. Tests for this restriction have been proposed for instance by Tse (2000) and Bera and Kim (2002), and several studies have found evidence that asset returns, in particular, are time-varying. Furthermore, correlation dynamics have been reported to have a relation to the state of the markets. Evidence that correlations increase during financial crises has been reported, for example, by King and Wadhwani (1990) and Lin, Engle, and Ito (1994). If correlations have a relation to market turbulence, there may exist one or several exogenous variables that signal correlation movements.

The paper *'Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations'* [3] introduces a new dynamic correlation MGARCH model, the Smooth Transition Conditional Correlation (STCC) GARCH model, in which conditional correlations are allowed to vary smoothly between two extreme states of constant correlations according to a transition variable. The transition variable is chosen by the modeller and depends on the process to be modelled. Possible choices include functions of past values of one or more of the series, time (as in Berben and Jansen (2005)), or exogenous variables such as business cycle indicators or volatility indices. The STCC–GARCH model provides a framework for testing for the relevance of a transition variable to the correlation dynamics. When testing the constant correlation hypothesis, the alternative hypothesis depends on the chosen transition variable. Therefore, rejecting constancy reveals that the transition variable is in fact informative about the dynamics in the correlations. Non-rejection, on the other hand, indicates that the variable contains no information about the time-varying correlations. However, this does not imply that the correlations would be constant. In the studies by Kroner and Ng (1998) and Longin and Solnik (2001) the correlations are found to show asymmetric behaviour in that they react more strongly to negative shocks than to positive ones. The STCC framework allows one to examine hypotheses of this kind in a flexible manner: first by testing for the relevance of a transition variable that is in accordance with the hypothesis, and then, in case of a rejection, by estimating the STCC–GARCH model to find out the direction of changes in correlations indicated by the transition variable. As the dimension of the model increases, the number of parameters increases quite rapidly. It may be that some of the correlations do not exhibit time-variation according to the chosen transition variable, and therefore it is of interest to test whether some of the correlation parameters are in fact constant while others are time-varying. The STCC–GARCH

[3] This paper is joint work with Timo Teräsvirta.

framework offers a way of finding such restrictions by testing partial constancy of the correlation matrix.

In the STCC-GARCH model, the dynamic correlations are allowed to vary according to one transition variable. However, it may be argued that the correlations can be influenced by several variables, such as ones reflecting the current turbulence in the market, whether the market is going up or down, or simply calendar time. The paper *'Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model'* [4] introduces the Double Smooth Transition Conditional Correlation (DSTCC) GARCH model, which extends the STCC–GARCH model in that it allows for two simultaneous transitions and thereby contains two transition variables. A special case of the DSTCC–GARCH model is the Time-Varying Smooth Transition Conditional Correlation (TVSTCC) GARCH model in which one of the transition variables is calendar time. This implies that the two states of correlations between which the correlations vary in the STCC–GARCH framework are allowed to smoothly shift to other levels over time. It may be of interest to allow the constant states of correlation to shift over time, since when MGARCH models are applied, the time series in question necessarily cover long periods of time. The DSTCC–GARCH framework offers a testing environment similar to the one within the STCC–GARCH model. Due to the relatively large number of parameters in a full DSTCC–GARCH model, the partial constancy tests become relevant tools in order to find a parsimonious specification.

An important aspect of any econometric model is their relative ease or difficulty of estimation. In GARCH models, the estimation of univariate models is generally quite straightforward. Most of the existing software programs have built-in estimation procedures that cover a wide variety of univariate GARCH models and provide several choices for algorithms to be used in the numerical optimization problem. However, the optimization problem is highly nonlinear even in the univariate setting, which may cause problems. This fact is exacerbated when considering multivariate GARCH models. Reviews of performance of MGARCH estimation routines have as yet focused on solutions that rely mostly on built-in procedures and require minimal programming skills. The so-called 'black box' approach means that the modeller simply presses a button and hopes that the optimizer knows what it is doing and, more importantly, handles possible failures or problematic situtations in a way that is in line with the model specification. This approach, however, is known most likely to result in unreliable model estimates. Therefore, in estimating an MGARCH model, it is of great importance to possess a thorough understanding, both from the theoretical and the computational point of view, of the properties of the model, the estimation algorithm, and their interaction. The paper *'Numerical aspects of the estimation of multivariate GARCH models'* points out several issues that are of importance when estimating an MGARCH, or any other highly nonlinear, model. Understanding not only the model to be estimated and the numerical procedures used, but also how the program reaches convergence is a definite prerequisite to successful and reliable estimation of the parameters of the model. In the case of highly nonlinear models, 'trivial' approaches to

[4] This paper is joint work with Timo Teräsvirta.

model estimation via maximizing a likelihood function rarely work, and it becomes necessary to modify the estimation algorithm to take into account the particular features of the model being estimated. Typical issues are, for example, multiple local maxima and boundary conditions. The paper also looks at the computational resources required to reliably run an estimation and draws attention to the practical considerations of software choice, efficient programming, and programming skills.

References

Bae, K.-H., G. A. Karolyi, and R. M. Stulz (2003): "A new approach to measuring financial contagion," *The Review of Financial Studies*, 16, 717–763.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79–109.

Bera, A. K., and S. Kim (2002): "Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns," *Journal of Empirical Finance*, 9, 171–195.

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

Black, F. (1976): "Studies of stock market volatility changes," Proc. 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section.

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

——— (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., R. F. Engle, and D. B. Nelson (1994): "ARCH models," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2959–3038. Elsevier Science, Amsterdam.

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

Brännäs, K., and J. G. de Gooijer (2004): "Asymmetries in conditional mean and variance: Modelling stock returns by asMA–asQGARCH," *Journal of Forecasting*, 23, 155–171.

Chou, R. Y. (1988): "Volatility persistence and stock valuations: some empirical evidence using GARCH," *Journal of Applied Econometrics*, 3, 279–294.

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

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1006.

Engle, R. F., C. W. J. Granger, and D. Kraft (1984): "Combining competing forecasts of inflation using a bivariate ARCH model," *Journal of Economic Dynamics and Control*, 8, 151–165.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987): "Estimating time varying risk premia in the term structure: the ARCH-M model," *Econometrica*, 55, 391–407.

Engle, R. F., and V. K. Ng (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–78.

Fama, E. F. (1965): "The behaviour of stock market prices," *Journal of Business*, 38, 34–105.

Hansson, B., and P. Hördahl (1998): "Testing the conditional CAPM using multivariate GARCH–M," *Applied Financial Economics*, 8, 377–388.

Kim, T.-H., and H. White (2004): "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, 1, 56–73.

King, M. A., and S. Wadhwani (1990): "Transmission of volatility between stock markets," *The Review of Financial Studies*, 3, 5–33.

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

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

Longin, F., and B. Solnik (2001): "Extreme correlation of international equity markets," *Journal of Finance*, 56, 649–676.

Mandelbrot, B. (1963): "The variation of certain speculative prices," *Journal of Business*, 36, 394–419.

Palm, F. C. (1996): "GARCH models of volatility," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 14, pp. 209–240. Amsterdam: Elsevier Sciences.

Peiró, A. (2002): "Skewness in individual stocks at different investment horizons," *Quantitative Finance*, 2, 139–146.

——— (2004): "Asymmetries and tails in stock index returns: are their distributions really asymmetric?," *Quantitative Finance*, 4, 37–44.

Schwert, G. W. (1989): "Why does stock market volatility change over time?," *Journal of Finance*, 44, 1115–1153.

Shephard, N. (1996): "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, pp. 1–67. Chapman and Hall, London.

Taylor, S. J. (1986): *Modelling Financial Time Series*. Wiley, Chichester.

Teräsvirta, T. (2007): "Univariate GARCH models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch. Springer, New York.

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

Tse, Y. K., and K. C. Tsui (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

Parameterizing unconditional skewness in models for financial time series

Parameterizing unconditional skewness in models for financial time series

Abstract

In this paper we consider the third-moment structure of a class of nonlinear time series models. It is often argued that the marginal distribution of financial time series such as returns is skewed. Therefore it is of importance to know what properties a model should possess if it is to accommodate unconditional skewness. We consider modelling the unconditional mean and variance using models that respond nonlinearly or asymmetrically to shocks. We investigate the implications of these models on the third-moment structure of the marginal distribution as well as conditions under which the unconditional distribution exhibits skewness and nonzero third-order autocovariance structure. In this respect, an asymmetric or nonlinear specification of the conditional mean is found to be of greater importance than the properties of the conditional variance. Several examples are discussed and, whenever possible, explicit analytical expressions provided for all third-order moments and cross-moments. Finally, we introduce a new tool, the shock impact curve, for investigating the impact of shocks on the conditional mean squared error of return series.

This paper is joint work with Timo Teräsvirta and Changli He.

We thank Tony Hall, Markku Lanne, Mika Meitz, Peter Phillips, and Pentti Saikkonen for helpful comments and suggestions. Participants of the conferences and workshops 'Econometrics and Computational Economics', Helsinki, November 2004, 14th meeting of the New Zealand Econometric Study Group, Christchurch, March 2005, RUESG Workshop on Financial Econometrics, Helsinki, August 2005, International Conference on Finance, Copenhagen, September 2005, and the European Meeting of the Econometric Society, Vienna, August 2006, also provided useful remarks. The responsibility for any errors and shortcomings in this paper remains ours.

1 Introduction

Financial series such as high-frequency asset returns have little forecastable structure in the mean. For this reason, and because volatility is used as a measure of risk, forecasting volatility and thus modelling the conditional variance has been the main concern of practitioners. The most popular family of volatility models, the GARCH family, see Bollerslev (1986) for the standard GARCH model, is used to characterize two important stylized facts of return series: fat tails of the marginal distribution of returns and volatility clustering, that is, higher-order dependence observed in the series.

Another feature of these series that has attracted attention is an asymmetric response of volatility to shocks. GARCH models that can take this into account include the Threshold GARCH (Zakoïan, 1994), the GJR–GARCH (Glosten, Jagannathan, and Runkle, 1993), and the Smooth Transition GARCH (Hagerud, 1997; González-Rivera, 1998) model. Pagan and Schwert (1990) and Engle and Ng (1993) have suggested a practical way of describing this response by the so-called News Impact Curve (NIC).

In addition, it has been observed that the marginal distribution of returns is sometimes skewed. Harvey and Siddique (1999), Chen, Hong, and Stein (2001), and Engle and Patton (2001), to mention a few contributions, report evidence for financial time series with asymmetric distributions. However, as pointed out by Peiro ((2002, 2004)), one should not go as far as stating the skewness of marginal distributions of returns as a stylized fact, nor rely solely on its traditional measurement under normality. One should investigate possible asymmetry of the distribution using not only traditional tests but distribution-free measurements as well, see also Kim and White (2004). Attempts to model this skewness through defining the concept of conditional skewness have been made, see for instance Harvey and Siddique (1999), Lambert and Laurent (2002) and references therein, Brännäs and Nordman (2003a), Brännäs and Nordman (2003b), and Harris, Küçüközmen, and Yilmaz (2004). This requires giving up a standard assumption in econometric work, namely, that noise sent through a parametric filter to generate the output has a symmetric distribution around zero. (Of course, modelling positive-valued series constitutes an exception.) Furthermore, in some cases, see Hansen (1994), one even gives up the assumption, otherwise invariably made in the context of GARCH processes, that the errors of the process are independent.

It may be conceptually difficult to understand why the noise that in principle should contain no information about the properties of the process should have a non-symmetric distribution and thus be informative about the output. For this reason, in this paper we make an effort to find out under which conditions the marginal distribution of returns can be skewed while the noise has a symmetric distribution. It turns out that for this purpose we have to study processes with a nonconstant conditional mean. In so doing, we shall be interested in the case where a shock can have a nonzero effect on both the conditional mean and conditional variance. This leads us to consider processes with a symmetric or asymmetric, and linear or nonlinear conditional mean. There is empirical evidence of some return series having

an asymmetric conditional mean; see Brännäs and de Gooijer (2004). An example of a nonlinear conditional mean is the well-known GARCH-in-mean (GARCH–M) process introduced by Engle, Lilien, and Robins (1987). Lanne and Saikkonen (2005) recently considered a GARCH–M model with an asymmetric error distribution. These asymmetries and nonlinearities are likely to affect the marginal distribution of the series to be modelled but, to the best of our knowledge, their effect on the third-moment structure of the process has not been investigated.

The starting-point of the present paper is a model whose first two conditional moments are parametric and the error distribution is symmetric around zero. The asymmetric moving average (asMA) model by Wecker (1981) serves as an example of such a model. It nests a linear moving average model, and for this reason the effect of the asymmetry in the conditional mean on the third-moment structure of the process can be easily investigated by imposing appropriate parameter restrictions on the model. Examining the role of the conditional mean as a whole in this framework is quite straightforward. Our other example will be the GARCH–M model. It is well known (Hong, 1991) that the GARCH–M model implies autocorrelated returns, but it is probably less well known that introducing a function of the conditional variance in the conditional mean makes the marginal distribution of the observations skewed.

For the purpose of deriving analytical expressions for unconditional third-order moments, parameterizing the conditional standard deviation is preferable to parameterizing the conditional variance. In the latter case, the definitions of moments would involve expectations that do not have analytic expressions. For this reason, we focus on the threshold GARCH (TGARCH) model that in turn nests the absolute-value GARCH (AVGARCH) model of Taylor (1986) and Schwert (1989). The TGARCH model has an asymmetric response to shocks, whereas the same response in the AVGARCH model is symmetric as it is in the standard GARCH model of Bollerslev (1986). General conclusions drawn from these two models of the conditional standard deviation are applicable to other GARCH models as well.

Recently, Brännäs and de Gooijer (2004) proposed a model that introduces asymmetry both in the conditional mean and the conditional variance. The variance is an extension of the QGARCH model of Sentana (1995). The authors considered the first and second moments of their asMA–asGARCH model but did not investigate the third-moment structure of their model. Because they parameterize the conditional variance, not the conditional standard deviation, finding analytical expressions for the third-order moments appears difficult. In fact, it seems that even lower-order moments may not have analytical expressions readily available. As already suggested, general conclusions from our models will be applicable to the asMA–asGARCH model.

It turns out that there is a rather large set of asMA–TGARCH parameter values such that the marginal distribution of the observations will be skewed. Not all of them are relevant in the sense that they would correspond to situations experienced in practice. For example, we may not expect the volatility to respond more strongly to positive than it does to negative shocks of the same size. In order to study the relevance of the parameter combinations in question we generalize the News Impact Curve (NIC) of Engle and Ng (1993) in order to account for the structure in the conditional mean. For this purpose we define a new concept, the Shock Impact Curve

(SIC), that describes the impact of a shock on the conditional mean squared error of the series, and apply it for our purposes.

The paper is organized as follows. The general model is introduced in Section 2 and its moment structure up to the third moments derived in Section 3. Special cases are presented in Section 4. The shock impact curve is defined and applied in Section 5. Conclusions from this study can be found in Section 6. Technical derivations and expressions of the moments are contained in the Appendix.

2 The model family

Let y_t be generated by

$$y_t = \mu_t + \varepsilon_t, \tag{1}$$

$$\varepsilon_t = z_t h_t \tag{2}$$

where μ_t is the conditional mean of y_t given $\mathcal{F}_{t-1}$ (the sigma-field generated by the available information until time $t-1$), h_t^2 is the conditional variance of y_t given $\mathcal{F}_{t-1}$, and $\{z_t\} \sim iid(0,1)$ with a distribution function that is symmetric around zero. The processes μ_t and h_t are measurable with respect to $\mathcal{F}_{t-1}$.

We consider a variety of examples from two classes of models for the conditional mean and especially focus on the ability of these models to exhibit asymmetric or nonlinear behaviour. For simplicity, we focus on first-order models which are empirically often found to be adequate. The equation

$$\mu_t = \phi\varepsilon_{t-1} + \phi^+(\varepsilon_{t-1}^+ - E\varepsilon_t^+) \tag{3}$$

where $\varepsilon_t^+ = \max(0, \varepsilon_t)$, defines the first-order asymmetric moving average (asMA) process of Wecker (1981). For $\phi^+ \neq 0$ the model is asymmetric and linear in its response to shocks. Note that model (3) nests an MA process which is symmetric and linear. If μ_t is a function of h_t such that

$$\mu_t = \phi(h_t^\delta - Eh_t^\delta), \qquad \delta = 1 \text{ or } 2 \tag{4}$$

we have the GARCH-in-mean (GARCH–M) model of Engle, Lilien, and Robins (1987). In this case the model for the conditional mean is nonlinear and the degree of asymmetry is controlled by the asymmetry of the GARCH process.

The error process $\{\varepsilon_t\}$ of (1) is assumed to be a conditionally heteroskedastic white noise sequence with

$$h_t^d = \omega + c_{t-1}h_{t-1}^d, \qquad d = 1 \text{ or } 2 \tag{5}$$

where $c_t = c_t(z_t)$ is a well-defined function and $h_t^d > 0$ for all t. To ensure this, suitable parameter restrictions must be imposed. The moment properties of the family of GARCH models defined by (5) are investigated in He and Teräsvirta (1999). It

nests many of the models in the family of GARCH models of Hentschel (1995). For instance, setting $d = 2$ in (5) and

$$c_t = \alpha z_t^2 + \beta$$

yields the standard GARCH model of Bollerslev (1986). Setting $d = 1$ and

$$c_t = \alpha \left| z_t \right| + \beta + \alpha^* z_t \tag{6}$$

equation (5) defines the threshold GARCH (TGARCH) model of Zakoïan (1994). By setting $\alpha^* = 0$, the model collapses into the AVGARCH model of Taylor (1986) and Schwert (1989). Furthermore, the GJR–GARCH model of Glosten, Jagannathan, and Runkle (1993) and the nonlinear GARCH (NLGARCH) model of Engle (1990) are nested in (5). Note that any GARCH model defined by equation (5) is symmetric in its response to shocks if and only if $c_t(z_t)$ in (5) is an even function of z_t. The following theorem states conditions under which the process $\{\varepsilon_t\}$ in (2) and (5) is strictly and *md*-order stationary.

Theorem 1 *If $E|z_t^{\lambda d}| < \infty$ and $Ec_t^{\lambda} < 1$ for some $\lambda \in (0, 1]$, then there exists a unique λd-order stationary solution to (2) and (5). The solution is strictly stationary and ergodic. If $E|z_t^{dm}| < \infty$, then the necessary and sufficient condition for the existence of the mdth moment of the solution $\{\varepsilon_t\}$ in (2) is $Ec_t^m < 1$ where m is a positive integer.*

For a proof, see Theorems 2.1 and 2.2 in Ling and McAleer (2002).

3 Moments

We begin by considering the moment structure of the general model (1) and (2). Assume that $\{y_t\}$ is strictly stationary with finite third-order moments and set $\gamma_i = E(y_t - Ey_t)^i$ and $\gamma_{ij}(k) = Cov(y_t^i, y_{t-k}^j)$, $i, j \geq 1$. The unconditional mean and variance are $Ey_t = E\mu_t$ and $\gamma_2 = Var\mu_t + E\varepsilon_t^2$, respectively. The autocovariances are $\gamma_{11}(k) = Cov(\mu_t, \mu_{t-k}) + E\mu_t\varepsilon_{t-k}$ where $k \geq 1$. Assuming μ_t is an asMA process in (3) and $\phi^+ \neq 0$ renders the autocovariances nonzero for $k \geq 1$, see Lemma 1 in the Appendix. The same holds if μ_t is a function of h_t^{δ}, $\delta = 1$ or 2, see Lemma 3.

The third moment and third-order cross-moments of y_t are given by

$$\begin{aligned}
\gamma_3 &= E(\mu_t - E\mu_t)^3 + 3Cov(\mu_t, \varepsilon_t^2) \\
\gamma_{21}(k) &= Cov(\mu_t^2, \mu_{t-k}) + Cov(\mu_{t-k}, \varepsilon_t^2) + E\mu_t^2\varepsilon_{t-k} + E\varepsilon_t^2\varepsilon_{t-k}, \qquad k \geq 1 \\
\gamma_{12}(k) &= Cov(\mu_t, \mu_{t-k}^2) + Cov(\mu_t, \varepsilon_{t-k}^2) + 2E\mu_t\mu_{t-k}\varepsilon_{t-k}, \qquad k \geq 1.
\end{aligned}$$

Define the unconditional skewness of y_t as $\kappa_3 = \gamma_3/(\gamma_2)^{3/2}$. The following proposition gives general conditions that yield zero skewness.

Proposition 1 *Consider the model in (1) and (2) that is third-order stationary. The conditions $E(\mu_t - E\mu_t)^3 = 0$ and $Cov(\mu_t, \varepsilon_t^2) = 0$ are sufficient for $\kappa_3 = 0$.*

When $\mu_t \equiv \mu$ (constant) in (1), the conditions in Proposition 1 are satisfied and thus $\gamma_3 = 0$. In this case, the only nonzero cross-moments are $\gamma_{21}(k) = E\varepsilon_t^2\varepsilon_{t-k}$, $k \geq 1$. Therefore, only assuming that the conditional second moment is time-varying does not imply nonzero unconditional skewness. A time-varying conditional mean is required for that.[1]

4 Examples

In this section we consider two specifications of the conditional mean in detail. The first one is the first-order asymmetric moving average model of Wecker (1981). The second one (GARCH–M) introduces the conditional standard deviation or variance into the conditional mean. We choose the TGARCH model for the error process $\{\varepsilon_t\}$. This choice is dictated by our goal which is to obtain analytical expressions for all unconditional third-order moments and cross-moments. Such expressions give an idea of how asymmetries and nonlinearities in conditional first and second moments contribute to the unconditional third moments. Since the TGARCH model is asymmetric in its response to shocks and nests the symmetric AVGARCH model, it is possible to isolate the effect of this asymmetry on the unconditional skewness. Analytical expressions for the moments can be found in Lemma 2 and the subsequent Corollaries and in Lemma 4 of the Appendix. These expressions are rather involved but yield quite straightforward conclusions.

Other GARCH models are likely to be similar to the TGARCH model in this respect but because most of them lack analytical expressions for third-order moments, one has to rely on simulations to obtain numerical values for them.[2] Whenever possible, we try to take examples of models such as the symmetric GARCH model of Bollerslev (1986), the asymmetric QGARCH one of Sentana (1995), or the GJR–GARCH model of Glosten, Jagannathan, and Runkle (1993). In fact, the results in the Appendix apply to the general family of GARCH models (5). However, fully explicit expressions for third moments are provided only for the TGARCH or the AVGARCH model and not for models for which $d = 2$ in (5). Note that the QGARCH model is not a member of the family defined by equation (5).

We begin by considering the complete asMA–TGARCH model and the effect of restricting the conditional standard deviation to be a symmetric AVGARCH model. Subsequently, we consider the third-moment structure of the model when the conditional mean is simplified to only contain either the asymmetric component ($\phi = 0$), the symmetric MA component ($\phi^+ = 0$), or neither ($\phi = \phi^+ = 0$). In all these cases we consider both the TGARCH and the AVGARCH specifications for the conditional

[1] One motivation for extending standard symmetric GARCH models to include the leverage effect has been to create asymmetric unconditional densities, see e.g. Lambert and Laurent (2002). Engle and Patton (2001) also write that 'the asymmetric structure of volatility generates skewed distributions of forecast prices'.

[2] The possibility of using quantile measures as in Kim and White (2004) for unconditional skewness is yet to be explored.

second moment. As a final example, the conditional mean is defined to be a function of the conditional second moment, which leads us to GARCH–M models.

4.1 First-order asMA model with TGARCH or AVGARCH conditional standard deviation

The third-moment structure of the first-order asMA–TGARCH model is characterized by Lemmas 2 and 5. It is apparent from the rather involved expressions that this model accommodates a rich variety of third-order moment structures.

For an asMA process combined with a model for conditional heteroskedasticity the autocovariances $\gamma_{11}(k)$ are nonzero for all lags $k \geq 0$, as also pointed out in Brännäs and de Gooijer (2004). This is the case for both symmetric and asymmetric GARCH processes. After a peak at the first lag, the values of autocorrelation are very low and the decay rate is slow. Brännäs and de Gooijer (2004) claim that it may be empirically possible to discriminate between an asMA(q) model with a constant conditional variance and one with a GARCH process for the conditional variance because of the difference in autocovariances. The former process has nonzero autocovariances up until lag q, and zero thereafter, whereas the latter has nonzero autocovariances for all lags. However, at least for the first-order asMA–GARCH model the autocovariances are very close to zero after first lag, which complicates distinguishing between them this way.

Since the expressions for the third-order moments are quite involved we illustrate the situation numerically. The following figures are produced using the standardized Gaussian error distribution. Figure 1 shows the amount of unconditional skewness that can be obtained from an invertible asMA–TGARCH process for certain parameter values of the conditional mean. Invertibility of the asMA process has to be checked using simulations, see e.g. Brännäs and de Gooijer (1995). Each curve represents the level of unconditional skewness for a fixed value of ϕ^+ and a suitable range of values for ϕ. When $|\phi|$ is large, the invertibility condition restricts the asymmetry parameter ϕ^+ and thus limits the achievable amount of skewness.

When the conditional standard deviation is restricted to follow the AVGARCH model, some of the expressions in Lemma 2 simplify, but not substantially. The resulting moment structure is given in Corollary 1. In this case the cross-moments $E\varepsilon_t^2\varepsilon_{t-k}$, $E\varepsilon_{t-k}\varepsilon_t^+$, $E\varepsilon_{t-k}\varepsilon_t^{+2}$, $E\varepsilon_{t-k}\varepsilon_{t-k-1}\varepsilon_t^+$, and $E\varepsilon_{t-k}\varepsilon_t^+\varepsilon_{t-k-1}^+$, $k \geq 1$, equal zero, which can be seen from the expressions in Lemma 5 and Corollary 5. Thus the third-moment structure is still rich as long as the conditional mean is defined by an asMA model. In particular, all third-order moments are nonzero, whether or not the conditional standard deviation exhibits asymmetry. However, there is a reduction on the amount of skewness when the conditional standard deviation is no longer asymmetric, which is seen by comparing the top and bottom panels in Figure 1. What seems to have an even larger effect is the increase in the persistence of the GARCH process, which increases the skewness of the marginal distribution whenever $\phi^+ \neq 0$.

Next consider the case where $\phi = 0$. It can be seen from the expressions in Corollary 2 that the third-moment structure is still rather rich and complex. It is, however, considerably simpler than in the case of $\phi \neq 0$. When the errors are restricted to follow a symmetric AVGARCH model, the expressions simplify somewhat, see Corollary 3, but again not very much. Expressions in Lemma 5 and Corollary 5 show that $E\varepsilon_t^2\varepsilon_{t-k} = E\varepsilon_{t-k}\varepsilon_t^+ = E\varepsilon_{t-k}\varepsilon_t^{+2} = E\varepsilon_{t-k}\varepsilon_t^+\varepsilon_{t-k-1}^+ = 0$, $k \geq 1$. Thus, regardless of the conditional standard deviation, an asymmetric conditional mean leads to a skewed marginal distribution for y_t and nonzero third-order cross-moments. An obvious conclusion is that asymmetry of the conditional mean plays a very influential role in determining both the sign and the amount of skewness in this distribution.

4.2 First-order MA model with TGARCH or AVGARCH conditional standard deviation

A rather simple third-moment structure follows when $\phi^+ = 0$ in (3) while the conditional standard deviation follows the TGARCH model. This is evident from the results in Corollary 4. In this case we still have $Ey_t^3 \neq 0$. An interesting feature is that $Ey_t^2 y_{t-k} \neq 0$, $k \geq 1$, whereas $Ey_t y_{t-k}^2 = 0$ for $k > 1$. It should also be noted that the only nonzero cross-moment of ε_t is now $E\varepsilon_t^2\varepsilon_{t-k}$. In fact, $\kappa_3 = 0$ if and only if $E\varepsilon_t^2\varepsilon_{t-1} = 0$, Furthermore, also assuming $\phi = 0$ in (3), i.e. having $\mu_t = 0$, forces the unconditional skewness to zero regardless of the asymmetry in the conditional second moment. In this case the only nonzero cross-moments are $Ey_t^2 y_{t-k} = E\varepsilon_t^2\varepsilon_{t-k}$, $k \geq 1$. These results may be useful in specifying asMA–TGARCH models.

The thick curve in Figure 1 represents the skewness as a function of ϕ for $\phi^+ = 0$. In the top panels it intersects the x-axis at $\phi = 0$ in accordance with the results mentioned after Proposition 1.

As an aside consider a model whose conditional mean specification is a first-order AR process: $\mu_t = \phi y_{t-1}$ in (1). In this case

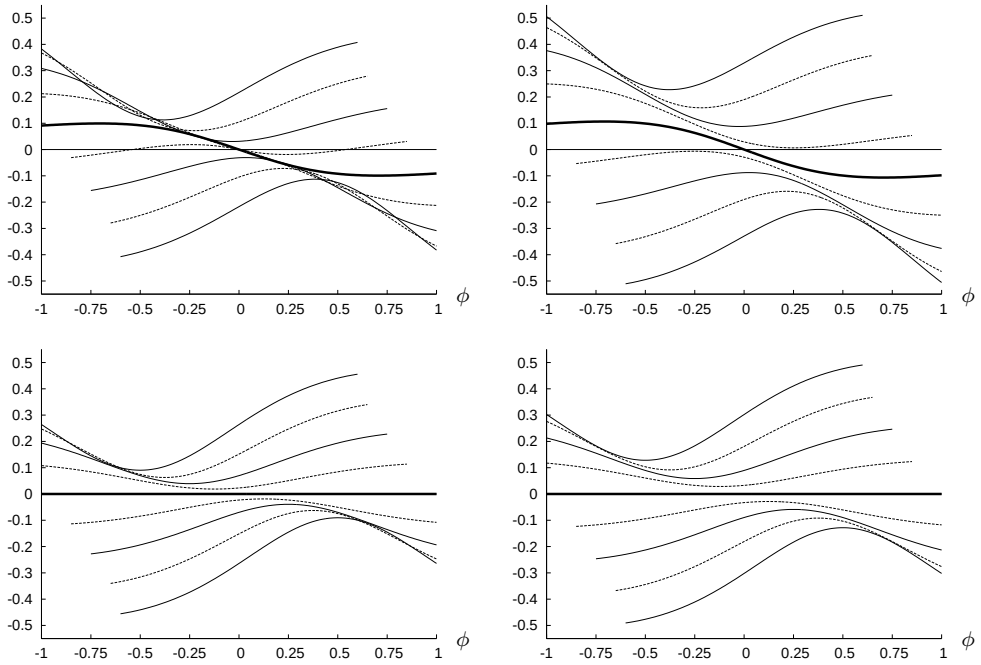
$$\kappa_3 \quad = \quad \frac{3\sum_{k=1}^{\infty}\phi^i E\varepsilon_t^2\varepsilon_{t-k}/(1-\phi^3)}{(Eh_t^2/(1-\phi^2))^{3/2}}.$$

Clearly $\kappa_3 = 0$ if and only if $E\varepsilon_t^2\varepsilon_{t-k} = 0$ for all $k \geq 1$. The unconditional skewness emerging from this model is very similar to that of the MA–TGARCH model already discussed. In Figure 2 the unconditional skewness is plotted as a function of the mean parameter ϕ for a range of values for β and keeping the other TGARCH parameter values fixed. It can be concluded that the amount of skewness obtained from a model with a linear and symmetric conditional mean is not large and the effect of increasing persistence in the GARCH process on skewness is negligible.

We now turn to the analytic form of the cross-moment $E\varepsilon_t^2\varepsilon_{t-k}$ in Corollary 4. When the conditional standard deviation is defined as a TGARCH process, it follows that

$$E\varepsilon_t^2\varepsilon_{t-k} = 2\omega\alpha^* Eh_t^2 \sum_{j=0}^{k-1}(Ec_t)^{k-1-j}(Ec_t^2)^j + 2\alpha^*(Ec_t^2)^{k-1}Ec_t|z_t|^2 Eh_t^3, \qquad k \geq 1,$$

Figure 1: asMA–TGARCH: Unconditional skewness of y_t as a function of ϕ for the following values of ϕ^+ and the TGARCH parameters: lines: $\phi^+ = 1.0, 0.75, \ldots, -1.0$ (top to bottom), thick line corresponds to $\phi^+ = 0$; TGARCH parameters: $\omega = 0.005$, $\alpha = 0.05$, $\beta = 0.90$ (left-hand panels), $\beta = 0.94$ (right-hand panels), $\alpha^* = -0.04$ (top panels), and $\alpha^* = 0$ (bottom panels).



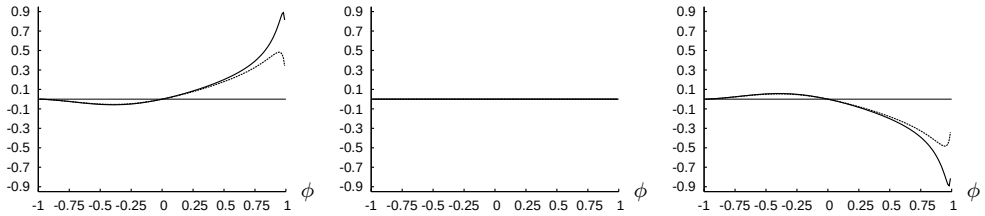
where the expressions for the moments of c_t and h_t are given in Lemma 5. Hence $E\varepsilon_t^2\varepsilon_{t-k} \neq 0$ for $\alpha^* \neq 0$. Assuming $\alpha^* = 0$ yields $E\varepsilon_t^2\varepsilon_{t-k} = 0$, $k \geq 1$, which implies that the third moment and all the third-order cross-moments in Corollary 4 are zero. In fact, any parameterization of h_t or h_t^2 that has the property $E\varepsilon_t^2\varepsilon_{t-k} = 0$ for $k \geq 1$ gives the same result. As stated in Corollary 5, all symmetric GARCH models belonging to the family (5) have that property. The standard GARCH model of Bollerslev (1986) is an example of such a case. As further examples, consider the nonlinear models where the errors are governed by a first-order QGARCH process $h_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2 + \alpha^*\varepsilon_{t-1}$ (Sentana, 1995) or by a GJR–GARCH process $h_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2 + \alpha^*\varepsilon_{t-1}^{+2}$ (Glosten, Jagannathan, and Runkle, 1993). If the errors follow a QGARCH process then the expression for $E\varepsilon_t^2\varepsilon_{t-k}$ is given by

$$E\varepsilon_t^2\varepsilon_{t-k} \quad = \quad \alpha^*(\alpha+\beta)^{k-1}Eh_t^2, \quad k \geq 1,$$

where

$$Eh_t^2 \quad = \quad \frac{\omega}{1-(\alpha+\beta)}.$$

Figure 2: AR–TGARCH: Unconditional skewness of y_t as a function of ϕ for the following values of the TGARCH parameters: x-axis: ϕ, y-axis: unconditional skewness, $\omega = 0.005$, $\alpha = 0.05$, $\beta = 0.94$ (solid line), $\beta = 0.90$ (dashed line), $\alpha^* = 0.04$ (left), $\alpha^* = 0$ (middle), and $\alpha^* = -0.04$ (right).



If the errors follow a GJR–GARCH process, $d = 2$ and $c_t = \alpha z_t^2 + \beta + \alpha^* z_t^{+2}$ in (5),

$$E\varepsilon_t^2 \varepsilon_{t-k} \quad = \quad \alpha^*(\alpha + \beta + \alpha^* E z_t^{+2})^{k-1} E z_t^{+3} E h_t^3, \quad k \geq 1.$$

An explicit expression for Eh_t^3 is not available but is known not to be trivially zero. Also in these cases, $E\varepsilon_t^2\varepsilon_{t-k} \neq 0$ if and only if $\alpha^* \neq 0$. Thus, if the conditional mean is symmetric and linear, and the conditional second moment is symmetric, the unconditional marginal distribution for y_t is symmetric around zero. In the bottom panels of Figure 1, the thick line corresponding to $\phi^+ = 0$ illustrates this finding. It is emphasized that the results just discussed are only obtained if $\phi \neq 0$. Thus, at least some linear dependence in $\{y_t\}$ is necessary for skewness in the unconditional distribution of y_t.

We may also note that in the case of the first-order asQGARCH model $h_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}^2 + \alpha^*\varepsilon_{t-1} + \alpha^{**}\varepsilon_{t-1}^+$ (Brännäs and de Gooijer, 2004), the expression for $E\varepsilon_t^2\varepsilon_{t-k}$ becomes very complicated, and it does not seem possible to derive an analytical form for it. However, this moment seems to be nonzero for $k \geq 1$ because its components are not trivially zero.[3]

4.3 GARCH-in-mean model

It is well known that when the conditional standard deviation, conditional variance or any other nontrivial function of these, enters the conditional mean, $\gamma_{11}(k) \neq 0$ for $k \geq 1$ if $\gamma_2 < \infty$; see Hong (1991). It may be less well known that in this case $\gamma_3 \neq 0$. As an example, consider the TGARCH–M model (1),(2), and (4)–(6) so that $Ey_t = 0$. Since h_t^δ is a positive-valued variable and its distribution is asymmetric, it follows that $Ey_t^3 \neq 0$. The third moment and third-order cross-moments of the third-order stationary TGARCH–M process are given by Lemma 4. The unconditional skewness from a TGARCH–M model with $\delta = 1$ and $\delta = 2$ are plotted in Figure 3 as a function of ϕ. The figure shows that the range of possible skewness increases with the persistence of the GARCH process. It is also seen that when the conditional

[3]In fact, it seems that there is no explicit expression for any moment $E\varepsilon_t^m$, $m > 1$, for this model – at least it seems that there does not exist an analytic form for Eh_t^2.

standard deviation enters the conditional mean, the distribution becomes more skewed than it would be if the conditional mean were a function of the conditional variance. Assuming $\phi = 0$ implies $\gamma_3 = 0$ regardless of any asymmetry in the conditional standard deviation or conditional variance. This is also seen from Figure 3 where all the lines intersect the x-axis at $\phi = 0$. In this case the only nonzero cross-moments are $Ey_t^2 y_{t-k} = Eh_t^2 \varepsilon_{t-k}$; see the discussion in the previous subsection.

Assuming that h_t is defined by the standard AVGARCH model, the expressions for the moments simplify somewhat. Then $Eh_t^2 \varepsilon_{t-k} = Eh_t \varepsilon_{t-k} = Eh_t h_{t-k} \varepsilon_{t-k} = 0$, $k \geq 1$, as can be seen from the expressions in Lemma 5 and Corollary 5. However, provided that $\phi \neq 0$, all third-order moments are nonzero regardless of whether the conditional standard deviation is symmetric or asymmetric. The amount of skewness in this case is considerably less than in the case of the TGARCH–M model, which can be seen by comparing the top and bottom panels of Figure 3.

This example demonstrates that the third-moment structure in the case of the TGARCH–M or AVGARCH–M model is richer than it is in MA–TGARCH and MA–AVGARCH models, respectively. It can be concluded that both asymmetric and nonlinear responses to shocks in the conditional mean play an important role in producing skewness in the marginal density of y_t.

5 Shock impact curves

Engle and Ng (1993) defined the news impact curve as a function that describes the impact of a shock ε_{t-1} on current volatility expressed as conditional variance h_t^2. The shock is the component of the return y_t that can be characterized as 'news' to the agents in the following model:

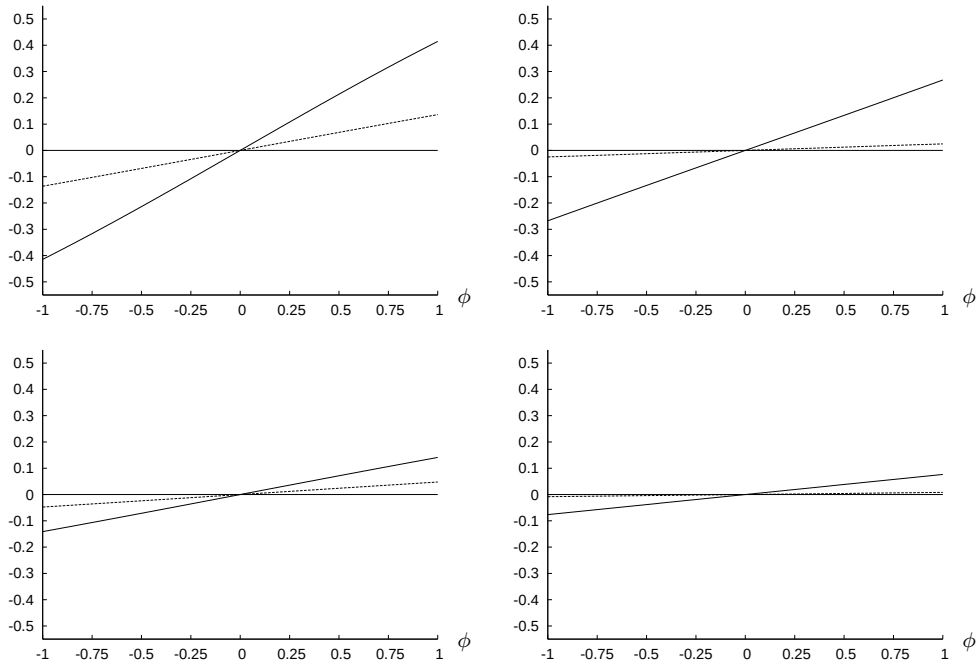
$$y_t = f(y_{t-j}, \varepsilon_{t-j}; j \geq 1) + \varepsilon_t.$$

In this model, the conditional mean $E_{t-1} y_t$ is not constant over time but is a function of past shocks. It is assumed that the conditional mean component is not news but rather structure known to the agents. For this reason, the NIC is measuring the impact of a shock on the conditional variance of the return. Nevertheless, for the purposes of this paper it will be useful to introduce a slight extension that also involves the shock coming through the conditional mean. It is called the Shock Impact Curve (SIC) and describes the impact of the shock on the conditional mean squared error of the return. The SIC is defined as follows:

$$E_{t-1}^{\mathrm{SIC}} y_t^2 = \mu_t^2(\sigma) + h_t^2(\sigma) \tag{7}$$

where $\mu(\sigma)$ and $h^2(\sigma)$ are the conditional mean and variance with elements in $\mathcal{F}_{t-2}$ replaced with their unconditional counterparts, for instance $Var_{t-k-1}\, y_{t-k} = h_{t-k}^2$, $k \geq 1$, is replaced with $\sigma^2 \stackrel{\text{def}}{=} Var\, y_t = E\mu_t^2 + Eh_t^2$. It may be argued that the correlation structure of $\{y_t\}$ is known to the agents, whereby that part of the response does not qualify as news. If this structure is weak, however, it may be difficult in

Figure 3: TGARCH–M: Unconditional skewness of y_t as a function of ϕ for the following values of δ and the TGARCH parameters: $\delta = 1$ (left-hand panels) and $\delta = 2$ (right-hand panels), TGARCH parameters: $\omega = 0.005$, $\alpha = 0.05$, $\beta = 0.94$ (solid line), $\beta = 0.90$ (dashed line), $\alpha^* = \pm 0.04$ (top panels), and $\alpha^* = 0$ (bottom panels).



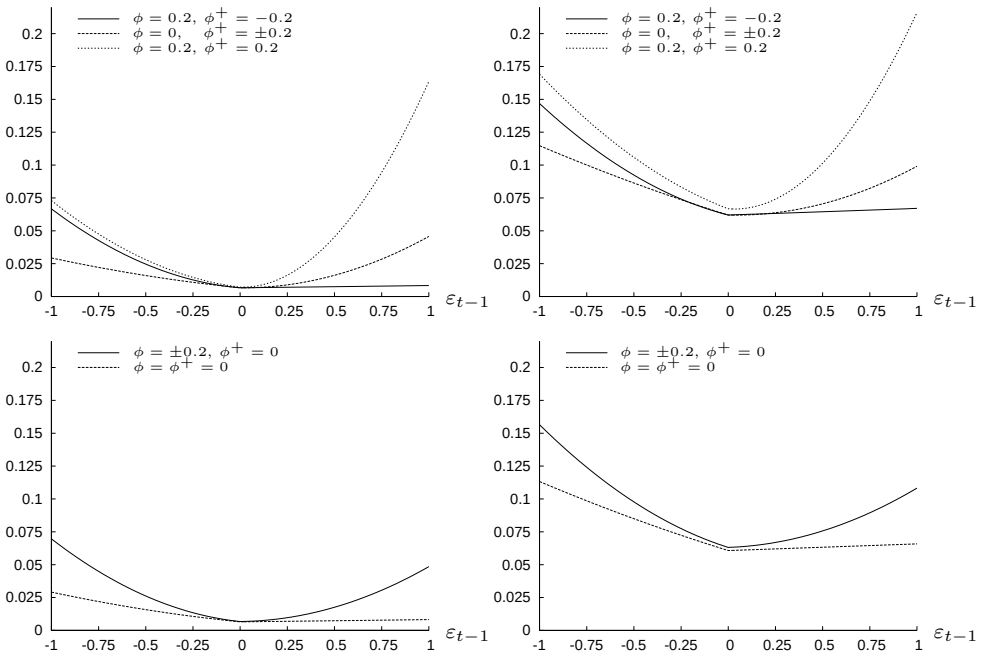
practice to separate this effect from the actual 'news'. If $\mu_t = 0$, SIC coincides with NIC. Conversely, the impact of 'news' and that of a 'shock' on the next return can have rather different shapes.

The SIC can be used to study the effect of a shock on both the conditional mean and the conditional variance. While one may expect negative 'news' to have a stronger effect on volatility than positive ones, it may be interesting to see what the situation is when the unconditional mean is assumed to have some structure. For the asMA process in (3)

$$\mu_t^2(\sigma) = \phi^{+2}(E\varepsilon_t^+)^2 + \phi^2\varepsilon_{t-1}^2 - 2\phi\phi^+\varepsilon_{t-1}E\varepsilon_t^+ + (\phi^{+2} + 2\phi\phi^+)\varepsilon_{t-1}^{+2} - 2\phi^{+2}\varepsilon_{t-1}^+E\varepsilon_t^+$$

In Figure 4 the top panels show the shock impact curves for a selection of parameters for the asMA process. The solid line represents the case in which the conditional mean only responds to negative shocks, whereas the other two curves represent models in which the effect of positive shocks is pronounced. Consider first the top right-hand panel, look at the dashed line ($\phi = 0$, $\phi^+ = \pm 0.2$) and compare it with the corresponding one in the top left-hand panel. Even if the conditional mean only responds to positive shocks, the impact of a shock can be larger for negative shocks than for positive ones as long as the persistence of the GARCH process is high. The

Figure 4: asMA–TGARCH: Shock impact curves for different values for parameters for the conditional mean, TGARCH parameters: $\omega = 0.005$, $\alpha = 0.05$, $\beta = 0.90$ (left-hand panels), $\beta = 0.94$ (right-hand panels), and $\alpha^* = -0.04$.



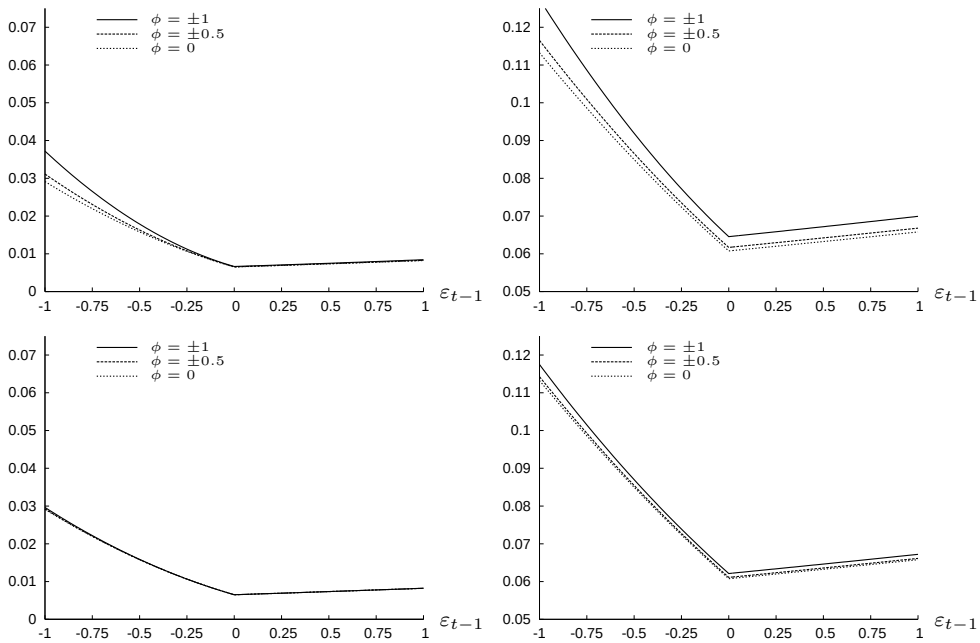
bottom panels show the SIC when the conditional mean either follows an MA process or is constant. If the conditional mean is a linear MA process, the response to shocks due to the conditional mean is symmetric, $\mu_t^2(\sigma) = \phi^2\varepsilon_{t-1}^2$. In this case the effect of the conditional mean dominates the effect of the conditional variance so that the impact of a shock on the mean squared error is almost symmetric even if the GARCH process is asymmetric. A comparison of the solid lines in the bottom left and right-hand panels results in a similar conclusion in that the increased persistence in the GARCH process emphasizes the role of the conditional variance in the shock impact curve. Replacing the TGARCH process with a different asymmetric GARCH process has virtually no effect on the shape of the curves in Figure 4. A symmetric GARCH process would somewhat dampen the impact of negative shocks, in which case the curves in the bottom panels would be symmetric around zero.

For the GARCH–M process in (4),

$$\mu_t^2(\sigma) = \phi^2(Eh_t^\delta)^2 + \phi^2 h_t^{2\delta} - 2\phi^2 h_t^\delta Eh_t^\delta$$

In this case, the effect entering through the conditional variance is the one controlling the response to shocks. In Figure 5 the shock impact curves are plotted for TGARCH–M model with $\delta = 1$ and 2 in (4). Comparing the left- and right-hand panels shows

Figure 5: TGARCH–M: Shock impact curves for different values for parameters for the conditional mean, $\delta = 1$ (top panels), $\delta = 2$ (bottom panels); TGARCH parameters: $\omega = 0.005$, $\alpha = 0.05$, $\beta = 0.90$ (left-hand panels), $\beta = 0.94$ (right-hand panels), and $\alpha^* = -0.04$.



that increased persistence in the TGARCH process magnifies the impact of shocks. The asymmetry of the impact is inherited from the GARCH process. Replacing the TGARCH process with a symmetric GARCH process produces shock impact curves that are symmetric around zero.

6 Conclusions

In this paper we show how different parameterizations of the conditional mean and variance contribute to the asymmetry in the unconditional distribution of y_t. This is important because marginal distributions of return series often appear skewed. It is thus useful to know the structure of the unconditional distribution implied by a model that has asymmetries or nonlinearities in the first and the second conditional moment.

The models we have considered in detail are the asMA–TGARCH and the TGARCH–M model. In the former model, both the conditional mean and the conditional standard deviation are asymmetric around zero. The latter model even has a nonlinear mean. We derive the analytic expressions for the third-order moment

structure of these models and consider various special cases of the asMA–TGARCH model in which the mean and/or the standard deviation specification is restricted to be symmetric. Similar considerations are made in the case of the TGARCH–M model. In general, we find that asymmetries or nonlinearities in the conditional mean are of greater importance than they are in the conditional standard deviation or variance when it comes to generating skewed marginal distributions. If the conditional mean is symmetric and linear, then the unconditional skewness can only follow from the asymmetry of the conditional standard deviation or variance. However, in that case the third-moment structure of the variable of interest is no longer particularly flexible. But then, if the conditional mean is asymmetric or nonlinear, the distribution of y_t can be even strongly skewed regardless of whether or not the conditional standard deviation or variance is symmetric or asymmetric.

It may be of interest to see how the past news affect not only the current volatility but magnitude of today's returns. We introduce a definition of the shock impact curve which describes the impact of a shock on the mean squared error of the return. It combines the effects of the conditional mean and the conditional standard deviation or variance on the squared returns. The conditional mean can strongly dominate the shape of the news impact curves.

It would be interesting to consider a wider variety of specifications for the conditional mean and variance and derive the corresponding expressions in these cases. However, for many models, such as the standard GARCH model and some of its extensions, analytical expressions for third-order moments are not available. Our simulation experiments show that the same conclusions can be drawn when the TGARCH or AVGARCH process is replaced with other GARCH models that parameterize the conditional variance instead of the conditional standard deviation.

Finally, because a skewed marginal distribution can be a result of some type of asymmetric or nonlinear behaviour in the process for the conditional mean, testing for asymmetries and nonlinearities in the conditional mean is important. If an asymmetric or nonlinear model is found suitable, this may have implications on the unconditional third-moment structure of the process. Of course, any comparison of the unconditional moments estimated from the data with the moments implied by the fitted model (plug-in estimation) is dependent on simulations whenever analytical expressions for the moments of interest are not available.

Appendix

Lemma 1 (First and second-moment structure of asMA) *Consider an asymmetric MA process (1)–(3) and (5) that is second order stationary. The unconditional first and second-order moments and cross-moments of y_t are given by*

$$
\begin{aligned}
Ey_t &= 0, && (8)\\
Var\, y_t &= E\mu_t^2 + E\varepsilon_t^2, && (9)\\
\gamma_{11}(k) &= E\mu_t\mu_{t-k} + E\mu_t\varepsilon_{t-k}, \quad k \geq 1, && (10)
\end{aligned}
$$

where

$$
\begin{aligned}
E\mu_t^2 &= \phi^2 E\varepsilon_t^2 + (2\phi\phi^+ + \phi^{+2})E\varepsilon_t^{+2} - \phi^{+2}(E\varepsilon_t^+)^2\\
E\mu_t\mu_{t-k} &= \phi\phi^+ E\varepsilon_{t-k}\varepsilon_t^+ + \phi^{+2}E\varepsilon_t^+\varepsilon_{t-k}^+ - \phi^{+2}(E\varepsilon_t^+)^2\\
E\mu_t\varepsilon_{t-k} &= \begin{cases} \phi E\varepsilon_t^2 + \phi^+ E\varepsilon_t^{+2}, & k = 1\\ \phi^+ E\varepsilon_{t-k}\varepsilon_{t-1}^+, & k > 1. \end{cases}
\end{aligned}
$$

The moments of ε_t and ε_t^+ can be found in Lemma 5.

Proof. The results are readily obtained by straightforward algebra. ■

Lemma 2 (Third-moment structure of asMA) *Consider an asymmetric MA process (1)–(3) and (5) that is third-order stationary. The unconditional third-order moments and cross-moments of y_t are given by*

$$
\begin{aligned}
\gamma_3 &= E\mu_t^3 + 3E\mu_t\varepsilon_t^2, && (11)\\
\gamma_{21}(k) &= E\mu_t^2\mu_{t-k} + E\mu_{t-k}\varepsilon_t^2 + E\mu_t^2\varepsilon_{t-k} + E\varepsilon_t^2\varepsilon_{t-k}, \quad k \geq 1 && (12)\\
\gamma_{12}(k) &= E\mu_t\mu_{t-k}^2 + E\mu_t\varepsilon_{t-k}^2 + 2E\mu_t\mu_{t-k}\varepsilon_{t-k}, \quad k \geq 1 && (13)
\end{aligned}
$$

where

$$
\begin{aligned}
E\mu_t^3 &= (3\phi^2\phi^+ + 3\phi\phi^{+2} + \phi^{+3})E\varepsilon_t^{+3} - (6\phi\phi^{+2} + 3\phi^{+3})E\varepsilon_t^{+2}E\varepsilon_t^+\\
&\quad -3\phi^2\phi^+ E\varepsilon_t^2 E\varepsilon_t^+ + 2\phi^{+3}(E\varepsilon_t^+)^3\\
E\mu_t^2\mu_{t-k} &= \phi^3 E\varepsilon_t^2\varepsilon_{t-k} + (2\phi^2\phi^+ + \phi\phi^{+2})E\varepsilon_{t-k}\varepsilon_t^{+2}\\
&\quad +\phi^2\phi^+(E\varepsilon_t^2\varepsilon_{t-k}^+ - E\varepsilon_t^2 E\varepsilon_t^+) + (2\phi\phi^{+2} + \phi^{+3})(E\varepsilon_t^{+2}\varepsilon_{t-k}^+ - E\varepsilon_t^{+2}E\varepsilon_t^+)\\
&\quad -2\phi\phi^{+2}E\varepsilon_{t-k}\varepsilon_t^+ E\varepsilon_t^+ - 2\phi^{+3}(E\varepsilon_t^+\varepsilon_{t-k}^+ E\varepsilon_t^+ - (E\varepsilon_t^+)^3)\\
E\mu_t\mu_{t-k}^2 &= \phi^2\phi^+(E\varepsilon_{t-k}^2\varepsilon_t^+ - E\varepsilon_t^2 E\varepsilon_t^+) + (2\phi\phi^{+2} + \phi^{+3})(E\varepsilon_t^+\varepsilon_{t-k}^{+2} - E\varepsilon_t^{+2}E\varepsilon_t^+)\\
&\quad -2\phi\phi^{+2}E\varepsilon_{t-k}\varepsilon_t^+ E\varepsilon_t^+ - 2\phi^{+3}(E\varepsilon_t^+\varepsilon_{t-k}^+ E\varepsilon_t^+ - (E\varepsilon_t^+)^3)\\
E\mu_t\varepsilon_t^2 &= \phi E\varepsilon_t^2\varepsilon_{t-1} + \phi^+(E\varepsilon_t^2\varepsilon_{t-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+)\\
E\mu_{t-k}\varepsilon_t^2 &= \phi E\varepsilon_t^2\varepsilon_{t-k-1} + \phi^+(E\varepsilon_t^2\varepsilon_{t-k-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+)
\end{aligned}
$$

$$E\mu_t\varepsilon_{t-k}^2 = \begin{cases} \phi^+(E\varepsilon_t^{+3} - E\varepsilon_t^2 E\varepsilon_t^+), & k=1 \\ \phi^+(E\varepsilon_{t-(k-1)}^2\varepsilon_t^+ - E\varepsilon_t^2 E\varepsilon_t^+), & k>1 \end{cases}$$

$$E\mu_t^2\varepsilon_{t-k} = \begin{cases} (2\phi\phi^+ + \phi^{+2})E\varepsilon_t^{+3} - 2\phi\phi^+ E\varepsilon_t^2 E\varepsilon_t^+ - 2\phi^{+2}E\varepsilon_t^{+2}E\varepsilon_t^+, & k=1 \\ (2\phi\phi^+ + \phi^{+2})E\varepsilon_{t-(k-1)}\varepsilon_t^{+2} - 2\phi^{+2}E\varepsilon_{t-(k-1)}\varepsilon_t^+ E\varepsilon_t^+ & \\ +\phi^2 E\varepsilon_t^2\varepsilon_{t-(k-1)}, & k>1 \end{cases}$$

$$E\mu_t\mu_{t-k}\varepsilon_{t-k} = \begin{cases} \phi\phi^+(E\varepsilon_t^2\varepsilon_{t-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+ + E\varepsilon_{t-1}\varepsilon_t^{+2}) & \\ +\phi^{+2}(E\varepsilon_t^{+2}\varepsilon_{t-1}^+ - E\varepsilon_t^{+2}E\varepsilon_t^+) + \phi^2 E\varepsilon_t^2\varepsilon_{t-1}, & k=1 \\ \phi\phi^+ E\varepsilon_{t-(k-1)}\varepsilon_{t-k}\varepsilon_t^+ & \\ +\phi^{+2}(E\varepsilon_{t-(k-1)}\varepsilon_t^+\varepsilon_{t-k}^+ - E\varepsilon_{t-(k-1)}\varepsilon_t^+ E\varepsilon_t^+), & k>1. \end{cases}$$

The moments of ε_t and ε_t^+ can be found in Lemma 5.

Proof. The results are readily obtained by tedious but straightforward algebra. ■

Corollary 1 (Third-moment structure of asMA with symmetric GARCH) *Consider an asymmetric MA process (1)–(3) and (5) that is third-order stationary. Furthermore, assume that c_t in (5) is even with respect to z_t. The unconditional third-order moments and cross-moments of y_t are given by (11)–(13) in Lemma 2 where*

$$\begin{aligned} E\mu_t^3 &= (3\phi^2\phi^+ + 3\phi\phi^{+2} + \phi^{+3})E\varepsilon_t^{+3} - (6\phi\phi^{+2} + 3\phi^{+3})E\varepsilon_t^{+2}E\varepsilon_t^+ \\ &\quad -3\phi^2\phi^+ E\varepsilon_t^2 E\varepsilon_t^+ + 2\phi^{+3}(E\varepsilon_t^+)^3 \\ E\mu_t^2\mu_{t-k} &= \phi^2\phi^+(E\varepsilon_t^2\varepsilon_{t-k}^+ - E\varepsilon_t^2 E\varepsilon_t^+) + (2\phi\phi^{+2} + \phi^{+3})(E\varepsilon_t^{+2}\varepsilon_{t-k}^+ - E\varepsilon_t^{+2}E\varepsilon_t^+) \\ &\quad -2\phi^{+3}(E\varepsilon_t^+\varepsilon_{t-k}^+ E\varepsilon_t^+ - (E\varepsilon_t^+)^3) \\ E\mu_t\mu_{t-k}^2 &= \phi^2\phi^+(E\varepsilon_{t-k}^2\varepsilon_t^+ - E\varepsilon_t^2 E\varepsilon_t^+) + (2\phi\phi^{+2} + \phi^{+3})(E\varepsilon_t^+\varepsilon_{t-k}^{+2} - E\varepsilon_t^{+2}E\varepsilon_t^+) \\ &\quad -2\phi^{+3}(E\varepsilon_t^+\varepsilon_{t-k}^+ E\varepsilon_t^+ - (E\varepsilon_t^+)^3) \\ E\mu_t\varepsilon_t^2 &= \phi^+(E\varepsilon_t^2\varepsilon_{t-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+) \\ E\mu_{t-k}\varepsilon_t^2 &= \phi^+(E\varepsilon_t^2\varepsilon_{t-k-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+) \\ E\mu_t\varepsilon_{t-k}^2 &= \begin{cases} \phi^+(E\varepsilon_t^{+3} - E\varepsilon_t^2 E\varepsilon_t^+), & k=1 \\ \phi^+(E\varepsilon_{t-k}^2\varepsilon_{t-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+), & k>1 \end{cases} \\ E\mu_t^2\varepsilon_{t-k} &= \begin{cases} (2\phi\phi^+ + \phi^{+2})E\varepsilon_t^{+3} - 2\phi_1\phi_1^+ E\varepsilon_t^2 E\varepsilon_t^+ - 2\phi_1^{+2}E\varepsilon_t^{+2}E\varepsilon_t^+, & k=1 \\ 0, & k>1 \end{cases} \\ E\mu_t\mu_{t-k}\varepsilon_{t-k} &= \begin{cases} \phi\phi^+(E\varepsilon_t^2\varepsilon_{t-1}^+ - E\varepsilon_t^2 E\varepsilon_t^+ + E\varepsilon_{t-1}\varepsilon_t^{+2}) & \\ +\phi^{+2}(E\varepsilon_t^{+2}\varepsilon_{t-1}^+ - E\varepsilon_t^{+2}E\varepsilon_t^+), & k=1 \\ 0, & k>1. \end{cases} \end{aligned}$$

The moments of ε_t and ε_t^+ can be found in Lemma 5.

Corollary 2 (Third-moment structure of asMA with $\phi = 0$) *Consider an asymmetric MA process (1)–(3) and (5) with $\phi = 0$ that is third-order stationary. The unconditional third-order moments and cross-moments of y_t are given by (11)–(13) in Lemma 2 where*

$$
\begin{aligned}
E\mu_t^3 &= \phi^{+3}(E\varepsilon_t^{+3} - 3E\varepsilon_t^{+2}E\varepsilon_t^{+} + 2(E\varepsilon_t^{+})^3) \\
E\mu_t^2\mu_{t-k} &= \phi^{+3}(E\varepsilon_t^{+2}\varepsilon_{t-k}^{+} - E\varepsilon_t^{+2}E\varepsilon_t^{+}) - 2\phi^{+3}(E\varepsilon_t^{+}\varepsilon_{t-k}^{+}E\varepsilon_t^{+} - (E\varepsilon_t^{+})^3) \\
E\mu_t\mu_{t-k}^2 &= \phi^{+3}(E\varepsilon_t^{+}\varepsilon_{t-k}^{+2} - E\varepsilon_t^{+2}E\varepsilon_t^{+}) - 2\phi^{+3}(E\varepsilon_t^{+}\varepsilon_{t-k}^{+}E\varepsilon_t^{+} - (E\varepsilon_t^{+})^3) \\
E\mu_t\varepsilon_t^2 &= \phi^{+}(E\varepsilon_t^2\varepsilon_{t-1}^{+} - E\varepsilon_t^2E\varepsilon_t^{+}) \\
E\mu_{t-k}\varepsilon_t^2 &= \phi^{+}(E\varepsilon_t^2\varepsilon_{t-k-1}^{+} - E\varepsilon_t^2E\varepsilon_t^{+}) \\
E\mu_t\varepsilon_{t-k}^2 &= \begin{cases} \phi^{+}(E\varepsilon_t^{+3} - E\varepsilon_t^2E\varepsilon_t^{+}), & k = 1 \\ \phi^{+}(E\varepsilon_{t-k}^2\varepsilon_{t-1}^{+} - E\varepsilon_t^2E\varepsilon_t^{+}), & k > 1 \end{cases} \\
E\mu_t^2\varepsilon_{t-k} &= \begin{cases} \phi^{+2}(E\varepsilon_t^{+3} - 2E\varepsilon_t^{+2}E\varepsilon_t^{+}), & k = 1 \\ \phi^{+2}(E\varepsilon_{t-k}\varepsilon_{t-1}^{+2} - 2E\varepsilon_{t-k}\varepsilon_{t-1}^{+}E\varepsilon_t^{+}), & k > 1 \end{cases} \\
E\mu_t\mu_{t-k}\varepsilon_{t-k} &= \begin{cases} \phi^{+2}(E\varepsilon_t^{+2}\varepsilon_{t-1}^{+} - E\varepsilon_t^{+2}E\varepsilon_t^{+}), & k = 1 \\ \phi^{+2}(E\varepsilon_{t-k}\varepsilon_{t-1}^{+}\varepsilon_{t-k-1}^{+} - E\varepsilon_{t-k}\varepsilon_{t-1}^{+}E\varepsilon_t^{+}), & k > 1. \end{cases}
\end{aligned}
$$

The moments of ε_t and ε_t^{+} can be found in Lemma 5.

Corollary 3 (Third-moment structure of asMA with $\phi = 0$ and symmetric GARCH) *Consider an asymmetric MA process (1)–(3) and (5) with $\phi = 0$ that is third-order stationary. Furthermore, assume that c_t in (5) is even with respect to z_t. The unconditional third-order moments and cross-moments of y_t are given by (11)–(13) in Lemma 2 where*

$$
\begin{aligned}
E\mu_t^3 &= \phi^{+3}(E\varepsilon_t^{+3} - 3E\varepsilon_t^{+2}E\varepsilon_t^{+} + 2(E\varepsilon_t^{+})^3) \\
E\mu_t^2\mu_{t-k} &= \phi^{+3}E\varepsilon_t^{+2}\varepsilon_{t-k}^{+} - \phi^{+3}E\varepsilon_t^{+2}E\varepsilon_t^{+} - 2\phi^{+3}(E\varepsilon_t^{+}\varepsilon_{t-k}^{+}E\varepsilon_t^{+} - (E\varepsilon_t^{+})^3) \\
E\mu_t\mu_{t-k}^2 &= \phi^{+3}E\varepsilon_t^{+}\varepsilon_{t-k}^{+2} - \phi^{+3}E\varepsilon_t^{+2}E\varepsilon_t^{+} - 2\phi^{+3}(E\varepsilon_t^{+}\varepsilon_{t-k}^{+}E\varepsilon_t^{+} - (E\varepsilon_t^{+})^3) \\
E\mu_t\varepsilon_t^2 &= \phi^{+}(E\varepsilon_t^2\varepsilon_{t-1}^{+} - E\varepsilon_t^2E\varepsilon_t^{+}) \\
E\mu_{t-k}\varepsilon_t^2 &= \phi^{+}(E\varepsilon_t^2\varepsilon_{t-k-1}^{+} - E\varepsilon_t^2E\varepsilon_t^{+}) \\
E\mu_t\varepsilon_{t-k}^2 &= \begin{cases} \phi^{+}(E\varepsilon_t^{+3} - E\varepsilon_t^2E\varepsilon_t^{+}), & k = 1 \\ \phi^{+}(E\varepsilon_{t-k}^2\varepsilon_{t-1}^{+} - E\varepsilon_t^2E\varepsilon_t^{+}), & k > 1 \end{cases}
\end{aligned}
$$

$$E\mu_t^2\varepsilon_{t-k} = \begin{cases} \phi^{+2}(E\varepsilon_t^{+3} - 2E\varepsilon_t^{+2}E\varepsilon_t^{+}), & k=1 \\ 0, & k>1 \end{cases}$$

$$E\mu_t\mu_{t-k}\varepsilon_{t-k} = \begin{cases} \phi^{+2}(E\varepsilon_t^{+2}\varepsilon_{t-1}^{+} - E\varepsilon_t^{+2}E\varepsilon_t^{+}), & k=1 \\ 0, & k>1. \end{cases}$$

The moments of ε_t and ε_t^+ can be found in Lemma 5.

Corollary 4 (Third-moment structure of MA) *Consider an asymmetric MA process (1)–(3) and (5) with $\phi^+ = 0$ that is third-order stationary. The unconditional third-order moments and cross-moments of y_t are given by (11)–(13) in Lemma 2 where*

$$E\mu_t^3 = E\mu_t\mu_{t-k}^2 = E\mu_t\varepsilon_{t-k}^2 = 0$$

and

$$\begin{aligned} E\mu_t\varepsilon_t^2 &= \phi E\varepsilon_t^2\varepsilon_{t-1} \\ E\mu_t^2\mu_{t-k} &= \phi^3 E\varepsilon_t^2\varepsilon_{t-k} \\ E\mu_{t-k}\varepsilon_t^2 &= \phi E\varepsilon_t^2\varepsilon_{t-(k+1)} \\ E\mu_t^2\varepsilon_{t-k} &= \begin{cases} 0, & k=1 \\ \phi^2 E\varepsilon_t^2\varepsilon_{t-(k-1)}, & k>1 \end{cases} \\ E\mu_t\mu_{t-k}\varepsilon_{t-k} &= \begin{cases} \phi^2 E\varepsilon_t^2\varepsilon_{t-1}, & k=1 \\ 0, & k>1. \end{cases} \end{aligned}$$

The moments of ε_t and ε_t^+ can be found in Lemma 5.

If the conditional second moment is parameterized such that $E\varepsilon_t\varepsilon_{t-k} = 0$ for all $k \geq 1$ (for instance c_t in (5) is symmetric with respect to z_t), then the third-order moments and cross-moments are zero.

Lemma 3 (First and second-moment structure of GARCH–M) *Consider a GARCH–M process (1), (2), (4), and (5) that is third-order stationary. The unconditional first and second-order moments and cross-moments of y_t are given by (8)–(10) in Lemma 1 where*

$$\begin{aligned} E\mu_t^2 &= \phi^2(Eh_t^{2\delta} - (Eh_t^{\delta})^2) \\ E\mu_t\mu_{t-k} &= \phi^2(Eh_t^{\delta}h_{t-k}^{\delta} - (Eh_t^{\delta})^2) \\ E\mu_t\varepsilon_{t-k} &= \phi Eh_t^{\delta}\varepsilon_{t-k}. \end{aligned}$$

The moments of ε_t and h_t can be found in Lemma 5.

Proof. The results are readily obtained by straightforward algebra. ■

Lemma 4 (Third-moment structure of GARCH–M) *Consider a GARCH–M process (1), (2), (4), and (5) that is third-order stationary. The unconditional third-order moments and cross-moments of y_t are given by (11)–(13) in Lemma 2 where*

$$\begin{aligned}
E\mu_t^3 &= \phi^3(Eh_t^{3\delta} - 3Eh_t^{2\delta}Eh_t^{\delta} + 2(Eh_t^{\delta})^3) \\
E\mu_t^2\mu_{t-k} &= \phi^3(Eh_t^{2\delta}h_{t-k}^{\delta} - 2Eh_t^{\delta}h_{t-k}^{\delta}Eh_t^{\delta} - Eh_t^{2\delta}Eh_t^{\delta} + 2(Eh_t^{\delta})^3) \\
E\mu_t\mu_{t-k}^2 &= \phi^3(Eh_t^{\delta}h_{t-k}^{2\delta} - 2Eh_t^{\delta}h_{t-k}^{\delta}Eh_t^{\delta} - Eh_t^{2\delta}Eh_t^{\delta} + 2(Eh_t^{\delta})^3) \\
E\mu_t^2\varepsilon_{t-k} &= \phi^2(Eh_t^{2\delta}\varepsilon_{t-k} - 2Eh_t^{\delta}\varepsilon_{t-k}Eh_t^{\delta}) \\
E\mu_t\mu_{t-k}\varepsilon_{t-k} &= \phi^2(Eh_t^{\delta}h_{t-k}^{\delta}\varepsilon_{t-k} - Eh_t^{\delta}\varepsilon_{t-k}Eh_t^{\delta}) \\
E\mu_t\varepsilon_t^2 &= \phi(Eh_t^{\delta+2} - Eh_t^{\delta}Eh_t^2) \\
E\mu_t\varepsilon_{t-k}^2 &= \phi(Eh_t^{\delta}\varepsilon_{t-k}^2 - Eh_t^{\delta}Eh_t^2) \\
E\mu_{t-k}\varepsilon_t^2 &= \phi(Eh_t^2h_{t-k}^{\delta} - Eh_t^{\delta}Eh_t^2).
\end{aligned}$$

The moments of ε_t and ε_t^+ can be found in Lemma 5. If the c_t in (5) is even with respect to z_t, then $E\mu_t^2\varepsilon_{t-k} = E\mu_t\mu_{t-k}\varepsilon_{t-k} = 0$.

Proof. The results are readily obtained by straightforward algebra. For the results for the symmetric GARCH we make use of Corollary 5. ■

Lemma 5 *Consider the GARCH model (2) and (5). Suppose that ε_t is stationary with time-invariant moments. Provided that the moments exist, the moments of ε_t, ε_t^+, and h_t are given by*

$$\begin{aligned}
E\varepsilon_t^m &= \begin{cases} d_{0m}Eh_t^m, & m \text{ even} \\ 0, & m \text{ odd} \end{cases} \\
E\varepsilon_t^{+m} &= d_{0m}^+Eh_t^m \\
E\varepsilon_t^m\varepsilon_{t-k}^{(+)n} &= \begin{cases} d_{0m}Eh_t^m\varepsilon_{t-k}^{(+)n}, & m \text{ even} \\ 0, & m \text{ odd} \end{cases} \\
E\varepsilon_t^{+m}\varepsilon_{t-k}^{(+)n} &= d_{0m}^+Eh_t^m\varepsilon_{t-k}^{(+)n} \\
E\varepsilon_{t-k}\varepsilon_t^+\varepsilon_{t-k-1}^{(+)} &= d_{01}^+Eh_t\varepsilon_{t-k}\varepsilon_{t-k-1}^{(+)}
\end{aligned}$$

where

$$\begin{aligned}
Eh_t^{dm} &= \frac{1}{1-d_{m0}}\sum_{j=1}^{m}\binom{m}{j}\omega^j d_{(m-j)0}Eh_t^{d(m-j)} \\
Eh_t^{dn}\varepsilon_{t-k}^{(+)m} &= \begin{cases} \sum_{j=0}^{n}\binom{n}{j}\omega^j d_{(n-j)m}^{(+)}Eh_t^{d(n-j)+m}, & k=1 \\ \sum_{j=0}^{n}\binom{n}{j}\omega^j d_{(n-j)0}Eh_t^{d(n-j)}\varepsilon_{t-(k-1)}^{(+)m}, & k>1 \end{cases} \\
Eh_t^d\varepsilon_{t-k}\varepsilon_{t-k-1}^{(+)} &= d_{10}^{k-1}d_{11}Eh_t^{d+1}\varepsilon_{t-1}^{(+)}
\end{aligned}$$

and

$$
\begin{aligned}
Eh_t^{dn} h_{t-k}^{dm} &= \begin{cases} \sum_{j=0}^{n} \binom{n}{j} \omega^j d_{(n-j)0} E h_t^{d(n-j+m)}, & k=1 \\ \sum_{j=0}^{n} \binom{n}{j} \omega^j d_{(n-j)0} E h_t^{d(n-j)} h_{t-(k-1)}^{dm}, & k>1 \end{cases} \\
Eh_t^d h_{t-k}^d \varepsilon_{t-k} &= d_{10}^{k-1} d_{11} E h_t^{2d+1}
\end{aligned}
$$

with $k \geq 1$ *and where*

$$
\begin{aligned}
d_{ij} &= E c_t^i z_t^j, \quad i \geq 0,\ j \geq 0 \\
d_{ij}^+ &= E c_t^i z_t^{+j}, \quad i \geq 0,\ j > 0
\end{aligned}
$$

In the expressions above the notation $(+)$ *means that* $+$ *is either included in or excluded from the equation in question. When* $d = 1$ *the recursions above yield analytically explicit expressions whereas for* $d = 2$ *some of them involve moments that have to be calculated numerically through simulations. If the conditional standard deviation follows the TGARCH process (5) and (6) then*

$$
\begin{aligned}
d_{ij} &= \sum_{\substack{0 \leq h_1, h_2, h_3 \leq i \\ h_1+h_2+h_3=i \\ h_3+j \ even}} \frac{i!}{h_1! h_2! h_3!} \alpha^{h_1} \beta^{h_2} \alpha^{*h_3} E|z_t|^{i+j-h_2}, \quad i \geq 1,\ j \geq 0 \\
d_{ij}^+ &= \sum_{\substack{0 \leq h_1, h_2, h_3 \leq i \\ h_1+h_2+h_3=i}} \frac{i!}{h_1! h_2! h_3!} \alpha^{h_1} \beta^{h_2} \alpha^{*h_3} E z_t^{+\ i+j-h_2}, \quad i \geq 1,\ j \geq 1
\end{aligned}
$$

For the AVGARCH process the expressions for d_{ij} *and* d_{ij}^+ *are obtained by setting* $\alpha^* = 0$*, restricting the index* $h_3 = 0$*, and defining* $0^0 = 1$*. Furthermore, if* $z_t \sim nid(0,1)$*, then the moments of* z_t*,* $|z_t|$*, and the censored variable* z_t^+ *are given in Lemma 6.*

Proof. The results are readily obtained by straightforward but tedious algebra. ■

Corollary 5 *Consider the GARCH model (2) and (5) in Lemma 5. If the process for the conditional second moment is symmetric in its response to shocks, then* $d_{nm} = 0$ *whenever* m *is odd. Hence, of the moments in Lemma 5,*

$$
Eh_t^{dn} \varepsilon_{t-k}^m = Eh_t^d h_{t-k}^d \varepsilon_{t-k} = Eh_t^d \varepsilon_{t-k} \varepsilon_{t-k-1}^{(+)} = 0
$$

and

$$
E\varepsilon_t^n \varepsilon_{t-k}^m = E\varepsilon_t^{+dn} \varepsilon_{t-k}^m = E\varepsilon_{t-k} \varepsilon_t^+ \varepsilon_{t-k-1}^{(+)} = 0
$$

where m *is odd.*

Proof. If $c_t(z_t)$ in (5) is an even function of z_t then the function $c_t^n(z_t) z_t^m$ is an odd function of z_t for any odd m and therefore $d_{nm} = 0$. ■

Lemma 6 *Assume $z_t \sim nid(0,1)$. Then the moments of z_t, $|z_t|$, and z_t^+ are given by*

$$
\begin{aligned}
Ez_t^m &= \begin{cases} \prod\limits_{\substack{i=1 \\ i \text{ odd}}}^{m-1} i, & m \text{ even} \\ 0, & m \text{ odd} \end{cases} \\
E|z_t|^m &= \begin{cases} Ez_t^m, & m \text{ even} \\ \prod\limits_{\substack{i=2 \\ i \text{ even}}}^{m-1} i, & m \text{ odd} \end{cases} \\
Ez_t^{+m} &= \frac{1}{2} E|z_t|^m.
\end{aligned}
$$

Proof. Straightforward by recursion and by noticing that the censored variable z_t^+ can be expressed as

$$z_t^+ = \max(0, z_t) = \frac{1}{2}(|z_t| + z_t).$$

Note that an empty product is defined to equal one. ■

References

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

Brännäs, K., and J. G. de Gooijer (1995): "Invertibility of nonlinear time series models," *Communications in Statistics: Theory and Methods*, 24, 2701–2714.

——— (2004): "Asymmetries in conditional mean and variance: Modelling stock returns by asMA–asQGARCH," *Journal of Forecasting*, 23, 155–171.

Brännäs, K., and N. Nordman (2003a): "An alternative conditional asymmetry specification for stock returns," *Applied Financial Economics*, 13, 537–541.

——— (2003b): "Conditional skewness modelling for stock returns," *Applied Economic Letters*, 10, 725–728.

Chen, J., H. Hong, and J. C. Stein (2001): "Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices," *Journal of Financial Economics*, 61, 345–381.

Engle, R. F. (1990): "Discussion: Stock market volatility and the crash of 87," *Review of Financial Studies*, 3, 103–106.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987): "Estimating time varying risk premia in the term structure: the ARCH-M model," *Econometrica*, 55, 391–407.

Engle, R. F., and V. K. Ng (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–78.

Engle, R. F., and A. J. Patton (2001): "What good is a volatility model?," *Quantitative Finance*, 1, 237–245.

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

González-Rivera, G. (1998): "Smooth-transition GARCH models," *Studies in Nonlinear Dynamics and Econometrics*, 3, 61–78.

Hagerud, G. E. (1997): "Specification tests for asymmetric GARCH," SSE/EFI Working Paper Series in Economics and Finance No. 163.

Hansen, B. E. (1994): "Autoregressive conditional density models," *International Economic Review*, 35, 705–730.

Harris, R., C. C. Küçüközmen, and F. Yilmaz (2004): "Skewness in the conditional distribution of daily equity returns," *Applied Financial Economics*, 14, 195–202.

Harvey, C. R., and A. Siddique (1999): "Autoregressive conditional skewness," *The Journal of Financial and Quantitative Analysis*, 34, 465–487.

He, C., and T. Teräsvirta (1999): "Fourth moment structure of the GARCH(p,q) process," *Econometric Theory*, 15, 824–846.

Hentschel, L. (1995): "All in the family: Nesting symmetric and asymmetric GARCH models," *Journal of Financial Economics*, 39, 71–104.

Hong, E. P. (1991): "The autocorrelation structure for the GARCH–M process," *Economics Letters*, 37, 129–132.

Kim, T.-H., and H. White (2004): "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, 1, 56–73.

Lambert, P., and S. Laurent (2002): "Modelling skewness dynamics in series of financial data using skewed location-scale distributions," Discussion Paper, Institut de Statistique, Louvain-la-Neuve.

Lanne, M., and P. Saikkonen (2005): "Modeling conditional skewness in stock returns," unpublished manuscript.

Ling, S., and M. McAleer (2002): "Stationarity and the existence of moments of a family of GARCH processes," *Journal of Econometrics*, 106, 109–117.

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

Peiró, A. (2002): "Skewness in individual stocks at different investment horizons," *Quantitative Finance*, 2, 139–146.

——— (2004): "Asymmetries and tails in stock index returns: are their distributions really asymmetric?," *Quantitative Finance*, 4, 37–44.

Schwert, G. W. (1989): "Why does stock market volatility change over time?," *Journal of Finance*, 44, 1115–1153.

Sentana, E. (1995): "Quadratic ARCH models," *The Review of Economic Studies*, 62, 639–661.

Taylor, S. J. (1986): *Modelling Financial Time Series*. Wiley, Chichester.

Wecker, W. E. (1981): "Asymmetric time series," *Journal of the American Statistical Association*, 76, 16–21.

Zakoïan, J.-M. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.

Multivariate GARCH models

Multivariate GARCH models

Abstract

This chapter contains a review of multivariate GARCH models. Most common GARCH models are presented and their properties considered. This also includes nonparametric and semiparametric models. Existing specification and misspecification tests are discussed. Finally, there is an empirical example in which several multivariate GARCH models are fitted to the same data set and the results compared.

This paper is joint work with Timo Teräsvirta.

We thank Robert Engle for providing the data for the empirical section, Markku Lanne and Pentti Saikkonen for sharing with us their program code for their Generalized Orthogonal Factor GARCH model, and Mika Meitz for programming assistance. The responsibility for any errors and shortcomings in this paper remains ours.

1 Introduction

Modelling volatility in financial time series has been the object of much attention ever since the introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) model in the seminal paper of Engle (1982). Subsequently, numerous variants and extensions of ARCH models have been proposed. A large body of this literature has been devoted to univariate models; see for example Bollerslev, Engle, and Nelson (1994), Palm (1996), and Shephard (1996).

While modelling volatility of the returns has been the main centre of attention, understanding the comovements of financial returns is of great practical importance. It is therefore important to extend the considerations to multivariate GARCH (MGARCH) models. For example, asset pricing depends on the covariance of the assets in a portfolio, and risk management and asset allocation relate for instance to finding and updating optimal hedging positions. For examples, see Bollerslev, Engle, and Wooldridge (1988), Ng (1991), and Hansson and Hördahl (1998). Multivariate GARCH models have also been used to investigate volatility and correlation transmission and spillover effects in studies of contagion, see Tse and Tsui (2002) and Bae, Karolyi, and Stulz (2003).

What then should the specification of an MGARCH model be like? On one hand, it should be flexible enough to be able to represent the dynamics of the conditional variances and covariances. On the other hand, as the number of parameters in an MGARCH model often increases rapidly with the dimension of the model, the specification should be parsimonious enough to allow for relatively easy estimation of the model and also allow for easy interpretation of the model parameters. However, parsimony often means simplification, and models with only a few parameters may not be able to capture the relevant dynamics in the covariance structure. Another feature that needs to be taken into account in the specification is imposing positive definiteness (as covariance matrices need, by definition, to be positive definite). One possibility is to derive conditions under which the conditional covariance matrices implied by the model are positive definite, but this is often infeasible in practice. An alternative is to formulate the model in a way that positive definiteness is implied by the model structure (in addition to some simple constraints).

Combining these needs has been the difficulty in the MGARCH literature. The first GARCH model for the conditional covariance matrices was the so-called VEC model of Bollerslev, Engle, and Wooldridge (1988), see Engle, Granger, and Kraft (1984) for an ARCH version. This model is a very general one, and a goal of the subsequent literature has been to formulate more parsimonious models. Furthermore, since imposing positive definiteness of the conditional covariance matrix in this model is difficult, formulating models with this feature has been considered important. Furthermore, constructing models in which the estimated parameters have direct interpretation has been viewed as beneficial.

In this paper, we survey the main developments of the MGARCH literature. For another such survey, see Bauwens, Laurent, and Rombouts (2006). This paper is organized as follows. In Section 2, several MGARCH specifications are reviewed. Statistical properties of the models are the topic of Section 3, whereas testing MGARCH

models is discussed in Section 4. An empirical comparison of a selection of the models is given in Section 5. Finally, some conclusions and directions for future research are in Section 6.

2 Models

Consider a stochastic vector process $\{\boldsymbol{r}_t\}$ with dimension $N \times 1$ such that $E\boldsymbol{r}_t = \boldsymbol{0}$. Let $\mathcal{F}_{t-1}$ denote the information set generated by the observed series $\{\boldsymbol{r}_t\}$ up to and including time $t-1$. We assume that $\boldsymbol{r}_t$ is conditionally heteroskedastic:

$$\boldsymbol{r}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t \tag{1}$$

given the information set $\mathcal{F}_{t-1}$, where the $N \times N$ matrix $\boldsymbol{H}_t = [h_{ijt}]$ is the conditional covariance matrix of $\boldsymbol{r}_t$ and $\boldsymbol{\eta}_t$ is an iid vector error process such that $E\boldsymbol{\eta}_t\boldsymbol{\eta}_t' = \boldsymbol{I}$. This defines the standard multivariate GARCH framework, in which there is no linear dependence structure in $\{\boldsymbol{r}_t\}$.

What remains to be specified is the matrix process $\boldsymbol{H}_t$. Various parametric formulations will be reviewed in the following subsections. We have divided these models into four categories. In the first one, the conditional covariance matrix $\boldsymbol{H}_t$ is modelled directly. This class includes, in particular, the VEC and BEKK models that were among the first parametric MGARCH models. The models in the second class, the factor models, are motivated by parsimony: the process $\boldsymbol{r}_t$ is assumed to be generated by a (small) number of unobserved heteroskedastic factors. Models in the third class are built on the idea of modelling the conditional variances and correlations instead of straightforward modelling of the conditional covariance matrix. Members of this class include the Constant Conditional Correlation (CCC) model and its extensions. The appeal of this class lies in the intuitive interpretation of the correlations, and models belonging to it have received plenty of attention in the recent literature. Finally, we consider semi- and nonparametric approaches that can offset the loss of efficiency of the parametric estimators due to misspecified structure of the conditional covariance matrices. Multivariate stochastic volatility models are discussed in a separate chapter, see Chib (2007).

Before turning to the models, we discuss some points that need attention when specifying an MGARCH model. As already mentioned, a problem with MGARCH models is that the number of parameters can increase very rapidly as the dimension of $\boldsymbol{r}_t$ increases. This creates difficulties in the estimation of the models, and therefore an important goal in constructing new MGARCH models is to make them reasonably parsimonious while maintaining flexibility. Another aspect that has to be imposed is the positive definiteness of the conditional covariance matrices. Ensuring positive definiteness of a matrix, usually through an eigenvalue-eigenvector-decomposition, is a numerically difficult problem, especially in large systems. Yet another difficulty with MGARCH models has to do with the numerical optimization of the likelihood function (in the case of parametric models). The conditional covariance (or correlation) matrix appearing in the likelihood depends on the time index t, and often has to be inverted

for all t in every iteration of the numerical optimization. When the dimension of the problem increases, this is a both time consuming and numerically unstable procedure. Avoiding excessive inversion of matrices is thus a worthy goal in designing MGARCH models. It should be emphasized, however, that practical implementation of all the models to be considered in this chapter is of course feasible, but the problem lies in devising easy to use, automated estimation routines that would make widespread use of these models possible.

2.1 Models of the conditional covariance matrix

The VEC-GARCH model of Bollerslev, Engle, and Wooldridge (1988) is a straightforward generalization of the univariate GARCH model. Every conditional variance and covariance is a function of all lagged conditional variances and covariances, as well as lagged squared returns and cross-products of returns. The model may be written as follows:

$$\mathrm{vech}(\boldsymbol{H}_t) = \boldsymbol{c} + \sum_{j=1}^{q} \boldsymbol{A}_j \mathrm{vech}(\boldsymbol{r}_{t-j}\boldsymbol{r}'_{t-j}) + \sum_{j=1}^{p} \boldsymbol{B}_j \mathrm{vech}(\boldsymbol{H}_{t-j}) \tag{2}$$

where $\mathrm{vech}(\cdot)$ is an operator that stacks the columns of the lower triangular part of its argument square matrix, $\boldsymbol{c}$ is an $N(N+1)/2 \times 1$ vector, and $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ are $N(N+1)/2 \times N(N+1)/2$ parameter matrices. In fact, the authors introduced a multivariate GARCH–in-mean model, but in this chapter we only consider its conditional covariance component. The generality of the VEC model has disadvantages. One is that there exist only sufficient, rather restrictive, conditions for $\boldsymbol{H}_t$ to be positive definite for all t, see Gouriéroux (1997, Section 6). Besides, the number of parameters equals $(p+q)(N(N+1)/2)^2 + N(N+1)/2$, which is large unless N is small. Furthermore, as will be discussed below, estimation of the parameters is computationally demanding.

Bollerslev, Engle, and Wooldridge (1988) present a simplified version of the model by assuming that $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ in (2) are diagonal matrices. In this case, it is possible to obtain conditions for $\boldsymbol{H}_t$ to be positive definite for all t, see Bollerslev, Engle, and Nelson (1994). Estimation is less difficult than in the complete VEC model because each equation can be estimated separately. But then, this "diagonal VEC" model that contains $(p+q+1)N(N+1)/2$ parameters seems too restrictive since no interaction is allowed between the different conditional variances and covariances.

A numerical problem is that estimation of parameters of the VEC model is computationally demanding. Assuming that the errors $\boldsymbol{\eta}_t$ follow a multivariate normal distribution, the log-likelihood of the model (1) has the following form:

$$\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) = c - (1/2)\sum_{t=1}^{T} \ln|\boldsymbol{H}_t| - (1/2)\sum_{t=1}^{T} \boldsymbol{r}'_t \boldsymbol{H}_t^{-1} \boldsymbol{r}_t. \tag{3}$$

The parameter vector $\boldsymbol{\theta}$ has to be estimated iteratively. It is seen from (3) that the conditional covariance matrix $\boldsymbol{H}_t$ has to be inverted for every t in each iteration, which

may be tedious when the number of observations is large. Another, an even more difficult problem, is how to ensure positive definiteness of the covariance matrices. In the case of the VEC model there does not seem to exist a general solution to this problem.

A model that can be viewed as a restricted version of the VEC model is the Baba-Engle-Kraft-Kroner (BEKK) defined in Engle and Kroner (1995). It has the attractive property that the conditional covariance matrices are positive definite by construction. The model has the form

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \sum_{j=1}^{q}\sum_{k=1}^{K} \boldsymbol{A}_{kj}'\boldsymbol{r}_{t-j}\boldsymbol{r}_{t-j}'\boldsymbol{A}_{kj} + \sum_{j=1}^{p}\sum_{k=1}^{K} \boldsymbol{B}_{kj}'\boldsymbol{H}_{t-j}\boldsymbol{B}_{kj} \tag{4}$$

where $\boldsymbol{A}_{kj}$, $\boldsymbol{B}_{kj}$, and $\boldsymbol{C}$ are $N \times N$ parameter matrices, and $\boldsymbol{C}$ is lower triangular. The decomposition of the constant term into a product of two triangular matrices is to ensure positive definiteness of $\boldsymbol{H}_t$. The BEKK model is covariance stationary if and only if the eigenvalues of $\sum_{j=1}^{q}\sum_{k=1}^{K} \boldsymbol{A}_{kj} \otimes \boldsymbol{A}_{kj} + \sum_{j=1}^{p}\sum_{k=1}^{K} \boldsymbol{B}_{kj} \otimes \boldsymbol{B}_{kj}$, where $\otimes$ denotes the Kronecker product of two matrices, are less than one in modulus. Whenever $K > 1$ an identification problem arises because there are several parameterizations that yield the same representation of the model. Engle and Kroner (1995) give conditions for eliminating redundant, observationally equivalent representations.

Interpretation of parameters of (4) is not easy. But then, consider the first order model

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{A}'\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}'\boldsymbol{A} + \boldsymbol{B}'\boldsymbol{H}_{t-1}\boldsymbol{B}. \tag{5}$$

Setting $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{D}$ where $\boldsymbol{D}$ is a diagonal matrix, (5) becomes

$$\boldsymbol{H}_t = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{A}'\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}'\boldsymbol{A} + \boldsymbol{D}E[\boldsymbol{A}'\boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}'\boldsymbol{A}|\mathcal{F}_{t-2}]\boldsymbol{D}. \tag{6}$$

It is seen from (6) that what is now modelled are the conditional variances and covariances of certain linear combinations of $\boldsymbol{r}_t$ or "portfolios". Kroner and Ng (1998) restrict $\boldsymbol{B} = \delta\boldsymbol{A}$ where $\delta > 0$ is a scalar.

A further simplified version of (5) in which $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal matrices has sometimes appeared in applications. This "diagonal BEKK" model trivially satisfies the equation $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{D}$. It is a restricted version of the diagonal VEC model such that the parameters of the covariance equations (equations for h_{ijt}, $i \neq j$) are products of the parameters of the variance equations (equations for h_{iit}). In order to obtain a more general model (that is, to relax these restrictions on the coefficients of the covariance terms) one has to allow $K > 1$. The most restricted version of the diagonal BEKK model is the scalar BEKK one with $\boldsymbol{A} = a\boldsymbol{I}$ and $\boldsymbol{B} = b\boldsymbol{I}$ where a and b are scalars.

Each of the BEKK models implies a unique VEC model, which then generates positive definite conditional covariance matrices. Engle and Kroner (1995) provide sufficient conditions for the two models, BEKK and VEC, to be equivalent. They also give a representation theorem that establishes the equivalence of diagonal VEC models (that have positive definite covariance matrices) and general diagonal BEKK models. When the number of parameters in the BEKK model is less than the corresponding number in the VEC model, the BEKK parameterization imposes restrictions

that makes the model different from that of VEC model. Increasing K in (4) eliminates those restrictions and thus increases the generality of the BEKK model towards the obtained from using pure VEC model. Engle and Kroner (1995) give necessary conditions under which all unnecessary restrictions are eliminated. However, too large a value of K will give rise to the identification problem mentioned earlier.

Estimation of a BEKK model still involves somewhat heavy computations due to several matrix inversions. The number of parameters, $(p+q)KN^2 + N(N+1)/2$ in the full BEKK model, or $(p+q)KN + N(N+1)/2$ in the diagonal one, is still quite large. Obtaining convergence may therefore be difficult because (4) is not linear in parameters. There is the advantage, however, that the structure automatically ensures positive definiteness of $\boldsymbol{H}_t$, so this does not need to be imposed separately. Partly because numerical difficulties are so common in the estimation of BEKK models, it is typically assumed $p = q = K = 1$ in applications of (4).

Parameter restrictions to ensure positive definiteness are not needed in the matrix exponential GARCH model proposed by Kawakatsu (2006). It is a generalization of the univariate exponential GARCH model (Nelson, 1991) and is defined as follows:

$$\text{vech}(\ln \boldsymbol{H}_t - \boldsymbol{C}) = \sum_{i=i}^{q} \boldsymbol{A}_i \boldsymbol{\eta}_{t-i} + \sum_{i=1}^{q} \boldsymbol{F}_i (|\boldsymbol{\eta}_{t-i}| - E|\boldsymbol{\eta}_{t-i}|) + \sum_{i=1}^{p} \boldsymbol{B}_i \text{vech}(\ln \boldsymbol{H}_{t-i} - \boldsymbol{C}) \quad (7)$$

where $\boldsymbol{C}$ is a symmetric $N \times N$ matrix, and $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, and $\boldsymbol{F}_i$ are parameter matrices of sizes $N(N+1)/2 \times N$, $N(N+1)/2 \times N(N+1)/2$, and $N(N+1)/2 \times N$, respectively. There is no need to impose restrictions on the parameters to ensure positive definiteness, because the matrix $\ln \boldsymbol{H}_t$ need not be positive definite. The positive definiteness of the covariance matrix $\boldsymbol{H}_t$ follows from the fact that for any symmetric matrix $\boldsymbol{S}$, the matrix exponential defined as

$$\exp(\boldsymbol{S}) = \sum_{i=0}^{\infty} \frac{\boldsymbol{S}^i}{i!}$$

is positive definite. Since the model contains a large number of parameters, Kawakatsu (2006) discusses a number of more parsimonious specifications. He also considers the estimation of the model, hypothesis testing, the interpretation of the parameters, and provides an application. How popular this model will turn out in practice remains to be seen.

2.2 Factor models

Factor models are motivated by economic theory. For instance, in the arbitrage pricing theory of Ross (1976) returns are generated by a number of common unobserved components, or factors; for further discussion see Engle, Ng, and Rothschild (1990) who introduced the first factor GARCH model. In this model it is assumed that the observations are generated by underlying factors that are conditionally heteroskedastic and possess a GARCH-type structure. This approach has the advantage that it

reduces the dimensionality of the problem when the number of factors relative to the dimension of $\boldsymbol{r}_t$ is small.

Engle, Ng, and Rothschild (1990) define a factor structure for the conditional covariance matrix as follows. They assume that $\boldsymbol{H}_t$ is generated by K ($< N$) underlying, not necessarily uncorrelated, factors $f_{k,t}$ as follows:

$$\boldsymbol{H}_t = \boldsymbol{\Omega} + \sum_{k=1}^{K} \boldsymbol{w}_k \boldsymbol{w}_k' f_{k,t} \tag{8}$$

where $\boldsymbol{\Omega}$ is an $N \times N$ positive semi-definite matrix, $\boldsymbol{w}_k$, $k = 1, \ldots, K$, are linearly independent $N \times 1$ vectors of factor weights, and the $f_{k,t}$'s are the factors. It is assumed that these factors have a first-order GARCH structure:

$$f_{k,t} = \omega_k + \alpha_k (\boldsymbol{\gamma}_k' \boldsymbol{r}_{t-1})^2 + \beta_k f_{k,t-1}$$

where ω_k, α_k, and β_k are scalars and $\boldsymbol{\gamma}_k$ is an $N \times 1$ vector of weights. The number of factors K is intended to be much smaller than the number of assets N, which makes the model feasible even for a large number of assets. Consistent but not efficient two-step estimation method using maximum likelihood is discussed in Engle, Ng, and Rothschild (1990). In their application, the authors consider two factor-representing portfolios as the underlying factors that drive the volatilities of excess returns of the individual assets. One factor consists of value-weighted stock index returns and the other one of average T-bill returns of different maturities. This choice is motivated by principal component analysis.

Diebold and Nerlove (1989) propose a model similar to the one Engle, Ng, and Rothschild (1990). However their model is rather a stochastic volatility model than a GARCH one, and hence we do not discuss its properties here; see Sentana (1998) for a comparison of this model with the factor GARCH one.

In the factor ARCH model of Engle, Ng, and Rothschild (1990) the factors are generally correlated. This may be undesirable as it may turn out that several of the factors capture very similar characteristics of the data. If the factors were uncorrelated, they would represent genuinely different common components driving the returns. Motivated by this consideration, several factor models with uncorrelated factors have been proposed in the literature. In all of them, the original observed series contained in $\boldsymbol{r}_t$ are assumed to be linked to unobserved, uncorrelated variables, or factors, $\boldsymbol{z}_t$ through a linear, invertible transformation $\boldsymbol{W}$:

$$\boldsymbol{r}_t = \boldsymbol{W} \boldsymbol{z}_t$$

where $\boldsymbol{W}$ is thus a nonsingular $N \times N$ matrix. Use of uncorrelated factors can potentially reduce their number relative to the approach where the factors can be correlated. The unobservable factors are estimated from the data through $\boldsymbol{W}$. The factors $\boldsymbol{z}_t$ are typically assumed to follow a GARCH process. Differences between the factor models are due to the specification of the transformation $\boldsymbol{W}$ and, importantly, whether the number of heteroskedastic factors is less than the number of assets or not.

In the Generalized Orthogonal (GO–) GARCH model of van der Weide (2002), the uncorrelated factors $\boldsymbol{z}_t$ are standardized to have unit unconditional variances, that is, $E\boldsymbol{z}_t\boldsymbol{z}_t' = \boldsymbol{I}$. This specification extends the Orthogonal (O–) GARCH model of Alexander and Chibumba (1997) in that $\boldsymbol{W}$ is not required to be orthogonal, only invertible. The factors are conditionally heteroskedastic with GARCH-type dynamics. The $N \times N$ diagonal matrix of conditional variances of $\boldsymbol{z}_t$ is defined as follows:

$$\boldsymbol{H}_t^z = (\boldsymbol{I} - \boldsymbol{A} - \boldsymbol{B}) + \boldsymbol{A} \odot (\boldsymbol{z}_{t-1}\boldsymbol{z}_{t-1}') + \boldsymbol{B}\boldsymbol{H}_{t-1}^z \tag{9}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are diagonal $N \times N$ parameter matrices and $\odot$ denotes the Hadamard (i.e. elementwise) product of two conformable matrices. The form of the constant term imposes the restriction $E\boldsymbol{z}_t\boldsymbol{z}_t' = \boldsymbol{I}$. Covariance stationarity of $\boldsymbol{r}_t$ in the models with uncorrelated factors is ensured if the diagonal elements of $\boldsymbol{A} + \boldsymbol{B}$ are less than one. Therefore the conditional covariance matrix of $\boldsymbol{r}_t$ can be expressed as

$$\boldsymbol{H}_t = \boldsymbol{W}\boldsymbol{H}_t^z\boldsymbol{W}' = \sum_{k=1}^{N} \boldsymbol{w}_{(k)}\boldsymbol{w}_{(k)}' h_{k,t}^z \tag{10}$$

where $\boldsymbol{w}_{(k)}$ are the columns of the matrix $\boldsymbol{W}$ and $h_{k,t}^z$ are the diagonal elements of the matrix $\boldsymbol{H}_t^z$. The difference between equations (8) and (10) is that the factors in (10) are uncorrelated but then, in the GO–GARCH model it is not possible to have fewer factors than there are assets. This is possible in the O–GARCH model but at the cost of obtaining conditional covariance matrices with a reduced rank.

Van der Weide (2002) constructs the linear mapping $\boldsymbol{W}$ by making use of the singular value decomposition of $E\boldsymbol{r}_t\boldsymbol{r}_t' = \boldsymbol{W}\boldsymbol{W}'$. That is,

$$\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}$$

where the columns of $\boldsymbol{U}$ hold the eigenvectors of $E\boldsymbol{r}_t\boldsymbol{r}_t'$ and the diagonal matrix $\boldsymbol{\Lambda}$ holds its eigenvalues, thus exploiting unconditional information only. Estimation of the orthogonal matrix $\boldsymbol{V}$ requires use of conditional information; see van der Weide (2002) for details.

Vrontos, Dellaportas, and Politis (2003) have suggested a related model. They state their Full Factor (FF–) GARCH model as above but restrict the mapping $\boldsymbol{W}$ to be an $N \times N$ invertible triangular parameter matrix with ones on the main diagonal. Furthermore, the parameters in $\boldsymbol{W}$ are estimated directly using conditional information only. Assuming $\boldsymbol{W}$ to be triangular simplifies matters but is restrictive because, depending on the order of the components in the vector $\boldsymbol{r}_t$, certain relationships between the factors and the returns are ruled out.

Lanne and Saikkonen (in press) put forth yet another modelling proposal. In their Generalized Orthogonal Factor (GOF–) GARCH model the mapping $\boldsymbol{W}$ is decomposed using the polar decomposition:

$$\boldsymbol{W} = \boldsymbol{C}\boldsymbol{V}$$

where $\boldsymbol{C}$ is a symmetric positive definite $N \times N$ matrix and $\boldsymbol{V}$ an orthogonal $N \times N$ matrix. Since $E\boldsymbol{r}_t\boldsymbol{r}_t' = \boldsymbol{W}\boldsymbol{W}' = \boldsymbol{C}\boldsymbol{C}'$, the matrix $\boldsymbol{C}$ can be estimated making use of the

spectral decomposition $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}'$ where the columns of $\boldsymbol{U}$ are the eigenvectors of $E\boldsymbol{r}_t\boldsymbol{r}_t'$ and the diagonal matrix $\boldsymbol{\Lambda}$ contains its eigenvalues, thus using unconditional information only. Estimation of $\boldsymbol{V}$ requires the use of conditional information, see Lanne and Saikkonen (in press) for details.

An important aspect of the GOF–GARCH model is that some of the factors can be conditionally homoskedastic. In addition to being parsimonious, this allows the model to include not only systematic but also idiosyncratic components of risk. Suppose K ($\leq N$) of the factors are heteroskedastic, while the remaining $N-K$ factors are homoskedastic. Without loss of generality we can assume that the K first elements of $\boldsymbol{z}_t$ are the heteroskedastic ones, in which case this restriction is imposed by setting that the $N-K$ last diagonal elements of $\boldsymbol{A}$ and $\boldsymbol{B}$ in (9) equal to zero. This results in the conditional covariance matrix of $\boldsymbol{r}_t$ of the following form (ref. eq. (10)):

$$\begin{aligned} \boldsymbol{H}_t &= \sum_{k=1}^{K} \boldsymbol{w}_{(k)}\boldsymbol{w}_{(k)}' h_{k,t}^z + \sum_{k=K+1}^{N} \boldsymbol{w}_{(k)}\boldsymbol{w}_{(k)}' \\ &= \sum_{k=1}^{K} \boldsymbol{w}_{(k)}\boldsymbol{w}_{(k)}' h_{k,t}^z + \boldsymbol{\Omega}. \end{aligned} \tag{11}$$

The expression (11) is very similar to the one in (8), but there are two important differences. In (11) the factors are uncorrelated, whereas in (8), as already pointed out, this is not generally the case. The role of $\boldsymbol{\Omega}$ in (11) is also different from that of $\boldsymbol{\Omega}$ in (8). In the factor ARCH model $\boldsymbol{\Omega}$ is required to be a positive semi-definite matrix and it has no particular interpretation. For comparison, the matrix $\boldsymbol{\Omega}$ in the GOF–GARCH model has a reduced rank directly related to the number of heteroskedastic factors. Furthermore, it is closely related to the unconditional covariance matrix of $\boldsymbol{r}_t$. This results to the model being possibly considerably more parsimonious than the factor ARCH model; for details and a more elaborate discussion, see Lanne and Saikkonen (in press). Therefore, the GOF–GARCH model can be seen as combining the advantages of both the factor models (having a reduced number of heteroskedastic factors) and the orthogonal models (relative ease of estimation due to the orthogonality of factors).

2.3 Models of conditional variances and correlations

Correlation models are based on the decomposition of the conditional covariance matrix into conditional standard deviations and correlations. The simplest multivariate correlation model that is nested in the other conditional correlation models, is the Constant Conditional Correlation (CCC–) GARCH model of Bollerslev (1990). In this model, the conditional correlation matrix is time-invariant, so the conditional covariance matrix can be expressed as follows:

$$\boldsymbol{H}_t = \boldsymbol{D}_t\boldsymbol{P}\boldsymbol{D}_t \tag{12}$$

where $\boldsymbol{D}_t = \mathrm{diag}(h_{1t}^{1/2}, \ldots, h_{Nt}^{1/2})$ and $\boldsymbol{P} = [\rho_{ij}]$ is positive definite with $\rho_{ii} = 1$, $i = 1, \ldots, N$. This means that the off-diagonal elements of the conditional covariance matrix are defined through the correlations of z_{it} and z_{jt}:

$$[\boldsymbol{H}_t]_{ij} = h_{it}^{1/2} h_{jt}^{1/2} \rho_{ij}, \quad i \neq j$$

where $1 \leq i, j \leq N$. The models for the processes $\{r_{it}\}$ are members of the class of univariate GARCH models. They are most often modelled as the GARCH(p,q) model, in which case the conditional variances can be written in a vector form

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \sum_{j=1}^{q} \boldsymbol{A}_j \boldsymbol{r}_{t-j}^{(2)} + \sum_{j=1}^{p} \boldsymbol{B}_j \boldsymbol{h}_{t-j} \tag{13}$$

where $\boldsymbol{\omega}$ is $N \times 1$ vector, $\boldsymbol{A_j}$ and $\boldsymbol{B_j}$ are diagonal $N \times N$ matrices, and $\boldsymbol{r}_t^{(2)} = \boldsymbol{r}_t \odot \boldsymbol{r}_t$. When the conditional correlation matrix $\boldsymbol{P}$ is positive definite and the elements of $\boldsymbol{\omega}$ and the diagonal elements of $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ positive, the conditional covariance matrix $\boldsymbol{H}_t$ is positive definite. Positivity of the diagonal elements of $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ is not, however, necessary for $\boldsymbol{P}$ to be positive definite unless $p = q = 1$, see Nelson and Cao (1992) for discussion of positivity conditions for h_{it} in univariate GARCH(p,q) models.

An extension to the CCC–GARCH model was introduced by Jeantheau (1998). In this Extended CCC– (ECCC–) GARCH model the assumption that the matrices $\boldsymbol{A}_j$ and $\boldsymbol{B}_j$ in (13) are diagonal is relaxed. This allows the past squared returns and variances of all series to enter the individual conditional variance equations. For instance, in the first-order ECCC–GARCH model, the ith variance equation is

$$h_{it} = \omega_i + a_{11} r_{1,t-1}^2 + \ldots + a_{1N} r_{N,t-1}^2 + b_{11} h_{1,t-1} + \ldots + b_{1N} h_{N,t-1}, \quad i = 1, \ldots, N.$$

An advantage of this extension is that it allows a considerably richer autocorrelation structure for the squared observed returns than the standard CCC–GARCH model. For example, in the univariate GARCH(1,1) model the autocorrelations of the squared observations decrease exponentially from the first lag. In the first-order ECCC–GARCH model, the same autocorrelations need not have a monotonic decline from the first lag. This has been shown by He and Teräsvirta (2004) who considered the fourth-moment structure of first- and second-order ECCC–GARCH models.

The estimation of MGARCH models with constant correlations is computationally attractive. Because of the decomposition (12), the log-likelihood in (3) has the following simple form:

$$\begin{aligned} \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) &= c - (1/2) \sum_{t=1}^{T} \sum_{i=1}^{N} \ln |h_{it}| - (1/2) \sum_{t=1}^{T} \log |\boldsymbol{P}| \\ &\quad -(1/2) \sum_{t=1}^{T} \boldsymbol{r}_t' \boldsymbol{D}_t^{-1} \boldsymbol{P}^{-1} \boldsymbol{D}_t^{-1} \boldsymbol{r}_t. \end{aligned} \tag{14}$$

From (14) it is apparent that during estimation, one has to invert the conditional correlation matrix only once per iteration. The number of parameters in the CCC–

and ECCC–GARCH models, in addition to the ones in the univariate GARCH equations, equals $N(N-1)/2$ and covariance stationarity is ensured if the roots of $\det(\boldsymbol{I} - \sum_{j=1}^{q} \boldsymbol{A}_j \lambda^j - \sum_{j=1}^{p} \boldsymbol{B}_j \lambda^j) = 0$ lie outside the unit circle.

Although the CCC–GARCH model is in many respects an attractive parameterization, empirical studies have suggested that the assumption of constant conditional correlations may be too restrictive. The model may therefore be generalized by retaining the previous decomposition but making the conditional correlation matrix in (12) time-varying. Thus,

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{P}_t \boldsymbol{D}_t. \tag{15}$$

In conditional correlation models defined through (15), positive definiteness of $\boldsymbol{H}_t$ follows if, in addition to the conditional variances h_{it}, $i = 1, \ldots, N$, being well-defined, the conditional correlation matrix $\boldsymbol{P}_t$ is positive definite at each point in time. Compared to the CCC–GARCH models, the advantage of numerically simple estimation is lost, as the correlation matrix has to be inverted for each t during every iteration.

Due to the intuitive interpretation of correlations, there exist a vast number of proposals for specifying $\boldsymbol{P}_t$. Tse and Tsui (2002) imposed GARCH type of dynamics on the conditional correlations. The conditional correlations in their Varying Correlation (VC–) GARCH model are functions of the conditional correlations of the previous period and a set of estimated correlations. More specifically,

$$\boldsymbol{P}_t = (1 - a - b)\boldsymbol{S} + a\boldsymbol{S}_{t-1} + b\boldsymbol{P}_{t-1}$$

where $\boldsymbol{S}$ is a constant, positive definite parameter matrix with ones on the diagonal, a and b are non-negative scalar parameters such that $a + b \leq 1$, and $\boldsymbol{S}_{t-1}$ is a sample correlation matrix of the past M standardized residuals $\hat{\boldsymbol{\varepsilon}}_{t-1}, \ldots, \hat{\boldsymbol{\varepsilon}}_{t-M}$ where $\hat{\boldsymbol{\varepsilon}}_{t-j} = \widehat{\boldsymbol{D}}_{t-j}^{-1} \boldsymbol{r}_{t-j}$, $j = 1, \ldots, M$. The positive definiteness of $\boldsymbol{P}_t$ is ensured by construction if $\boldsymbol{P}_0$ and $\boldsymbol{S}_{t-1}$ are positive definite. A necessary condition for the latter to hold is $M \geq N$. The definition of the "intercept" $1 - a - b$ corresponds to the idea of "variance targeting" in Engle and Mezrich (1996).

Kwan, Li, and Ng (2005) proposed a threshold extension to the VC–GARCH model. Within each regime, indicated by the value of an indicator or threshold variable, the model has a VC–GARCH specification. Specifically, the authors partition the real line into R subintervals, $r_0 = -\infty < l_1 < \ldots < l_{R-1} < l_R = \infty$, and define an indicator variable $s_t \in \mathcal{F}_{t-1}$. The rth regime is defined by $l_{r-1} < s_t \leq l_r$, and both the univariate GARCH models and the dynamic correlations have regime-specific parameters. Kwan, Li, and Ng (2005) also apply the same idea to the BEKK model and discuss estimation of the number of regimes. In order to estimate the model consistently, one has to make sure that each regime contains a sufficient number of observations.

Engle (2002) introduced a Dynamic Conditional Correlation (DCC–) GARCH model whose dynamic conditional correlation structure is similar to that of the VC–GARCH model. Engle considered a dynamic matrix process

$$\boldsymbol{Q}_t = (1 - a - b)\boldsymbol{S} + a\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} + b\boldsymbol{Q}_{t-1}$$

where a is a positive and b a non-negative scalar parameter such that $a + b < 1$, $\boldsymbol{S}$ is the unconditional correlation matrix of the standardized errors $\boldsymbol{\varepsilon}_t$, and $\boldsymbol{Q}_0$ is positive definite. This process ensures positive definiteness but does not generally produce valid correlation matrices. They are obtained by rescaling $\boldsymbol{Q}_t$ as follows:

$$\boldsymbol{P}_t = (\boldsymbol{I} \odot \boldsymbol{Q}_t)^{-1/2} \boldsymbol{Q}_t (\boldsymbol{I} \odot \boldsymbol{Q}_t)^{-1/2}$$

Both the VC– and the DCC–GARCH model extend the CCC–GARCH model but do it with few extra parameters. In each correlation equation, the number of parameters is $N(N-1)/2 + 2$ for the VC–GARCH model and two for in the DCC–GARCH one. This is a strength of these models but may also be seen as a weakness when N is large, because all $N(N-1)/2$ correlation processes are restricted to have the same dynamic structure.

In the VC–GARCH as well as the DCC–GARCH model, the dynamic structure of the time-varying correlations is a function of past returns. There is another class of models that allows the dynamic structure of the correlations to be controlled by an exogenous variable. This variable may be either an observable variable, a combination of observable variables, or a latent variable that represents factors that are difficult to quantify. One may argue that these models are not pure vector GARCH models because the conditioning set in them can be larger than in VC–GARCH or DCC–GARCH models. The first one of these models to be considered here is the Smooth Transition Conditional Correlation (STCC–) GARCH model.

In the STCC–GARCH model of Silvennoinen and Teräsvirta (2005) the conditional correlation matrix varies smoothly between two extreme states according to a transition variable. The following dynamic structure is imposed on the conditional correlations:

$$\boldsymbol{P}_t = (1 - G(s_t))\boldsymbol{P}_{(1)} + G(s_t)\boldsymbol{P}_{(2)}$$

where $\boldsymbol{P}_{(1)}$ and $\boldsymbol{P}_{(2)}$, $\boldsymbol{P}_{(1)} \neq \boldsymbol{P}_{(2)}$, are positive definite correlation matrices that describe the two extreme states of correlations, and $G(\cdot) : \mathbb{R} \to (0, 1)$, is a monotonic function of an observable transition variable $s_t \in \mathcal{F}^*_{t-1}$. The authors define $G(\cdot)$ as the logistic function

$$G(s_t) = \left(1 + e^{-\gamma(s_t - c)}\right)^{-1}, \quad \gamma > 0 \tag{16}$$

where the parameter γ determines the velocity and c the location of the transition. In addition to the univariate variance equations, the STCC–GARCH model has $N(N-1)+2$ parameters. The sequence $\{\boldsymbol{P}_t\}$ is a sequence of positive definite matrices because each $\boldsymbol{P}_t$ is a convex combination of two positive definite correlation matrices. The transition variable s_t is chosen by the modeller to suit the application at hand. If there is uncertainty about an appropriate choice of s_t, testing the CCC–GARCH model can be used as tool for judging the relevance of a given transition variable to the dynamic conditional correlations. A special case of the STCC–GARCH model is obtained when the transition variable is calendar time. The Time Varying Conditional Correlation (TVCC–) GARCH model was in its bivariate form introduced by Berben and Jansen (2005).

A recent extension of the STCC–GARCH model, the Double Smooth Transition Conditional Correlation (DSTCC–) GARCH model by Silvennoinen and Teräsvirta (2006) allows for another transition around the first one:

$$\begin{aligned} \boldsymbol{P}_t &= (1-G_2(s_{2t}))\left\{(1-G_1(s_{1t}))\boldsymbol{P}_{(11)}+G_1(s_{1t})\boldsymbol{P}_{(21)}\right\} \\ &\quad +G_2(s_{2t})\left\{(1-G_1(s_{1t}))\boldsymbol{P}_{(12)}+G_1(s_{1t})\boldsymbol{P}_{(22)}\right\} \end{aligned} \tag{17}$$

For instance, one of the transition variables can simply be calendar time. If this the case, one has the Time Varying Smooth Transition Conditional Correlation (TVSTCC–) GARCH model that nests the STCC–GARCH as well as the TVCC–GARCH model. The interpretation of the extreme states is the following: At the beginning of the sample, $\boldsymbol{P}_{(11)}$ and $\boldsymbol{P}_{(21)}$ are the two extreme states between which the correlations vary according to the transition variable s_{1t} and similarly, $\boldsymbol{P}_{(12)}$ and $\boldsymbol{P}_{(22)}$ are the corresponding states at the end of the sample. The TVSTCC–GARCH model allows the extreme states, constant in the STCC–GARCH framework, to be time-varying, which introduces extra flexibility when modelling long time series. The number of parameters, excluding the univariate GARCH equations, is $2N(N-1)+4$ which restricts the use of the model in very large systems.

The Regime Switching Dynamic Correlation (RSDC–) GARCH model introduced by Pelletier (2006) falls somewhere between the models with constant correlations and the ones with correlations changing continuously at every period. The model imposes constancy of correlations within a regime while the dynamics enter through switching regimes. Specifically,

$$\boldsymbol{P}_t = \sum_{r=1}^{R} \mathbb{I}_{\{\Delta_t = r\}} \boldsymbol{P}_{(r)}$$

where Δ_t is a (usually first-order) Markov chain independent of $\boldsymbol{\eta}_t$ that can take R possible values and is governed by a transition probability matrix $\boldsymbol{\Pi}$, $\mathbb{I}$ is the indicator function, and $\boldsymbol{P}_{(r)}$, $r = 1, \ldots, R$, are positive definite regime-specific correlation matrices. Correlation component of the model has $RN(N-1)/2 - R(R-1)$ parameters. A version that involves fewer parameters is obtained by restricting the R possible states of correlations to be linear combinations of a state of zero correlations and that of possibly high correlations. Thus,

$$\boldsymbol{P}_t = (1-\lambda(\Delta_t))\boldsymbol{I} + \lambda(\Delta_t)\boldsymbol{P}$$

where $\boldsymbol{I}$ is the identity matrix ("no correlations"), $\boldsymbol{P}$ is a correlation matrix representing the state of possibly high correlations, and $\lambda(\cdot) : \{1, \ldots, R\} \to [0, 1]$ is a monotonic function of Δ_t. The number of regimes R is not a parameter to be estimated. The conditional correlation matrices are positive definite at each point in time by construction both in the unrestricted and restricted version of the model.

2.4 Nonparametric and semiparametric approaches

Non- and semiparametric models form an alternative to parametric estimation of the conditional covariance structure. These approaches have the advantage of not imposing a particular (possibly misspecified) structure on the data. One advantage of at least a few fully parametric multivariate GARCH models is, however, that they offer an interpretation of the dynamic structure of the conditional covariance or correlation matrices. Another is that the quasi-maximum likelihood estimator is consistent when the errors are assumed multivariate normal. However, there may be considerable efficiency losses if the returns are not normally distributed. Semiparametric models combine the advantages of a parametric model in that they reach consistency and sustain the interpretability, and those of a nonparametric model which is robust against distributional misspecification.

One alternative is to specify a parametric model for the conditional covariance structure but estimate the error distribution nonparametrically, thereby attempting to offset the efficiency loss of the quasi-maximum likelihood estimator compared to the maximum likelihood estimator of the correctly specified model. In the semiparametric model of Hafner and Rombouts (in press) the data are generated by any particular parametric MGARCH model and the error distribution is unspecified but estimated nonparametrically. Their approach of leads to the log-likelihood

$$\sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) = c - (1/2) \sum_{t=1}^{T} \ln |\boldsymbol{H}_t| + \sum_{t=1}^{T} \ln g(\boldsymbol{H}_t^{-1/2} \boldsymbol{r}_t). \tag{18}$$

where $g(\cdot)$ is an unspecified density function of the standardized residuals $\boldsymbol{\eta}_t$ such that $E[\boldsymbol{\eta}_t] = \mathbf{0}$ and $E[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \boldsymbol{I}$. This model may be seen as a multivariate extension of the semiparametric GARCH model by Engle and González-Rivera (1991). A flexible error distribution blurs the line between the parametric structure and the distribution of the errors. For example, if the correlation structure of a semiparametric GARCH model is misspecified, a nonparametric error distribution may absorb some of the misspecification. The nonparametric method for estimating the density g is discussed in detail in Hafner and Rombouts (in press). They assume that g belongs to the class of spherical distributions. Even with this restriction their semiparametric estimator remains more efficient than the maximum likelihood estimator if the errors z_t are non-normal.

Long and Ullah (2005) introduce an approach a similar to the previous one in that the model is based on any parametric MGARCH model. After estimating a parametric model, the estimated standardized residuals $\widehat{\boldsymbol{\eta}}_t$ are extracted. When the model is not correctly specified, these residuals may have some structure in them, and Long and Ullah (2005) use nonparametric estimation to extract this information. This is done by estimating the conditional covariance matrix using the Nadaraya-Watson estimator

$$\boldsymbol{H}_t = \widehat{\boldsymbol{H}}_t^{1/2} \frac{\sum_{\tau=1}^{T} \widehat{\boldsymbol{\eta}}_\tau \widehat{\boldsymbol{\eta}}_\tau' K_h(s_\tau - s_t)}{\sum_{\tau=1}^{T} K_h(s_\tau - s_t)} \widehat{\boldsymbol{H}}_t^{1/2} \tag{19}$$

where $\widehat{\boldsymbol{H}}_t$ is the conditional covariance matrix estimated parametrically from an

MGARCH model, $s_t \in \mathcal{F}^*_{t-1}$ is a conditioning variable, $\hat{\boldsymbol{\varepsilon}}_t = \widehat{\boldsymbol{D}}_t^{-1}\boldsymbol{r}_t$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is the bandwidth parameter. Positive definiteness of $\widehat{\boldsymbol{H}}_t$ ensures positive definiteness of the semiparametric estimator $\boldsymbol{H}_t$.

In the Semi-Parametric Conditional Correlation (SPCC–) GARCH model of Hafner, van Dijk, and Franses (2005) the conditional variances are modelled parametrically by any choice of univariate GARCH model, where $\hat{\boldsymbol{\varepsilon}}_t = \widehat{\boldsymbol{D}}_t^{-1}\boldsymbol{r}_t$ is the vector consisting of the standardized residuals. The conditional correlations $\boldsymbol{P}_t$ are then estimated using a transformed Nadaraya-Watson estimator:

$$\boldsymbol{P}_t = (\boldsymbol{I} \odot \boldsymbol{Q}_t)^{-1/2}\boldsymbol{Q}_t(\boldsymbol{I} \odot \boldsymbol{Q}_t)^{-1/2}$$

where

$$\boldsymbol{Q}_t = \frac{\sum_{\tau=1}^{T} \hat{\boldsymbol{\varepsilon}}_\tau \hat{\boldsymbol{\varepsilon}}'_\tau K_h(s_\tau - s_t)}{\sum_{\tau=1}^{T} K_h(s_\tau - s_t)}. \tag{20}$$

In (20), $s_t \in \mathcal{F}^*_{t-1}$ is a conditioning variable, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is the bandwidth parameter.

Long and Ullah (2005) also suggest to estimating the covariance structure in a fully nonparametric fashion so that the model is not an MGARCH model but merely a parameter-free multivariate volatility model. The estimator of the conditional covariance matrix is

$$\boldsymbol{H}_t = \frac{\sum_{\tau=1}^{T} \boldsymbol{r}_\tau \boldsymbol{r}'_\tau K_h(s_\tau - s_t)}{\sum_{\tau=1}^{T} K_h(s_\tau - s_t)}$$

where s_t is a conditioning variable, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a kernel function, and h is the bandwidth parameter. This approach ensures positive definiteness of $\boldsymbol{H}_t$.

The choice of the kernel function is not important and it could be any probability density function, whereas the choice of the bandwidth parameter h is crucial, see for instance Pagan and Ullah (1999, Sections 2.4.2 and 2.7). Long and Ullah (2005) consider the choice of an optimal fixed bandwidth, whereas Hafner, van Dijk, and Franses (2005) discuss a way of choosing a dynamic bandwidth parameter such that the bandwidth is larger close to the tails of the marginal distribution of the conditioning variable s_t than it is in the mid-region of the distribution.

3 Statistical properties

Statistical properties of multivariate GARCH models are only partially known. For the development of statistical estimation and testing theory, it would be desirable to have conditions for strict stationarity and ergodicity of a multivariate GARCH process, as well as conditions for consistency and asymptotic normality of the quasi-maximum likelihood estimator. The results that are available establish these properties in special cases and sometimes under strong conditions.

Jeantheau (1998) considers the statistical properties and estimation theory of the ECCC–GARCH model he proposes. He provides sufficient conditions for the existence of a weakly stationary and ergodic solution, which is also strictly stationary.

This is done by assuming $E\boldsymbol{r}_t\boldsymbol{r}_t' < \infty$. It would be useful to have both a necessary and a sufficient condition for the existence of a strictly stationary solution, but this question remains open. Jeantheau (1998) also proves the strong consistency of the QML estimator for the ECCC–GARCH model. Ling and McAleer (2003) complement Jeantheau's results and also prove the asymptotic normality of the QMLE in the case of the ECCC–GARCH model. For the global asymptotic normality result, the existence of the sixth moment of $\boldsymbol{r}_t$ is required. The statistical properties of the second-order model are also investigated in He and Teräsvirta (2004), who provide sufficient conditions for the existence of fourth moments, and, furthermore, give expressions for the fourth moment as well as the autocorrelation function of squared observations as functions of the parameters.

Comte and Lieberman (2003) study the statistical properties of the BEKK model. Relying on a result in Boussama (1998), they give sufficient, but not necessary, conditions for strict stationarity and ergodicity. Applying Jeantheau's results, they provide conditions for the strong consistency of the QMLE. Furthermore, they also prove the asymptotic normality of the QMLE, for which they assume the existence of the eighth moment of $\boldsymbol{r}_t$. The fourth-moment structure of the BEKK and VEC models is investigated by Hafner (2003), who gives necessary and sufficient conditions for the existence of the fourth moments as well as provides expressions for them. These expressions are not functions of the parameters of the model. As the factor models listed in Section 2.2 are special cases of the BEKK model, the results of Comte and Lieberman (2003) and Hafner (2003) also apply to them.

4 Hypothesis testing

Testing the adequacy of estimated models is an important part of model building. Existing tests of multivariate GARCH models may be divided into two broad categories: general misspecification tests and specification tests. The purpose of the tests belonging to the former category is to check the adequacy of an estimated model. Specification tests are different in the sense that they are designed to test the model against a parametric extension. Such tests have been constructed for the CCC–GARCH model but obviously not for other models. We first review general misspecification tests.

4.1 General misspecification tests

Ling and Li (1997) derived a rather general misspecification test for multivariate GARCH models. It is available for many families of GARCH models. The test statistic has the following form:

$$Q(k) = T\boldsymbol{\gamma}_k'\widehat{\boldsymbol{\Omega}}_k^{-1}\boldsymbol{\gamma}_k \tag{21}$$

where $\boldsymbol{\gamma}_k = (\gamma_1, ..., \gamma_k)'$ with

$$\gamma_j = \frac{\sum_{t=j+1}^{T}(\boldsymbol{r}_t'\widehat{\boldsymbol{H}}_t^{-1}\boldsymbol{r}_t - N)(\boldsymbol{r}_{t-j}'\widehat{\boldsymbol{H}}_{t-j}^{-1}\boldsymbol{r}_{t-j} - N)}{\sum_{t=1}^{T}(\boldsymbol{r}_t'\widehat{\boldsymbol{H}}_t^{-1}\boldsymbol{r}_t - N)^2} \tag{22}$$

$j = 1, \ldots, k$, $\widehat{\boldsymbol{H}}_t$ has been estimated from a GARCH model, and $\widehat{\boldsymbol{\Omega}}_k$ is the estimated covariance matrix of $\boldsymbol{\gamma}_k$, see Ling and Li (1997) for details. Under the null hypothesis that the GARCH model is correctly specified, that is, $\boldsymbol{\eta}_t \sim \mathrm{IID}(\mathbf{0}, \boldsymbol{I})$, statistic (21) has an asymptotic χ^2 distribution with k degrees of freedom. Under H_0, $E\boldsymbol{r}_t'\boldsymbol{H}_t^{-1}\boldsymbol{r}_t = N$, expression (22) is the jth-order sample autocorrelation between $\mathbf{r}_t'\boldsymbol{H}_t^{-1}\boldsymbol{r}_t = \boldsymbol{\eta}_t'\boldsymbol{\eta}_t$ and $\boldsymbol{r}_{t-j}'\boldsymbol{H}_{t-j}^{-1}\boldsymbol{r}_{t-j} = \boldsymbol{\eta}_{t-j}'\boldsymbol{\eta}_{t-j}$. The test may thus be viewed as a generalization of the portmanteau test of Li and Mak (1994) for testing the adequacy of a univariate GARCH model. In fact, when $N = 1$, (21) collapses into the Li and Mak statistic. The McLeod and Li (1983) statistic (Ljung-Box statistic applied to squared residuals), frequently used for evaluating GARCH models, is valid neither in the univariate nor in the multivariate case, see Li and Mak (1994) for the univariate case.

A simulation study by Tse and Tsui (1999) indicates that the Ling and Li portmanteau statistic (22) often has low power. The authors show examples of situations in which a portmanteau test based on autocorrelations of pairs of individual standardized residuals performs better. The drawback of this statistic is, however, that its asymptotic null distribution is unknown, compare this with McLeod and Li (1983), and the statistic tends to be undersized. Each test is based only on a single pair of residuals.

Another generalization of univariate tests can be found in Kroner and Ng (1998). Their misspecification tests are suitable for any multivariate GARCH model. Let

$$\boldsymbol{G}_t = \boldsymbol{r}_t\boldsymbol{r}_t' - \widehat{\boldsymbol{H}}_t$$

where $\widehat{\boldsymbol{H}}_t$ has been estimated from a GARCH model. The elements of $\boldsymbol{G}_t = [g_{ijt}]$ are "generalized residuals". When the model is correctly specified, they form a matrix of martingale difference sequences with respect to the information set $\mathcal{F}_{t-1}$ that contains the past information until $t-1$. Thus any variable $x_s \in \mathcal{F}_{t-1}$ is uncorrelated with the elements of $\boldsymbol{G}_t$. Tests based on these misspecification indicators may now be constructed. This is done for each g_{ijt} separately. The suggested tests are generalizations of the sign-bias and size-bias tests of Engle and Ng (1993). The test statistics have an asymptotic χ^2 distribution with one degree of freedom when the null hypothesis is valid. If the dimension of the model is large and there are several misspecification indicators, the number of available tests may be very large.

Testing the adequacy of the CCC–GARCH model has been an object of interest since it was found that the assumption of constant correlations may sometimes be too restrictive in practice. Tse (2000) constructed an LM test of the CCC–GARCH model against the following alternative, $\boldsymbol{P}_t$, to constant correlations:

$$\boldsymbol{P}_t = \boldsymbol{P} + \boldsymbol{\Delta} \odot \boldsymbol{r}_{t-1}\boldsymbol{r}_{t-1}' \tag{23}$$

where $\boldsymbol{\Delta}$ is a symmetric parameter matrix with the main diagonal elements equal to zero. This means that the correlations are changing as functions of the previous

observations. The null hypothesis is $\mathrm{H}_0 : \boldsymbol{\Delta} = \mathbf{0}$ or, expressed as a vector equation, $\mathrm{vecl}(\boldsymbol{\Delta}) = 0$.[1] Equation (23) does not define a particular alternative to conditional correlations as $\boldsymbol{P}_t$ is not necessarily a positive definite matrix for every t. For this reason we interpret the test as a general misspecification test.

Bera and Kim (2002) present a test of a bivariate CCC–GARCH model against the alternative that the correlation coefficient is stochastic. The test is an Information Matrix test and as such an LM or score test. It is designed for a bivariate model, which restricts its usefulness in applications.

4.2 Tests for extensions of the CCC–GARCH model

The most popular extension of the CCC–GARCH model to-date is the DCC–GARCH model of Engle (2002). However, there does not seem to be any published work on developing tests of constancy of correlations directly against this model.

As discussed in Section 2.3, Silvennoinen and Teräsvirta (2005) extend the CCC–GARCH into a STCC–GARCH model in which the correlations fluctuate according to a transition variable. They construct an LM test for testing the constant correlation hypothesis against the smoothly changing correlations. Since the STCC–GARCH model is only identified when the correlations are changing, standard asymptotic theory is not valid. A good discussion of this problem can be found in Hansen (1996). The authors apply the technique in Luukkonen, Saikkonen, and Teräsvirta (1988) in order to circumvent the identification problem. The null hypothesis is $\gamma = 0$ in (16), and a linearization of the correlation matrix $\boldsymbol{P}_t$ by the first-order Taylor expansion of (16) yields

$$\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)} - s_t \boldsymbol{P}_{(2)}^*.$$

Under H_0, $\boldsymbol{P}_{(2)}^* = \mathbf{0}$ and the correlations are thus constant. The authors use this fact to build their LM-type test on the transformed null hypothesis H_0': $\mathrm{vecl}(\boldsymbol{P}_{(2)}^*) = \mathbf{0}$ (the diagonal elements of $\boldsymbol{P}_{(2)}^*$ equal zero by definition). When H_0' holds, the test statistic has an asymptotic χ^2 distribution with $N(N-1)/2$ degrees of freedom. The authors also derive tests for the constancy hypothesis under the assumption that some of the correlations remain constant also under the alternative. Silvennoinen and Teräsvirta (2006) extend the Taylor expansion based test to the situation where the STCC–GARCH model is the null model and the alternative the DSTCC–GARCH one. This test collapses into the test of the CCC–GARCH model against STCC–GARCH model when $G_1(s_{1t}) \equiv 1/2$ in (17).

5 An application

In this section we compare some of the multivariate GARCH models considered in previous sections by fitting them to the same data set. In order to keep the comparison

[1] The operator $\mathrm{vecl}(\cdot)$ stacks the columns of the strictly lower triangular part (excluding main diagonal elements) of its argument matrix.

transparent, we only consider bivariate models. Our observations are the daily returns of S&P 500 index and 10-year bond futures from January 1990 to August 2003. This data set has been analyzed by Engle and Colacito (2006). [2] There is no consensus in the literature about how stock and long term bond returns are related. Historically, the long-run correlations have been assumed constant, an assumption that has led to contradicting conclusions because evidence for both positive and negative correlation has been found over the years (short-run correlations have been found to be affected, among other things, by news announcements). From a theoretical point of view, the long-run correlation between the two should be state-dependent, driven by macroeconomic factors such as growth, inflation, and interest rates. The way the correlations respond to these factors may, however, change over time.

For this reason it is interesting to see what the correlations between the two asset returns obtained from the models are and how they fluctuate over time. The focus of reporting results will therefore be on conditional correlations implied by the estimated models, that is, the BEKK, GOF–, DCC–, DSTCC–, and SPCC–GARCH ones. In the last three models, the individual GARCH equations are simply symmetric first-order ones. The BEKK model is also of order one with $K = 1$. All computations have been performed using Ox, version 4.02, see Doornik (2002), and our own source code.

Estimation of the BEKK model turned out to be cumbersome. Convergence problems were encountered in numerical algorithms, but the iterations seemed to suggest diagonality of the coefficient matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. A diagonal BEKK model was eventually estimated without difficulty.

In the estimation of the GOF–GARCH model it is essential to obtain good initial estimates of the parameters; for details, see Lanne and Saikkonen (in press). Having done that, we experienced no difficulties in the estimation of this model with a single factor. Similarly, no convergence problems were encountered in the estimation of the DCC model of Engle (2002).

The DSTCC–GARCH model makes use of two transition variables. Because the DSTCC framework allows one to test for relevance of a variable, or variables, to the description of the dynamic structure of the correlations, we relied on the tests in Silvennoinen and Teräsvirta (2005, 2006), described in Section 4.2, to select relevant transition variables. Out of a multitude of variables, including both exogenous ones and variables constructed from the past observations, prices or returns, the Chicago Board Options Exchange volatility index (VIX) that represents the market expectation of 30-day volatility turned out to have the best performance. Calendar time seemed to be another obvious transition variable. As a result, the first-order TVSTCC–GARCH model was fitted to the bivariate data.

The semiparametric model of Hafner, van Dijk, and Franses (2005) also requires a choice of an indicator variable. Because the previous test results indicated that VIX is informative about the dynamics of the correlations, we chose VIX as the indicator variable. The SPCC–GARCH model was estimated using a standard kernel smoother

[2] The data set in Engle and Colacito (2006) begins in August 1988, but our sample starts from January 1990 because we also use the time series for a volatility index that is available only from that date onwards.

with an optimal fixed bandwidth, see Pagan and Ullah (1999, Sections 2.4.2 and 2.7) for discussion on the choice of constant bandwidth.

The estimated conditional correlations are presented in Figure 1, whereas Table 1 shows the sample correlation matrix of the estimated time-varying correlations. The correlations from the diagonal BEKK model and the DCC–GARCH model are very strongly positively correlated, which is also obvious from Figure 1. The second-highest correlation of correlations is the one between the SPCC–GARCH and the GOF–GARCH model. The time-varying correlations are mostly positive during the 1990's and negative after the turn of the century. In most models, correlations seem to fluctuate quite randomly, but the TVSTCC–GARCH model constitutes an exception. This is due to the fact that one of the transition variables is calendar time. Interestingly, in the beginning of the period the correlation between the S&P 500 and bond futures is only mildly affected by the expected volatility (VIX) and remains positive. Towards the end, not only does the correlation gradually turn negative, but expected volatility seems to affect it very strongly. Rapid fluctuations are a consequence of the fact that the transition function with VIX as the transition variable has quite a steep slope. After the turn of the century, high values of VIX generate strongly negative correlations.

Although the estimated models do not display fully identical correlations, the general message in them remains more or less the same. It is up to the user to select the model he wants to use in portfolio management and forfecasting. A way of comparing the models consists of inserting the estimated covariance matrices $\boldsymbol{H}_t$, $t = 1, \ldots, T$, into the Gaussian log-likelihood function (3) and calculate the maximum value of log-likelihood. These values for the estimated models appear in Table 1.

The models that are relatively easy to estimate seem to fit the data less well than the other models. The ones with a more complicated structure and, consequently, an estimation procedure that requires care, seem to attain higher likelihood values. However, the models do not make use of the same information set and, besides, they do not contain the same number of parameters. Taking this into account suggests the use of model selection criteria for assessing the performance of the models. Nevertheless, rankings by Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) are the same as the likelihood-based ranking; see Table 1. Note that in theory, rankings based on a model selection criterion favour the SPCC model. This is because no penalty is imposed on the nonparametric correlation estimates that improve the fit compared to constant correlations.

Nonnested testing as a means of comparison is hardly a realistic option here since the computational effort would be quite substantial.

6 Final remarks

In this review, we have considered a number of multivariate GARCH models and highlighted their features. It is obvious that the original VEC model contains too many parameters to be easily applicable, and research has been concentrated on

	diag BEKK	GOF	DCC	DSTCC	SPCC
diag BEKK	1.0000				
GOF	0.7713	1.0000			
DCC	0.9875	0.7295	1.0000		
TVSTCC	0.7577	0.7381	0.7690	1.0000	
SPCC	0.6010	0.8318	0.5811	0.7374	1.0000
log-likelihood	-6130	-6091	-6166	-6006	-6054
AIC	12275	12198	12347	12041	12120
BIC	12286	12211	12359	12062	12130

Table 1: Sample correlations of the estimated conditional correlations. The lower part of the table shows the log-likelihood values and the values of the corresponding model selection criteria.

finding parsimonious alternatives to it. Two lines of development are visible. First, there are the attempts to impose restrictions on the parameters of the VEC model. The BEKK model and the factor models are examples of this. Second, there is the idea of modelling conditional covariances through conditional variances and correlations. It has led to a number of new models, and this family of conditional correlation models appears to be quite popular right now.

As previously discussed, there is no statistical theory covering all MGARCH models. This may be expected, since models in the two main categories differ substantially from each other. Progress has been made in some special occasions, and these cases have been considered in previous sections.

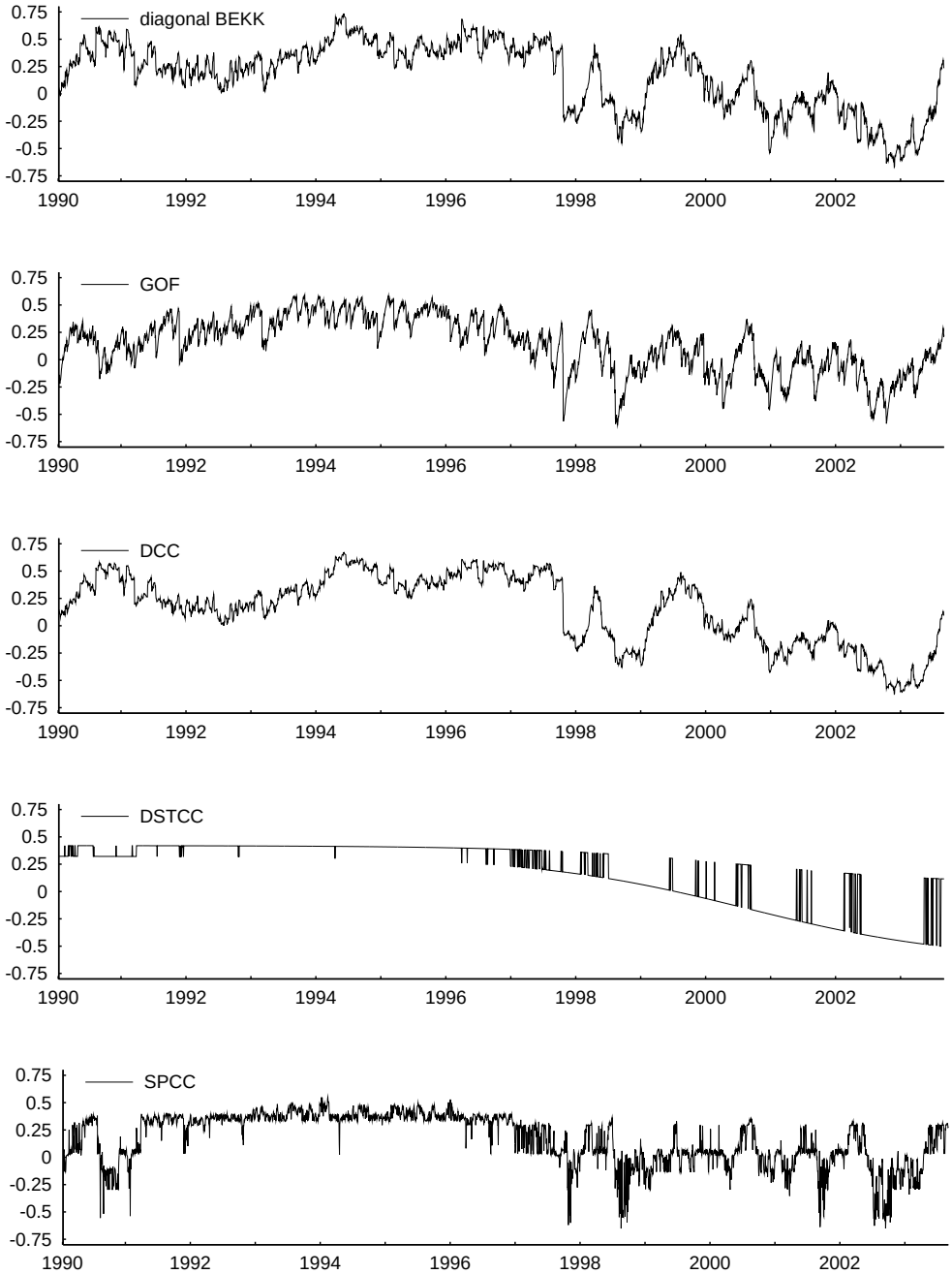
Estimation of multivariate GARCH models is not always easy. BEKK models appear more difficult to estimate than the CCC-GARCH model and its generalizations. It has not been possible in this review to cover algorithms for performing the necessary iterations. Brooks, Burke, and Persand (2003) compared four software packages for estimating MGARCH models. They used a single bivariate dataset and only fitted a first-order VEC-GARCH model to the data. A remarkable thing is that already the parameter estimates resulting from these packages are quite different, not to mention standard deviation estimates. The estimates give rather different ideas of the persistence of conditional volatility. These differences do not necessarily say very much about properties of the numerical algorithms used in the packages. It is more likely that they reflect the estimation difficulties. The log-likelihood function may contain a large number of local maxima, and different starting-values may thus lead to different outcomes. Attention should also be paid on the implementation of the estimation procedure in the software program used.

Not much has been done as yet to construct tests for evaluating MGARCH models. A few tests do exist, and a number of them have been considered in this review.

It may be that VEC and BEKK models, with the possible exception of factor

models, have already matured and there is not much that can be improved. The situation may be different for conditional correlation models. The focus has hitherto been on modelling the possibly time-varying correlations. Less emphasis has been put on the GARCH equations that typically have been GARCH(1,1) specifications. Designing diagnostic tools for testing and improving GARCH equations may be one of the challenges for the future.

Figure 1: Conditional correlations implied by the estimated models: Diagonal BEKK, GOF–GARCH, DCC–GARCH, DSTCC–GARCH, and SPCC–GARCH.



References

Alexander, C. O., and A. M. Chibumba (1997): "Multivariate orthogonal factor GARCH," University of Sussex Discussion Papers in Mathematics.

Bae, K.-H., G. A. Karolyi, and R. M. Stulz (2003): "A new approach to measuring financial contagion," *The Review of Financial Studies*, 16, 717–763.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79–109.

Bera, A. K., and S. Kim (2002): "Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns," *Journal of Empirical Finance*, 9, 171–195.

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., R. F. Engle, and D. B. Nelson (1994): "ARCH models," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2959–3038. Elsevier Science, Amsterdam.

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

Boussama, F. (1998): "Ergodicité, mélange et estimation dans les modèles GARCH," Thèse de l'Université Paris 7.

Brooks, C., S. P. Burke, and G. Persand (2003): "Multivariate GARCH models: software choice and estimation issues," *Journal of Applied Econometrics*, 18, 725–734.

Chib, S. (2007): "Multivariate stochastic volatility models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch. Springer, New York.

Comte, F., and O. Lieberman (2003): "Asymptotic theory for multivariate GARCH processes," *Journal of Multivariate Analysis*, 84, 61–84.

Diebold, F. X., and M. Nerlove (1989): "The dynamics of exchange rate volatility: a multivariate latent factor ARCH model," *Journal of Applied Econometrics*, 4, 1–21.

Doornik, J. A. (2002): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, 3rd edn., see also www.doornik.com.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1006.

——— (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. F., and R. Colacito (2006): "Testing and valuing dynamic correlations for asset allocation," *Journal of Business and Economic Statistics*, 24, 238–253.

Engle, R. F., and G. González-Rivera (1991): "Semiparametric ARCH models," *Journal of Business and Economic Statistics*, 9, 345–359.

Engle, R. F., C. W. J. Granger, and D. Kraft (1984): "Combining competing forecasts of inflation using a bivariate ARCH model," *Journal of Economic Dynamics and Control*, 8, 151–165.

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

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

Engle, R. F., and V. K. Ng (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–78.

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

Gouriéroux, C. (1997): *ARCH Models and Financial Applications*. Springer-Verlag, New York.

Hafner, C. M. (2003): "Fourth moment structure of multivariate GARCH models," *Journal of Financial Econometrics*, 1, 26–54.

Hafner, C. M., and J. V. K. Rombouts (in press): "Semiparametric multivariate volatility models," *Econometric Theory*.

Hafner, C. M., D. van Dijk, and P. H. Franses (2005): "Semi-parametric modelling of correlation dynamics," in *Advances in Econometrics*, ed. by T. Fomby, C. Hill, and D. Terrell, vol. 20/A, pp. 59–103. Amsterdam: Elsevier Sciences.

Hansen, B. E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413–430.

Hansson, B., and P. Hördahl (1998): "Testing the conditional CAPM using multivariate GARCH–M," *Applied Financial Economics*, 8, 377–388.

He, C., and T. Teräsvirta (2004): "An extended constant conditional correlation GARCH model and its fourth-moment structure," *Econometric Theory*, 20, 904–926.

Jeantheau, T. (1998): "Strong consistency of estimators for multivariate ARCH models," *Econometric Theory*, 14, 70–86.

Kawakatsu, H. (2006): "Matrix exponential GARCH," *Journal of Econometrics*, 134, 95–128.

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

Kwan, C. K., W. K. Li, and K. Ng (2005): "A multivariate threshold GARCH model with time-varying correlations," Preprint series 2005–12, National University of Singapore, Institution of Mathematical Sciences.

Lanne, M., and P. Saikkonen (in press): "A multivariate generalized orthogonal factor GARCH model," *Journal of Business and Economic Statistics*.

Li, W. K., and T. K. Mak (1994): "On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity," *Journal of Time Series Analysis*, 15, 627–636.

Ling, S., and W. K. Li (1997): "Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors," *Journal of Time Series Analysis*, 18, 447–464.

Ling, S., and M. McAleer (2003): "Asymptotic theory for a vector ARMA–GARCH model," *Econometric Theory*, 19, 280–310.

Long, X., and A. Ullah (2005): "Nonparametric and semiparametric multivariate GARCH model," Unpublished manuscript.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

McLeod, A. I., and W. K. Li (1983): "Diagnostic checking ARMA time series models using squared-residual autocorrelations," *Journal of Time Series Analysis*, 4, 269–273.

Nelson, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347–70.

Nelson, D. B., and C. Q. Cao (1992): "Inequality Constraints in the Univariate GARCH Model," *Journal of Business and Economic Statistics*, 10, 229–235.

Ng, L. (1991): "Tests of the CAPM with time-varying covariances: a multivariate GARCH approach," *The Journal of Finance*, 46, 1507–1521.

Pagan, A., and A. Ullah (1999): *Nonparametric Econometrics*. Cambridge University Press.

Palm, F. C. (1996): "GARCH models of volatility," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 14, pp. 209–240. Amsterdam: Elsevier Sciences.

Pelletier, D. (2006): "Regime switching for dynamic correlations," *Journal of Econometrics*, 131, 445–473.

Ross, S. A. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341–360.

Sentana, E. (1998): "The relation between conditionally heteroskedastic factor models and factor GARCH models," *Econometrics Journal*, 1, 1–9.

Shephard, N. (1996): "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, pp. 1–67. Chapman and Hall, London.

Silvennoinen, A., and T. Teräsvirta (2005): "Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations," SSE/EFI Working Paper Series in Economics and Finance No. 577.

——— (2006): "Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model," Unpublished manuscript.

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

Tse, Y. K., and K. C. Tsui (1999): "A note on diagnosing multivariate conditional heteroscedasticity models," *Journal of Time Series Analysis*, 20, 679–691.

——— (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

van der Weide, R. (2002): "GO–GARCH: A multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17, 549–564.

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

Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations

Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations

Abstract

In this paper we propose a new multivariate GARCH model with time-varying conditional correlation structure. The approach adopted here is based on the decomposition of the covariances into correlations and standard deviations. The time-varying conditional correlations change smoothly between two extreme states of constant correlations according to an endogenous or exogenous transition variable. An LM–test is derived to test the constancy of correlations and LM– and Wald tests to test the hypothesis of partially constant correlations. Analytical expressions for the test statistics and the required derivatives are provided to make computations feasible. An empirical example based on daily return series of five frequently traded stocks in the Standard & Poor 500 stock index completes the paper. The model is estimated for the full five-dimensional system as well as several subsystems and the results discussed in detail.

This paper is joint work with Timo Teräsvirta.

We thank Tim Bollerslev, Robert Engle, Markku Lanne, W. K. Li, Michael McAleer, and Y. K. Tse for helpful comments and suggestions. Participants of the conferences and workshops 'Econometrics and Computational Economics', Helsinki, November 2003, 14th Meeting of the New Zealand Econometric Study Group, Christchurch, March 2005, International Workshop on Econometrics and Statistics, Perth, April 2005, 4th Annual International Conference 'Forecasting Financial Markets and Decision-Making', Łódź , May 2005, 1st Symposium on Econometric Theory and Application, Taipei, May 2005, 3rd Nordic Econometric Meeting, Helsinki, May 2005, Economic Modelling Workshop, Brisbane, July 2005, Econometric Society World Congress, London, August 2005, and NSF/NBER Time Series Conference, Heidelberg, September 2005, also provided useful remarks. We are grateful to Mika Meitz for programming assistance. The responsibility for any errors and shortcomings in this paper remains ours.

1 Introduction

"...During major market events, correlations change dramatically ..." Bookstaber (1997)

Financial decision makers usually deal with many financial assets simultaneously. Modelling individual time series separately is thus an insufficient method as it leaves out information about comovements and interactions between the instruments of interest. Investors are facing risks that affect the assets in their portfolio in various ways which encourages them to find a position that allows to hedge against losses. In practice, this is often done by trying to diversify, possibly internationally, on many stock markets. When forming an efficient or optimal portfolio, correlations among, say, international stock returns are needed to determine gains from international portfolio diversification, and also the calculation of minimum variance hedge ratio needs updated correlations between assets in the hedge. Evidence that the correlations between national stock markets increase during financial crises but remain more or less unaffected during other times can be found for instance in King and Wadhwani (1990), Lin, Engle, and Ito (1994), de Santis and Gerard (1997), and Longin and Solnik (2001). As further examples, options depending on many underlying assets are very sensitive to correlations among those assets, and asset pricing models as well as some risk measures need measures of covariance between the assets in a portfolio. It is clear that there is an obvious need for a fexible and accurate model that can incorporate the information of possible comovements between the assets.

Volatility in multivariate financial data has been typically modelled applying the concept of conditional heteroskedasticity originally introduced by Engle (1982); see Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2007) for recent reviews. In the multivariate context, one also has to model the conditional covariances, not only the conditional variances. One possibility is to model the former directly, VEC and BEKK as well as F–GARCH, O–GARCH, GO–GARCH, and GDC–GARCH models may serve as examples. Another alternative is to model them through conditional correlations. One of the most frequently used multivariate GARCH models is the Constant Conditional Correlation (CCC) GARCH model of Bollerslev (1990). In this model comovements between heteroskedastic time series are modelled by allowing each series to follow a separate GARCH process while restricting the conditional correlations between the GARCH processes to be constant. The estimation of parameters of the CCC–GARCH model is relatively simple and the model has thus become popular among practitioners. An extension to CCC–GARCH model allowing dynamic interactions between the conditional variance equations, which creates a richer and more flexible autocorrelation structure than the one in the standard CCC–GARCH model, was introduced by Jeantheau (1998), and its moment structure was considered by He and Teräsvirta (2004).

In practice, the assumption of constant conditional correlations has often been found too restrictive, especially in the context of asset returns. Tests developed by Tse (2000) and Bera and Kim (2002) often reject the constancy of conditional correlations. There is evidence that the correlations are not only dependent on time

but also on the state of uncertainty in the markets. The conditional correlations can thus fluctuate over time and, in particular, they have been reported to increase during periods of market turbulence.

Tse and Tsui (2002) and Engle (2002) defined dynamic conditional correlation GARCH models (VC–GARCH and DCC–GARCH, respectively) that impose GARCH-type dynamics on the conditional correlations as well as on the conditional variances. These models are flexible enough to capture many kinds of heteroskedastic behaviour in multivariate series. The number of parameters in Engle's DCC–GARCH model remains relatively low because all conditional correlations are generated by first-order GARCH processes with identical parameters. In this model, the GARCH-type correlation processes are not linked to the individual GARCH processes of the return series. Recently, Kwan, Li, and Ng (2005) proposed an extension to the VC–GARCH model using a threshold approach.

Pelletier (2006) recently proposed a model with a regime-switching correlation structure driven by an unobserved state variable that follows a two-dimensional first-order Markov chain. The regime-switching model asserts that the correlations remain constant in each regime and the change between the states is abrupt and governed by transition probabilities. This model is motivated by the empirical finding that the correlations among asset returns tend to increase during periods of distress whereas the series behave in a more independent manner in tranquil periods.

In this paper we introduce another way of modelling comovements in the returns. The Smooth Transition Conditional Correlation (STCC) GARCH model allows the conditional correlations to change smoothly from one state to another as a function of a transition variable. This continuous variable may be a combination of observable stochastic variables, or a function of a lagged error term or terms. The empirical performance of the STCC-GARCH model thus depends on the ability of the transition variable to represent the forces affecting the conditional correlations.

A distinguishing feature of the STCC–GARCH model is that there is interaction between the volatility and correlation structures of the model. The model also has the appealing feature that it provides a framework in which constancy of the correlations and thus the adequacy of the model can be tested in a straightforward fashion. A special case of the STCC–GARCH model was independently introduced by Berben and Jansen (2005). Their model is bivariate, and the variable controlling the transition between the extreme regimes is simply the time.

The paper is organized as follows. In Section 2 the model is introduced and the estimation of its parameters considered. Section 3 is devoted to tests of constant correlations and partially constant correlations. Results of simulation experiments are reported in Section 4, and an application to illustrate the capabilities of the model is discussed in Section 5. Section 6 concludes. Technical derivations of the test statistics presented in the paper can be found in the Appendix.

2 The Smooth Transition Conditional Correlation GARCH model

2.1 The general multivariate GARCH model

Consider the following stochastic N-dimensional vector process with the standard representation

$$\boldsymbol{y}_t = E\left[\boldsymbol{y}_t \mid \mathcal{F}_{t-1}\right] + \boldsymbol{\varepsilon}_t \qquad t = 1, 2, \ldots, T \tag{1}$$

where $\mathcal{F}_{t-1}$ is the sigma-field generated by all the information until time $t-1$. Each of the univariate error processes has the specification

$$\varepsilon_{it} = h_{it}^{1/2} z_{it},$$

where the errors z_{it} form a sequence of independent random variables with mean zero and variance one, for each $i = 1, \ldots, N$. The conditional variance h_{it} follows a univariate GARCH process

$$h_{it} = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} \varepsilon_{i,t-j}^2 + \sum_{j=1}^{p} \beta_{ij} h_{i,t-j} \tag{2}$$

with the non-negativity and stationarity restrictions imposed. The first and second conditional moments of the vector $\boldsymbol{z}_t$ are given by

$$\begin{aligned} E\left[\boldsymbol{z}_t \mid \mathcal{F}_{t-1}\right] &= \boldsymbol{0}, \\ E\left[\boldsymbol{z}_t \boldsymbol{z}_t' \mid \mathcal{F}_{t-1}\right] &= \boldsymbol{P}_t. \end{aligned} \tag{3}$$

Furthermore, the standardized errors $\boldsymbol{\eta}_t = \boldsymbol{P}_t^{-1/2} \boldsymbol{z}_t \sim iid(\boldsymbol{0}, \boldsymbol{I}_N)$. Since z_{it} has unit variance for all i, $\boldsymbol{P}_t = [\rho_{ij,t}]_{i,j=1,\ldots,N}$ is the conditional correlation matrix for the $\boldsymbol{\varepsilon}_t$ where

$$\begin{aligned} \rho_{ij,t} &= E\left[z_{it} z_{jt} \mid \mathcal{F}_{t-1}\right] = \frac{E\left[z_{it} z_{jt} \mid \mathcal{F}_{t-1}\right]}{\sqrt{E\left[z_{it}^2 \mid \mathcal{F}_{t-1}\right] E\left[z_{jt}^2 \mid \mathcal{F}_{t-1}\right]}} \\ &= \frac{E\left[\varepsilon_{it} \varepsilon_{jt} \mid \mathcal{F}_{t-1}\right]}{\sqrt{E\left[\varepsilon_{it}^2 \mid \mathcal{F}_{t-1}\right] E\left[\varepsilon_{jt}^2 \mid \mathcal{F}_{t-1}\right]}} = Corr\left[\varepsilon_{it}, \varepsilon_{jt} \mid \mathcal{F}_{t-1}\right]. \end{aligned} \tag{4}$$

The correlations $\rho_{ij,t}$ are allowed to be time-varying in a manner that will be defined later on. It will, however, be assumed that $\boldsymbol{P}_t \in \mathcal{F}_{t-1}$.

To establish the connection to the approach often used in context of conditional correlation models, let us denote the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$ as

$$E\left[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' \mid \mathcal{F}_{t-1}\right] = \boldsymbol{H}_t = \boldsymbol{S}_t \boldsymbol{P}_t \boldsymbol{S}_t$$

where $\boldsymbol{P}_t$ is the conditional correlation matrix as in equation (3) and $\boldsymbol{S}_t = diag(h_{1t}^{1/2}, .., h_{Nt}^{1/2})$ with elements defined in (2). For the positive definiteness of $\boldsymbol{H}_t$ it

is sufficient to require the correlation matrix $\boldsymbol{P}_t$ to be positive definite at each point in time. It follows that the error process in (1) can be written as

$$\boldsymbol{\varepsilon}_t = \boldsymbol{H}_t^{1/2}\boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim iid\left(\mathbf{0}, \boldsymbol{I}_N\right).$$

2.2 Smooth transitions in conditional correlations

In order to complete the definition of the model we have to specify the time-varying structure of the conditional correlations in (4). We propose the Smooth Transition Conditional Correlation GARCH (STCC–GARCH) model, in which the conditional correlations are assumed to change smoothly over time depending on a transition variable. In the simplest case there are two extreme states of nature with state-specific constant correlations among the variables. The correlation structure changes smoothly between the two extreme states of constant correlations as a function of the transition variable. More specifically, the conditional correlation matrix $\boldsymbol{P}_t$ is defined as follows:

$$\boldsymbol{P}_t = \left(1 - G_t\right)\boldsymbol{P}_1 + G_t\boldsymbol{P}_2 \tag{5}$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are positive definite correlation matrices. Furthermore, G_t is a transition function whose values are bounded between 0 and 1. This structure ensures $\boldsymbol{P}_t$ to be positive definite with probability one, because it is a convex combination of two positive definite matrices.

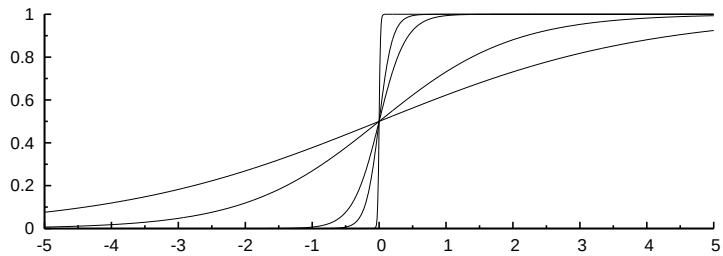
The transition function is chosen to be the logistic function

$$G_t = \left(1 + e^{-\gamma(s_t - c)}\right)^{-1}, \quad \gamma > 0 \tag{6}$$

where s_t is the transition variable, c determines the location of the transition and $\gamma > 0$ the slope of the function, that is, the speed of transition. The typical shape of the transition function is illustrated in Figure 1. Increasing γ, increases the speed of transition from 0 to 1 as a function of s_t, and the transition between the two extreme correlation states becomes abrupt as $\gamma \to \infty$. For simplicity, the parameters c and γ are assumed to be the same for all correlations. This assumption may sometimes turn out to be restrictive, but letting different parameters control the location and the speed of transition in correlations between different series may cause conceptual difficulties. This is because then $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ being positive definite does not imply the positive definiteness of every $\boldsymbol{P}_t$.

The choice of transition variable s_t depends on the process to be modelled. An important feature of the STCC–GARCH model is that the investigator can choose s_t to fit his research problem. In some cases, economic theory proposals may determine the transition variable, in some others available empirical information may be used for the purpose. Possible choices include time as in Berben and Jansen (2005), or functions of past values of one or more of the return series. Yet another option would be to use an exogenous variable, which is a natural idea for example when

Figure 1: The logistic transition function for $\gamma = 0.5$, 1, 5, 10, and 100 and location $c = 0$.



co-movements of individual stock returns are linked to the behaviour of the stock market itself. In that case, s_t could be a function of lagged values of the whole index. One could use the past conditional variance of the index returns, which Lanne and Saikkonen (2005) suggested when they constructed a univariate smooth transition GARCH model.

Another point worth considering in this context is the number of parameters. It increases rapidly with the number of series in the model, although the current parameterization is still quite parsimonious. However, if one wishes to model the dynamic behaviour between the series, a very small number of parameters may not be enough. Simplifications that are too radical are likely to lead to models that do not capture the behaviour that was to be modelled in the first place. It is possible to simplify the STCC–GARCH model at least to some extent such that it may still be useful in certain applications. As an example, one may restrict one of the two extreme correlation states to be that of complete independence ($\boldsymbol{P}_j = \boldsymbol{I}_N$, $j = 1$ or 2). This is a special case of a model where $\boldsymbol{P}_j = [\rho_{ij}]$ such that $\rho_{ij} = \rho$, $i \neq j$. Another possibility is to allow some of the conditional correlations to be time-varying, while the others remain constant over time. Examples of this will be discussed both in connection with testing and in the empirical application.

2.3 Estimation of the STCC–GARCH model

For the maximum likelihood estimation of parameters we assume joint conditional normality of the errors:

$$\boldsymbol{z}_t \mid \mathcal{F}_{t-1} \sim N\left(\boldsymbol{0}, \boldsymbol{P}_t\right).$$

Denoting by $\boldsymbol{\theta}$ the vector of all the parameters in the model, the log-likelihood for observation t is

$$l_t\left(\boldsymbol{\theta}\right) = -\frac{N}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{i=1}^{N}\log h_{it} - \frac{1}{2}\log\left|\boldsymbol{P}_t\right| - \frac{1}{2}\boldsymbol{z}_t'\boldsymbol{P}_t^{-1}\boldsymbol{z}_t, \quad t = 1, \ldots, T \tag{7}$$

and maximizing $\sum_{t=1}^{T} l_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ yields the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T$.

Asymptotic properties of the maximum likelihood estimators in the present case remain to be established. Bollerslev and Wooldridge (1992) provided a proof of consistency and asymptotic normality of the quasi maximum likelihood estimators in the context of general dynamic multivariate models. Recently, Ling and McAleer (2003) considered a class of vector ARMA–GARCH models and established strict stationarity and ergodicity as well as consistency and asymptotic normality of the QMLE under some reasonable moment conditions. In their model, however, the conditional correlations are assumed to be constant. Extending their results to cover the present situation would be interesting but is beyond the scope of this paper. For inference we merely assume that the asymptotic distribution of the ML-estimator is normal, that is,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathfrak{I}^{-1}(\boldsymbol{\theta}_0)\right)$$

where $\boldsymbol{\theta}_0$ is the true parameter and $\mathfrak{I}(\boldsymbol{\theta}_0)$ the population information matrix evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Before estimating the STCC–GARCH model, however, it is necessary to first test the hypothesis that the conditional correlations are constant. The reason for this is that some of the parameters of the STCC–GARCH model are not identified if the true model has constant conditional correlations. Estimating an STCC–GARCH model without first testing the constancy hypothesis could thus lead to inconsistent parameter estimates. The same is true if one wishes to increase the number of transitions in an already estimated model. Testing constancy of conditional correlations will be discussed in the next section.

Maximization of the log-likelihood (7) with respect to all the parameters at once can be difficult due to numerical problems. For the DCC–GARCH model, Engle (2002) proposed a two-step estimation procedure based on the decomposition of the likelihood into a volatility and a correlation component. The univariate GARCH models are estimated first, independently of each other, and the correlations thereafter, conditionally on the GARCH parameter estimates. This implies that the dynamic behaviour of each return series, characterized by an individual GARCH process, is not linked to the time-varying correlation structure. Under this assumption, the parameter estimates of the DCC–GARCH model are consistent under reasonable regularity conditions; see Engle (2002) and Engle and Sheppard (2001) for discussion. For comparison, in the STCC–GARCH model the dynamic conditional correlations form a channel of interaction between the volatility processes. Parameter estimation accommodates this fact: the parameters are estimated simultaneously by conditional maximum likelihood.

Due to the large number of parameters in the model, estimation of the STCC-GARCH model is carried out iteratively by concentrating the likelihood. The parameters are divided into three sets: parameters in the GARCH equations, correlations, and parameters of the transition function, and the log-likelihood is maximized by sequential iteration over these sets. After the first completed iteration, the parameter estimates correspond to the estimates obtained by a two-step estimation procedure. Even if the parameter estimates do not change much during the sequence of iterations, the iterative method increases efficiency by yielding smaller standard errors than the

two-step method. Furthermore, convergence is generally reached with a reasonable number of iterations.

It should be pointed out, however, that estimation requires care. The log-likelihood may have several local maxima, so estimation should be initiated from a set of different starting-values and achieved maxima compared before settling for final estimates.

3 Testing constancy of correlations

3.1 Test of constant conditional correlations

As already mentioned, the modelling of time-varying conditional correlations has to begin by testing the hypothesis of constant correlations. Tse (2000), Bera and Kim (2002), and Engle and Sheppard (2001) already proposed tests for this purpose. We shall present an LM–type test of constant conditional correlations against the STCC–GARCH alternative. A rejection of the null hypothesis supports the hypothesis of time-varying correlations or other types of misspecification but does not imply that the data have been generated from an STCC–GARCH model. For this reason our LM–type test can also be seen as a general misspecification test of the CCC–GARCH model.

In order to derive the test, consider an N-variate case where we wish to test the assumption of constant conditional correlations against conditional correlations that are time-varying with a simple transition of type (5) with a transition function defined by (6). For simplicity, assume that the conditional variance of each of the individual series follows a GARCH$(1,1)$ process and let $\boldsymbol{\omega}_i = (\alpha_{i0}, \alpha_i, \beta_i)'$ be the vector of parameters for conditional variance h_{it}. Generalizing the test to other types of GARCH models for the individual series is straightforward. The STCC–GARCH model collapses into a constant correlation model under the null hypothesis

$$H_0: \quad \gamma = 0$$

in (6). When this restriction holds, however, some of the parameters of the model are not identified. To circumvent this problem, we follow Luukkonen, Saikkonen, and Teräsvirta (1988) and consider an approximation of the alternative hypothesis. It is obtained by a first-order Taylor approximation around $\gamma = 0$ to the transition function G_t:

$$\begin{aligned} G_t &= \left(1 + e^{-\gamma(s_t - c)}\right)^{-1} \\ &\doteq 1/2 + (1/4)\,(s_t - c)\,\gamma. \end{aligned} \tag{8}$$

Applying (8) to (5) linearizes the time-varying correlation matrix $\boldsymbol{P}_t$ as follows:

$$\boldsymbol{P}_t^* = \boldsymbol{P}_1^* - s_t \boldsymbol{P}_2^*$$

where

$$\begin{aligned} \boldsymbol{P}_1^* &= \frac{1}{2}(\boldsymbol{P}_1 + \boldsymbol{P}_2) + \frac{1}{4}c(\boldsymbol{P}_1 - \boldsymbol{P}_2)\gamma, \\ \boldsymbol{P}_2^* &= \frac{1}{4}(\boldsymbol{P}_1 - \boldsymbol{P}_2)\gamma. \end{aligned} \tag{9}$$

If $\gamma = 0$, then $\boldsymbol{P}_2^* = 0$ and the correlations are constant. Thus we construct an auxiliary null hypothesis $H_0^{\text{aux}} : \boldsymbol{\rho}_2^* = 0$ where $\boldsymbol{\rho}_2^* = vecl\boldsymbol{P}_2^*$.[1]

This null hypothesis can be tested by an LM–test. Note that when H_0 holds, there is no approximation error because then $G_t \equiv 1/2$, and the standard asymptotic theory remains valid. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_N', \boldsymbol{\rho}_1^{*\prime}, \boldsymbol{\rho}_2^{*\prime})'$, where $\boldsymbol{\rho}_j^* = vecl\boldsymbol{P}_j^*$, $j = 1, 2$, be the vector of all parameters of the model. Under standard regularity conditions, the LM–statistic

$$LM_{CCC} = T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\rho}_2^{*\prime}}\right)\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_2^*,\boldsymbol{\rho}_2^*)}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\rho}_2^{*}}\right), \tag{10}$$

evaluated at the maximum likelihood estimators under the restriction $\boldsymbol{\rho}_2^* = 0$, has an asymptotic χ^2 distribution with $N(N-1)/2$ degrees of freedom. In expression (10), $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho}_2^*,\boldsymbol{\rho}_2^*)}$ is the south-east $\frac{N(N-1)}{2} \times \frac{N(N-1)}{2}$ block of the inverse of $\hat{\mathfrak{I}}_T$, where $\hat{\mathfrak{I}}_T$ is a consistent estimator for the asymptotic information matrix. For derivation and details of the statistic, as well as the suggested consistent estimator for the asymptotic information matrix, see the Appendix.

A straightforward extension is to test the constancy of conditional correlations against partially constant correlations:

$$\begin{aligned} H_0 &: \quad \gamma = 0 \\ H_1 &: \quad \rho_{ij,1} = \rho_{ij,2} \quad \text{for } (i,j) \in N_1 \end{aligned}$$

where $N_1 \subset \{1, \ldots, N\} \times \{1, \ldots, N\}$. Under the null hypothesis we again face the identification problem which is solved by linearizing the transition function. Details are given in the Appendix.

These tests involve a particular transition variable. Thus a failure to reject the null of constant correlations is just an indication that there is no evidence of time-varying correlations, given this transition variable. Evidence of time-varying correlations may still be found in case of another indicator. This highlights the importance of choosing a relevant transition variable for the data at hand, and in practice it may be useful to consider several alternatives unless restrictions implied by economic theory make the choice unique.

It should be mentioned that Berben and Jansen (2005) have in a bivariate context coincidentally proposed a test of the correlations being invariant with respect to calendar time. Their test is derived using an approach similar to ours, but they choose

[1] The *vecl* operator stacks the columns of the strict lower diagonal (obtained by excluding the diagonal elements) of the square argument matrix.

a different estimator for the information matrix in (10). Based on our simulation experiments, the estimator they use is substantially less efficient in finite samples than ours, especially when the number of series in the model is large.

3.2 Test of partially constant conditional correlations

The LM–statistic (10) is designed to test the null hypothesis of constant conditional correlations against the STCC–GARCH model. After estimating the model, it is possible to test the constancy of conditional correlations between a subset of return series such that the other conditional correlations remain time-varying both under the null hypothesis and the alternative. We derive both a Lagrange multiplier and a Wald test for this purpose. Both have a partially constant STCC–GARCH model as the null hypothesis, meaning that some of the correlations are constrained to be constant under

$$H_0: \quad \rho_{ij,1} = \rho_{ij,2} \quad \text{for } (i,j) \in N_0$$

where $N_0 \subset \{1,\ldots,N\} \times \{1,\ldots,N\}$. The alternative hypothesis is an unrestricted STCC–GARCH model. The identification problem encountered when testing whether the complete model has constant correlations is not present here. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}_1',\ldots,\boldsymbol{\omega}_N',\boldsymbol{\rho}_1',\boldsymbol{\rho}_2',c,\gamma)'$, where $\boldsymbol{\rho}_1 = vecl\boldsymbol{P}_1$ and $\boldsymbol{\rho}_2 = vecl\boldsymbol{P}_2$. Under standard regularity conditions, the usual Wald-statistic

$$W_{PCCC} = T\boldsymbol{a}(\hat{\boldsymbol{\theta}})' \left(\boldsymbol{A} \left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}}) \right]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})} \boldsymbol{A}' \right)^{-1} \boldsymbol{a}(\hat{\boldsymbol{\theta}}), \tag{11}$$

evaluated at the maximum likelihood estimators of the full STCC model, has an asymptotic χ^2 distribution with degrees of freedom equal to the number of constraints to be tested. Furthermore, in expression (11), $\boldsymbol{A} = \partial\boldsymbol{a}/\partial\boldsymbol{\rho}'$ where $\boldsymbol{\rho} = (\boldsymbol{\rho}_1',\boldsymbol{\rho}_2')'$ and $\boldsymbol{a}$ is the vector of constraints, and $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})}$ is the block corresponding to the correlation parameters of the inverse of $\hat{\mathfrak{I}}_T$.

It is also possible to apply the LM principle to this problem. Under the null hypothesis, the statistic

$$LM_{PCCC} = T^{-1}\boldsymbol{q}(\hat{\boldsymbol{\theta}})' \left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}}) \right]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})} \boldsymbol{q}(\hat{\boldsymbol{\theta}}) \tag{12}$$

evaluated at the restricted maximum likelihood estimators, has an asymptotic χ^2 distribution with the number of degrees of freedom equal to the number of constraints to be tested. In (12), $\boldsymbol{q}$ is the block of the score vector corresponding to the correlation parameters, and $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})}$ is again the block corresponding to the correlation parameters of the inverse of $\hat{\mathfrak{I}}_T$. The derivation and details of the Wald and LM–statistics as well as of the consistent estimator of the asymptotic information matrix are given in the Appendix.

The most important difference between the two test statistics is that the Wald statistic is evaluated at the estimates obtained from the estimation of the full STCC–GARCH model, whereas the LM–statistic makes use of restricted estimates. Choosing

the test has implications when the computation time of the statistic is concerned. For example, when one half of the correlation parameters are restricted, the estimation time is reduced by more than one half, compared to the estimation of the unrestricted model. Furthermore, since the models in question are complicated and nonlinear, it is preferable to first estimate the restricted model and then evaluate the need for a more general specification.

A straightforward extension to these tests is to allow some of the correlations to be constant even in the alternative model. This leads to the following pair of hypotheses:

$$\begin{aligned} H_0 \quad &: \quad \rho_{ij,1} = \rho_{ij,2} \quad \text{for } (i,j) \in N_0 \\ H_1 \quad &: \quad \rho_{ij,1} = \rho_{ij,2} \quad \text{for } (i,j) \in N_1 \end{aligned}$$

where $N_1 \subset N_0 \subset \{1, \ldots, N\} \times \{1, \ldots, N\}$. This may be useful when the dimension of the model is high and constancy of at least some of the conditional correlations would be an appropriate initial simplification. The details for this case can be found in the Appendix.

4 Simulation experiments

4.1 Size study

We study the finite-sample properties of the test of constant conditional correlations in a small simulation experiment. We generate data from a bivariate GARCH$(1,1)$ model with normal errors and choose a transition variable external to the model. The values of this variable are generated from a univariate GARCH$(1,1)$ model. Furthermore, the strength of the constant conditional correlation between the two series in the model is varied throughout from -0.9 to 0.9. The three series are parameterized as follows:

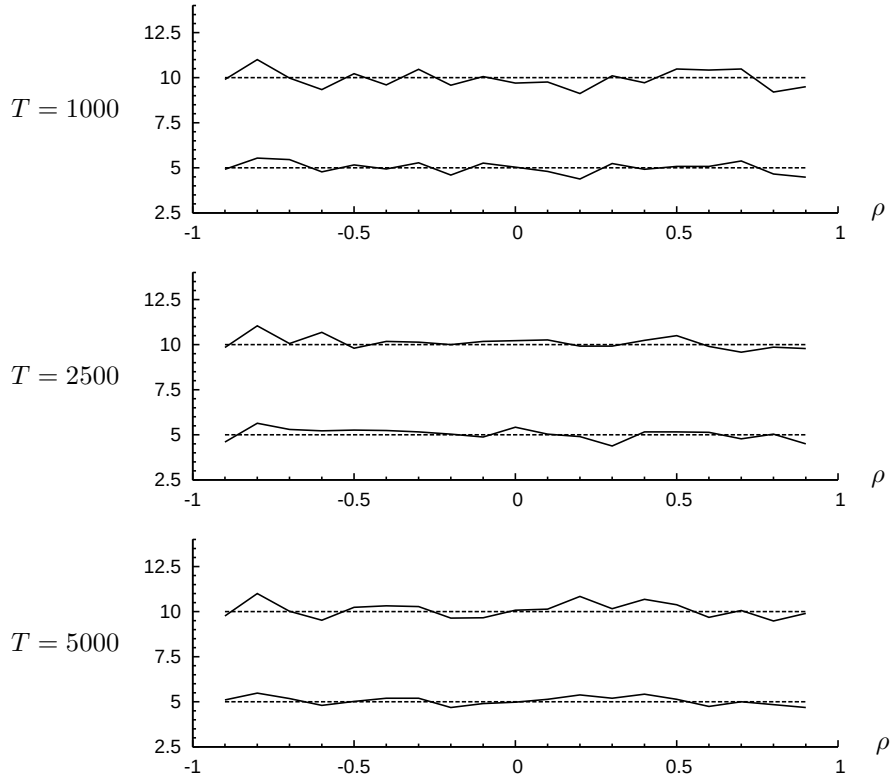
i	α_{i0}	α_{i1}	β_{i1}
1	0.02	0.04	0.95
2	0.01	0.03	0.96
ext	0.002	0.06	0.91

where the values for the GARCH parameters are chosen to be representative for two stocks and the S&P500 index. The sample sizes are 1000, 2500, and 5000, and the number of replications is 5000. To eliminate initialization effects, the first 1000 observations are removed from the series before generating the actual observations.

Values of the test statistic (10) are calculated using the analytical expression for the estimator of the information matrix. The rejection frequencies at the asymptotic significance levels 0.05 and 0.10 are shown in Figure 2. The actual size of the test seems to be quite close to the nominal size already for the sample size of 1000.

We also carried out experiments with the partial constancy tests, albeit with fewer replications, because the computational burdens were substantial. The results were, however, quite satisfactory, suggesting that the tests do not suffer from size distortion.

Figure 2: Actual size of the test of constant correlation plotted against different values of correlation using simulated data (5000 replicates), and sample sizes 1000, 2500 and 5000. Rejection frequencies are in percentages. The test is calculated with covariance estimator (14).



4.2 Power study

There is no direct benchmark to which to compare our constancy test. It may be interesting, however, to find out how powerful the test is when the data are generated by the DCC-GARCH model of Engle (2002). The VC-GARCH model of Tse and Tsui (2002) would be a comparable choice of an alternative. In this case, it is natural to assume the transition variable to be a function of generated observations, because correlations generated by the DCC-GARCH model are not influenced by exogenous information. The power of our test depends on the choice of the transition variable, and for this reason choosing a variable that is not informative about the change in the correlations yields low or no power. This is confirmed by our simulations. The power turns out to be very close to the nominal size when the transition variable carries no information of the variability of the correlations even when the data are generated from the STCC–GARCH model.

Our first choice of a transition variable is a linear combination of lags of squared

3–variate

i	α_{i0}	α_{i1}	β_{i1}
1	0.02	0.04	0.95
2	0.01	0.03	0.96
3	0.002	0.06	0.93

$\overline{R}$	y_1	y_2
y_2	0.6	
y_3	0.5	0.4

α	β
0.05	0.90

5–variate

i	α_{i0}	α_{i1}	β_{i1}
1	0.02	0.04	0.95
2	0.01	0.03	0.96
3	0.002	0.06	0.93
4	0.007	0.02	0.97
5	0.03	0.05	0.94

$\overline{R}$	y_1	y_2	y_3	y_4
y_2	0.8			
y_3	0.75	0.65		
y_4	0.7	0.6	0.5	
y_5	0.65	0.55	0.45	0.35

α	β
0.05	0.90

Table 1: Parameters for the DCC–GARCH model used in the power simulations: α_{i0}, α_{i1}, and β_{i1} are the GARCH parameters for series i, $\overline{R}$ is the unconditional correlation matrix, and α and β are the parameters governing the dynamics of the conditional correlations.

returns. The transition variable equals $s_t = \boldsymbol{w}'\bar{\boldsymbol{y}}_t^{(2)}$ where $\boldsymbol{w}$ is 5×1 vector of weights and $\bar{\boldsymbol{y}}_t^{(2)}$ is 5×1 vector whose kth element is $\frac{1}{N}\sum_{i=1}^{N} y_{i,t-k}^2$. The following weighting schemes are considered:

$$\begin{array}{ll} \text{equal:} & \boldsymbol{w} = (0.2, 0.2, 0.2, 0.2, 0.2)' \\ \text{arithmetic:} & \boldsymbol{w} = (0.3, 0.25, 0.2, 0.15, 0.1)' \\ \text{geometric:} & \boldsymbol{w} = (0.5, 0.25, 0.125, 0.0625, 0.0625)' \end{array}$$

The power simulations are performed on three- and five-dimensional models parameterized as shown in Table 1. As in the size study, the sample sizes are 1000, 2500, and 5000 and the number of replications 5000. The rejection frequencies are presented in Table 2.

As can be seen, the power increases with the number of the series. This may not be surprising because the DCC–GARCH model imposes the same dynamic structure on all cross-products $z_{it}z_{jt}$, and the time-varying structure thus becomes more evident as the dimension of the model increases. The power results suggest that the average squared returns over the past five days are quite informative indicators of correlations generated by the DCC–GARCH model. An STCC-GARCH model with this transition variable could be a reasonable substitute for the (true) DCC-GARCH model.

We have also simulated the case in which the transition variable used in the test is a function of past returns such that it accounts for both the sign and the size of returns. In this experiment our test has low power when the data are generated by the DCC-GARCH model. Thus, if we believe that both the direction and the strength of

	$T = 1000$		$T = 2500$		$T = 5000$	
3–variate	5%	10%	5%	10%	5%	10%
equal	0.474	0.586	0.783	0.851	0.960	0.975
arithmetic	0.461	0.571	0.771	0.841	0.952	0.974
geometric	0.361	0.480	0.669	0.763	0.911	0.942
5–variate	5%	10%	5%	10%	5%	10%
equal	0.803	0.873	0.987	0.991	1.000	1.000
arithmetic	0.783	0.855	0.983	0.991	1.000	1.000
geometric	0.641	0.744	0.963	0.976	0.998	0.999

Table 2: Rejection frequencies for the test of constant conditional correlations when the data generating processes are the DCC–GARCH models in Table 1.

the movement of the stock price jointly affect the conditional correlations, the DCC–GARCH model may not be our favourite model. If the effects moving the correlations are essentially 'market effects', they can be taken into account by the STCC–GARCH model.

5 Application to daily stock returns

The data set of our application consists of daily returns of five S&P 500 composite stocks traded at the New York Stock Exchange and the S&P 500 index itself. The main criterion for choosing the stocks is that they are frequently traded and that the trades are often large. The stocks are Ford, General Motors, Hewlett-Packard, IBM, and Texas Instruments, and the observation period begins January 3, 1984 and ends December 31, 2003. As usual, closing prices are transformed into returns by taking natural logarithms, differencing, and multiplying by 100, which gives a total of 5038 observations for each of the series. To avoid problems in the estimation of the GARCH equations, the observations in the series are truncated such that extremely large negative returns are set to a common value of -10. This is preferred to removing them altogether, because we do not want to remove the information in comovements related to very large negative returns. It turns out that the truncated observations lie more than ten standard deviations below the mean. Descriptive statistics of the return series can be found in Table 3. To give an idea of how truncation affects the overall properties of the series, the values of the statistics have been tabulated both before and after truncation.

5.1 Choosing the transition variable

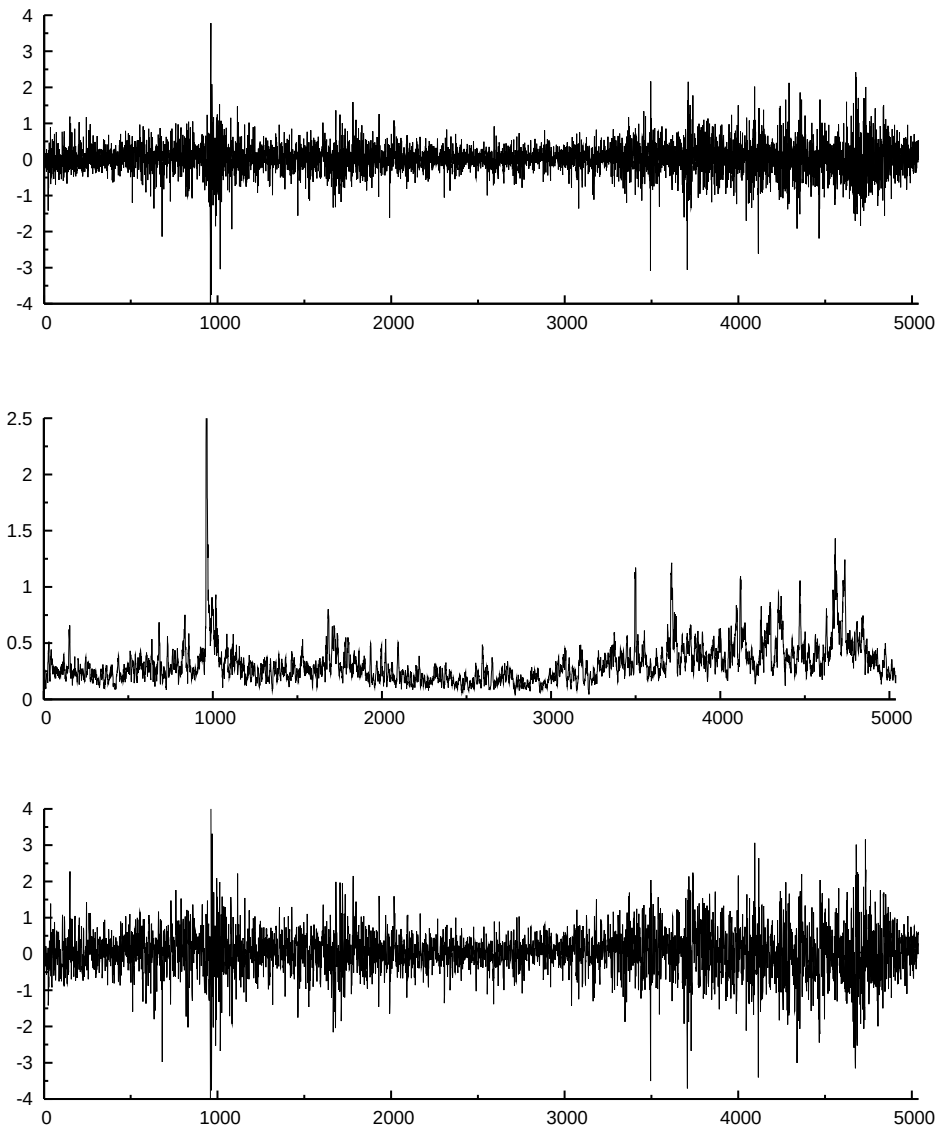
We consider the possibility that common shocks affect conditional correlations between daily returns. The transition variable in the transition function is a function of lagged returns of the S&P 500 index. As already discussed in Section 2.2, several choices are available. A question frequently investigated, see for instance Andersen, Bollerslev, Diebold, and Labys (2001) and Chesnay and Jondeau (2001), is whether comovements in the returns are stronger during general market turbulence than they are during more tranquil times. In that case, a lagged squared or absolute daily return, or a sum of lags of either ones, would be an obvious choice. Following Lanne and Saikkonen (2005), one could also consider the conditional variance of the S&P 500 returns. A model-based estimate of this quantity may be obtained by specifying and estimating an adequate GARCH model for the S&P 500 return series.

We restricted our attention to different functions of lagged squared and absolute returns of the index. Specifically, we considered the unweighted averages of the both lagged squared and lagged absolute returns over periods ranging from one to twenty days, and weighted averages of the same quantities with exponentially decaying weights with the discount ratios 0.9, 0.7, 0.5, and 0.3. The constant conditional correlations hypothesis was then tested using each of the 48 transition variables in the complete five-dimensional model as well as in every one of its submodels. The clearly strongest overall rejection occurred (these results are not reported here) when the transition variable was the seven-day average of lagged absolute returns. The graph of this transition variable is presented in Figure 3.

Table 4 contains the p–values of the constancy test based on this transition variable for all bivariate, trivariate, four-variable, and the full five-variable CCC–GARCH model. The test rejects the null hypothesis of constant correlations at significance level 0.01 for all bivariate, eight out of ten trivariate and four out of five four-variate models. The full five-variate CCC–GARCH model is rejected against STCC–GARCH as well. Generally, the rejections grow stronger as the dimension of the model increases, but the decrease of p–values is not monotonic.

If the interest lies in finding out whether the direction of the price movement as well as its strength affect conditional correlations, a function of lagged returns that preserves the sign of the returns is an appropriate transition variable. In order to accommodate this possibility, we considered the following two sets of lagged returns: $\{r_{t-j} : j = 1, \ldots, 10\}$ and $\{\sum_{i=1}^{j} r_{t-i} : j = 2, \ldots, 10\}$; note that $\sum_{i=1}^{j} r_{t-i} = 100(p_{t-1} - p_{t-(j+1)})$ where p_t is the log-price of the stock. The constant conditional correlations hypothesis was tested using these nineteen choices of transition variables. The strongest rejection most frequently occurred (results not reported here) when the transition variable was a lagged two-day return $100(p_{t-1} - p_{t-3})$. In this case, all CCC–GARCH models except the bivariate Ford-General Motors one, were rejected at the 0.01 level, most of them very strongly. The transition variable is depicted in Figure 3.

Figure 3: The S&P 500 returns from January 3, 1984 to December 31, 2003. The upper panel shows the returns (one observation falls outside the presented range), the middle panel shows the average of the absolute value returns over seven days (two observations fall outside the presented range), and the lower panel shows the log of the price difference over two days, or return over two days (five observations fall outside the presented range).



5.2 Effects of market turbulence on conditional correlations

We shall first investigate the case in which conditional correlations are assumed to fluctuate as a result of time-varying market distress which is measured by lagged seven-day averages of absolute S&P 500 returns. Four remarks are in order. First, we only consider first-order STCC–GARCH models. One may want to argue that more effort should be invested in the specification of the individual volatility models. However, in the context of financial returns data it is often found that the first-order GARCH model performs sufficiently well, and for the time being we settle for this the option. Second, estimation of parameters is carried out both by the iterative maximum likelihood and by the two-step procedure.[2] The standard errors of the parameter estimates of the STCC–GARCH model are calculated using numerical second derivatives for all estimates except the estimate of the velocity of transition parameter γ. This exception is due to the fact that in most of the cases the sequence of estimates of γ is converging towards some very large value, which causes numerical problems in calculating standard errors. As a consequence of slow convergence and the numerical problems encountered, the maximum value of γ is constrained to 100. This serves as an adequate approximation, since the transition function changes little beyond $\gamma \geq 100$, as Figure 1 indicates. Note, however, that in these cases the estimated standard errors are conditional on assuming $\gamma = 100$. Third, STCC–GARCH models are only estimated for data for which the constant correlations hypothesis is rejected. Finally, all computations have been performed using Ox, version 3.30; see Doornik (2002).

Instead of presenting all estimation and testing results we focus on specific examples that illustrate the behaviour of the dynamic interactive structure of the models. We begin by considering the bivariate models. Results from higher-dimensional models are discussed thereafter. The estimation and testing results for selected combinations of assets are presented in Tables 4 – 7.

When the transition variable is a function of lagged absolute S&P 500 returns, positive and negative returns of the same size have the same effect on the correlations and the absolute magnitude of the returns carries all the information of possible comovements in the returns. Small seven-day averages of absolute returns are associated with the conditional correlation matrix $\boldsymbol{P}_1$, whereas the large ones are related to $\boldsymbol{P}_2$.

In the F-GM model the estimated location of the transition is 0.24 which leaves 37% of the average absolute S&P 500 returns below the estimated location parameter c. In all the other estimated bivariate models the transition takes place around 0.5–0.8. This is a range such that 2–11% of the average absolute S&P 500 returns exceed the estimate of c. Except for the F-GM model, a feature common to every estimated bivariate model is that during market turbulence, indicated by a high value of the seven-day average absolute index return, the correlations are considerably higher than during calm periods. As to Ford and General Motors, the returns on these two stocks have fairly high conditional correlations in general, especially during periods when

[2] Without presenting the results, we note that the parameter estimates obtained from the two-step method differ somewhat from the estimates from the iterative method when an STCC–GARCH model is estimated, whereas the results are fairly similar under the assumption of constant correlations.

markets are *not* turbulent.

Turning to three-dimensional models in Table 6, the relationship between Ford and General Motors seems to be quite strong, and it dictates the estimated location to be around 0.24 as in the bivariate F-GM model. But then, in models that contain only one of the two automotive companies, the transition is estimated to take place around 0.7. The estimated correlations in the trivariate models behave in the same way as their counterparts in bivariate models whenever the estimated location of the transition coincides with the corresponding estimates from the two bivariate models. If this not the case, one pair of correlations dominates and the constancy of the remaining ones may not be rejected by a the partial constancy test. This is obvious from results reported in Table 6. Table 7 shows that the same pattern repeats itself in the four-variate models as well as in the five-variate model. In the five-variate model many of the correlations seem to be constant. Tests of partial constancy do not reject the null hypothesis that all the correlations except the ones between Ford and General Motors and Ford and Hewlett-Packard, respectively, are constant. The p–values for joint LM– and Wald tests of this hypothesis against the full STCC–GARCH model are 0.12 and 0.13, respectively. The parameter estimates of the restricted STCC–GARCH model can also be found in Table 7.

These correlations are quite different from the ones obtained from four-variate models with only one automotive company (Ford is excluded from the model given in Table 7). In the four-variate model, the correlations increase with the degree of market turbulence. It seems that the F-GM relationship is so strong that it prevents the investigator from seeing other interesting details in the data. As a whole, comparing the four- and five-variate models suggests that a single transition function may not be enough when one wants to characterize time-variation in these correlations. Extensions to the model are possible but, however, less parsimonious than the original model. Besides, the standard STCC-GARCH model ensures each correlation matrix to be positive definite as long as $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are positive definite matrices. This property becomes difficult to retain if the number of transitions exceeds one, unless strong restrictions, such as assuming $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ block diagonal, are placed on these matrices.

The estimated time-varying conditional correlations from the five-variate model appear in Figure 4. Note that most of the correlations in this model are constant around 0.3–0.4. The F-GM correlations are considerably higher when the market is calm than when it is turbulent, whereas the other two remain almost constant over time at the level of about 0.3. These correlations are in marked contrast with the correlations from the four-variate GM-HPQ-IBM-TXN model shown in Figure 5. When the markets are calm, the correlations are constant around 0.3–0.4. When there is strong turbulence, the correlations increase substantially. The results from this model support the notion of strong comovement of returns during times of distress.

5.3 Effects of shock asymmetry on conditional correlations

In order to investigate how possible asymmetry in the way S&P 500 returns affect the correlations $100(p_{t-1}-p_{t-3})$ is selected to be the transition variable in the model. The results of the constancy tests appear in Table 4. The tests of constant correlations reject constancy for each model except for the bivariate F-GM one. An STCC–GARCH$(1,1)$ model is thus estimated for all the other combinations. The S&P 500 two-day returns that are lower than the estimated location imply a correlation state approaching that of $\boldsymbol{P}_1$, whereas the returns greater than the estimated location result in correlations closer to the other extreme state, $\boldsymbol{P}_2$.

The results for the bivariate models can be found in Table 8. The models not involving Texas Instruments have the location of transition such that less than 1% of the S&P 500 two-day returns have values below the estimate of the location parameter c. In these models large negative shocks induce strongly positive conditional correlations between the returns; otherwise the correlations remain positive but are less strong. When Texas Instruments is combined with Ford or General Motors the transition is estimated to take place close to zero. Negative two-day index returns induce correlations slightly higher than positive two-day index returns. Combining Texas Instruments with Hewlett-Packard or IBM results in models where the estimated location is such that 30% of the S&P 500 two-day returns are larger than the estimated parameter c. In those two models the correlations are slightly weaker but still positive when the two-day returns of the index exceed the estimated location.

Modelling higher-dimensional combinations is complicated by the restriction that only one location for change is allowed in the model. A selection of the results from the three-variate models appear in Table 9. The strongest relationship typically determines the estimate of the location parameter in the transition function, and the estimated correlations are adapted to that location. For example, the estimated location in the trivariate HPQ-IBM-TXN model is close to the corresponding estimates from the HPQ-TXN and IBM-TXN models and repeats the correlation patterns in those models. This means that the correlations between HPQ and IBM are now based on a transition function that is different from the one in the bivariate model for these two return series. Similarly, in the F-HPQ-TXN and GM-HPQ-TXN models the strongest relationship appears to be the one between F and HPQ in the former and GM and HPQ in the latter model as the location parameter obtains about the same estimate as in these bivariate models. Yet another outcome is the one where none of the relationships is much stronger than the others. This may result in a completely new location for the transition. Consider the F-GM-TXN model. It is seen from Table 9 that the corresponding bivariate models have very different correlation structures. Besides, constancy of the conditional correlations between F and GM was not rejected when tested. The estimated location for the transition equals 0.17, a value different from any other location estimate in bivariate and trivariate models. It can be seen from Table 10 that c is estimated close to this value in the five-variate model as well. This is also true for all four-variate models involving TXN (not reported in the table). Furthermore, there is a local maximum in the likelihood of the F-GM-HPQ-IBM model corresponding to that value of the location parameter.

It the second highest maximum found: the estimates corresponding to the highest maximum are reported in the table.

We again graph some of the conditional correlations. Figure 6 contains the correlations from the trivariate HPQ-IBM-TXN model. They are positive and fluctuate between 0.3 and 0.5. A completely different picture emerges from Figure 7 where F has been added to this trivariate combination to form a four-variate model. The correlations mostly remain unchanged, expect for a few occasions when the market has received a very strong negative shock. Note, however, that the F-GM correlation fluctuates very little, which is in line with the fact that the constancy of this conditional correlation was not rejected when tested in the bivariate framework. In Figure 8, the patterns of estimated correlations from the complete five-variate model are close to the ones shown in the trivariate model of Figure 6.

In theory, as a solution to the 'multilocation problem' one could generalize the STCC–GARCH model such that it would allow different slope and location parameters for each pair of correlations. However, as already mentioned, such an extension entails the statistical problem of ensuring positive definiteness of the correlation matrix at each point of time.

5.4 Comparison

We conclude this section with a brief informal comparison of the time-varying correlations implied by the STCC– and DCC–GARCH models. The former models are the four- and five-dimensional ones reported in the previous subsections. The corresponding five-dimensional DCC–GARCH$(1,1)$ model is estimated using the two-step estimation method of Engle (2002). The estimated GARCH equations in the DCC–GARCH model differ slightly from the ones in the STCC–GARCH models due to the two-step procedure, whereas the correlation dynamics are very persistent (for conciseness we do not present the estimation results).

As can be seen from Figure 5, in the four-variate STCC–GARCH model periods of strongly volatile periods are reflected by increased correlation levels. For example, during the stock market crash in November 1987 the correlations are high for a short period but return quickly to lower values. These shifts are also visible in Figure 7, indicating that they are, more precisely, resulting from large negative price movements rather than just high volatility. Figures 6 and 8 show that the higher the index returns, the lower the correlations. The rather turbulent period beginning in the late 1990s results in correlations that are more often in the 'high regime' than the previous ones.

Contrary to these results, the DCC–GARCH model suggests a very persistent response of correlations to large shocks. For example, the events of November 1987 lead to increased correlations, but the return to lower levels is very slow. Another notable fact is that the turbulent period from the late 1990s onwards does not seem to have much of an effect on the correlations, except that the correlations generally display an upward tendency at the very end of the observation period.

These two approaches thus lead to rather different conclusions about the conditional correlations between the return series. Since these correlations cannot be ob-

served, it is not possible to decide whether the STCC–GARCH or the DCC–GARCH models yield results that are closer to the 'truth'. In theory, testing the models against each other may be possible but at the same time computationally quite demanding. These models may also be compared by investigating their out-of-sample forecasting performance, which has not been done in this study.

6 Conclusions

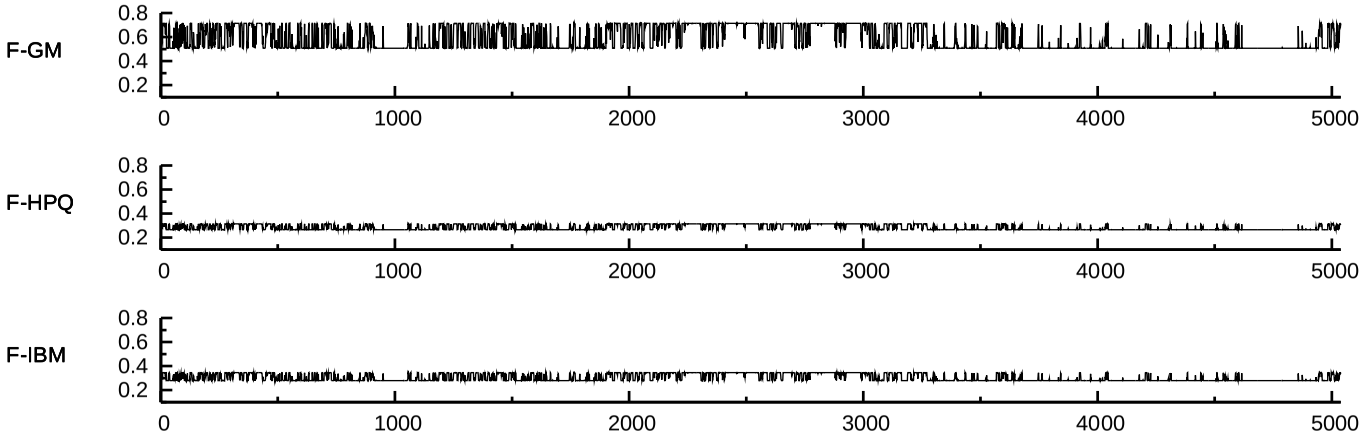
We have proposed a new multivariate conditional correlation model with time-varying correlations, the STCC–GARCH model. The conditional correlations are changing smoothly between two extreme states according to a transition variable that can be exogenous to the system. These correlations are weighted averages of two sets of constant correlations, which means that the corresponding time-varying correlation matrices are always positive definite on the condition that the two constant correlation matrices are positive definite.

The transition variable controlling the time-varying correlations can be chosen quite freely, depending on the modelling problem at hand. The STCC–GARCH model may thus be used for investigating the effects of numerous potential factors, endogenous as well as exogenous, on conditional correlations. In this respect the model differs from most other dynamic conditional correlation models such as the ones proposed by Tse and Tsui (2002), Engle (2002), and Pelletier (2006).

The STCC–GARCH model is applied to up to a five-variable set of daily returns of frequently traded stocks included in the S&P 500 index. When using the seven-day lagged average of the daily absolute return of the index as the transition variable we find that the conditional correlations are substantially higher during periods of high volatility than otherwise. Asymmetric response of correlations to shocks is examined using the one-day lag of the two-day index returns. In that case very large negative returns on the index imply very high conditional correlations between the volatilities.

In its present form the model allows for a single transition with location and smoothness parameters common to all series. In theory this restriction can be relaxed, but finding a useful way of doing it is left for future work. The model may be further refined by allowing specifications of the univariate GARCH equations beyond the standard GARCH$(1,1)$ model. An extension in the spirit of the multivariate CCC–GARCH model of Jeantheau (1998) would be an interesting alternative. Then the squared returns would be linked not only through the conditional correlations but also through the GARCH equations. Another point worth considering is incorporating higher-frequency data into the model. Recent research has emphasized the importance of information that is present in the high-frequency data but lost in aggregation. One such possibility would be to use the realized volatility or bipower variation of stock index returns over a day or a number of days as the transition variable in a model for stock returns. This possibility is left for future research.

Figure 4: The estimated time-varying conditional correlations from the five-variable STCC–model when the transition variable is lagged absolute S&P 500 returns averaged over seven days, where some of the correlations restricted constant, see Table 7. The estimated location is $\hat{c} = 0.24$.



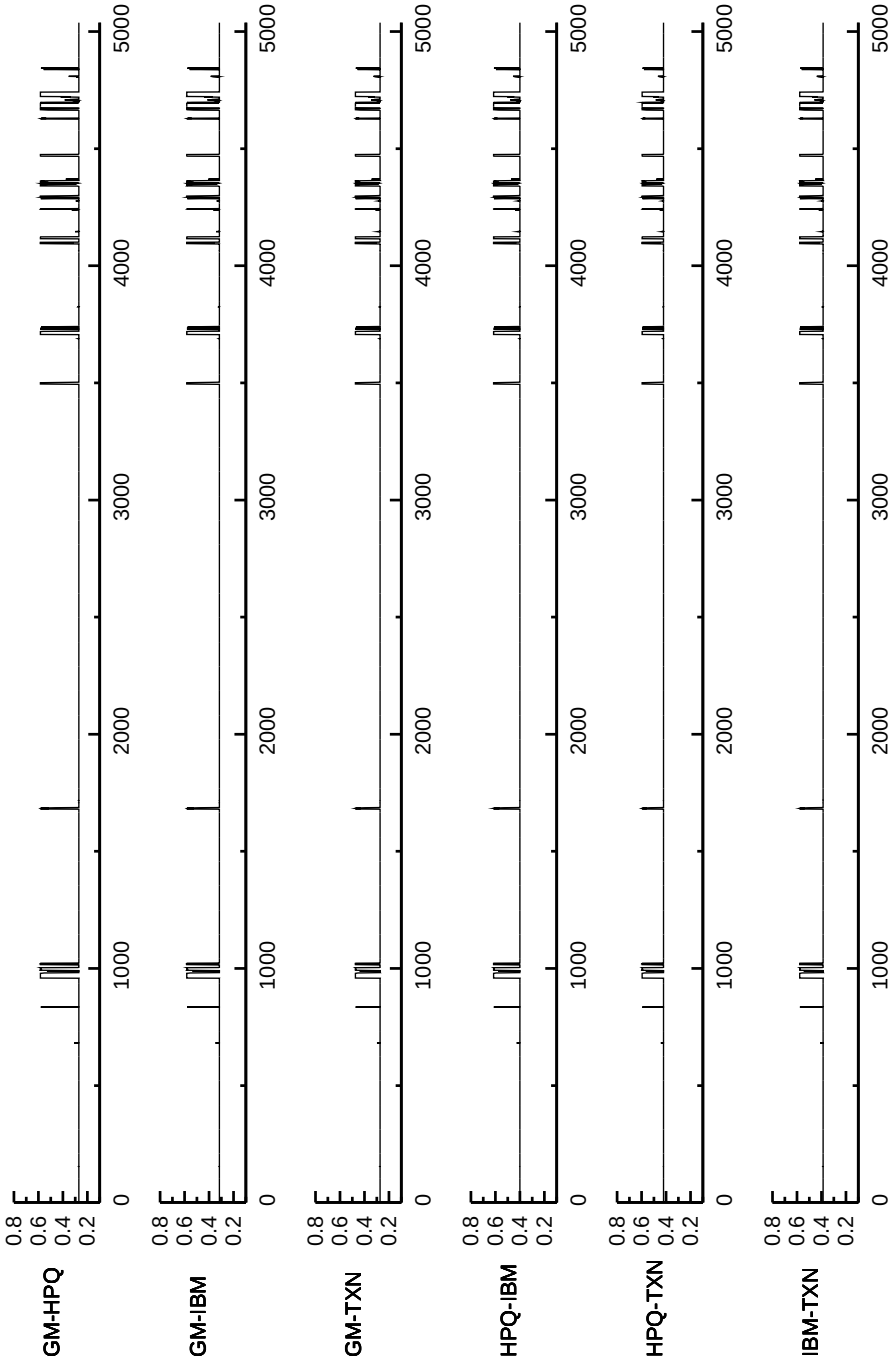
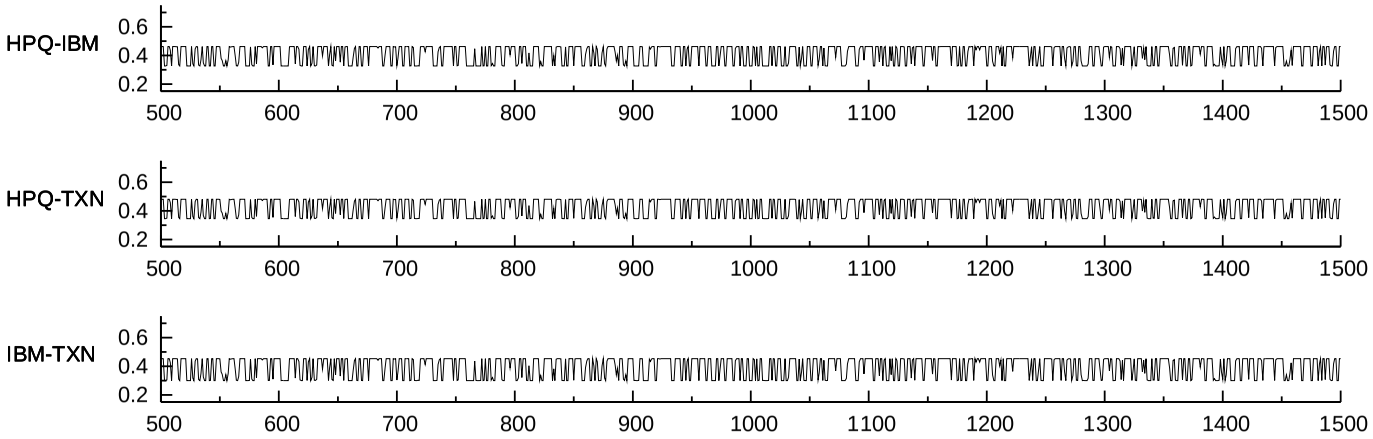


Figure 5: The estimated conditional correlations from the four-variate (GM-HPQ-IBM-TXN) STCC–GARCH model model from Table 7 when the transition variable is lagged absolute S&P 500 returns averaged over seven days. The estimated location is $\hat{c} = 0.71$.

Figure 6: The estimated conditional correlations from the three-variable (HPQ-IBM-TXN) STCC–model when the transition variable is a lagged S&P 500 return over two days, see Table 9. The estimated location is $\hat{c} = 0.26$. The time period covers the years from 1986 to 1989.



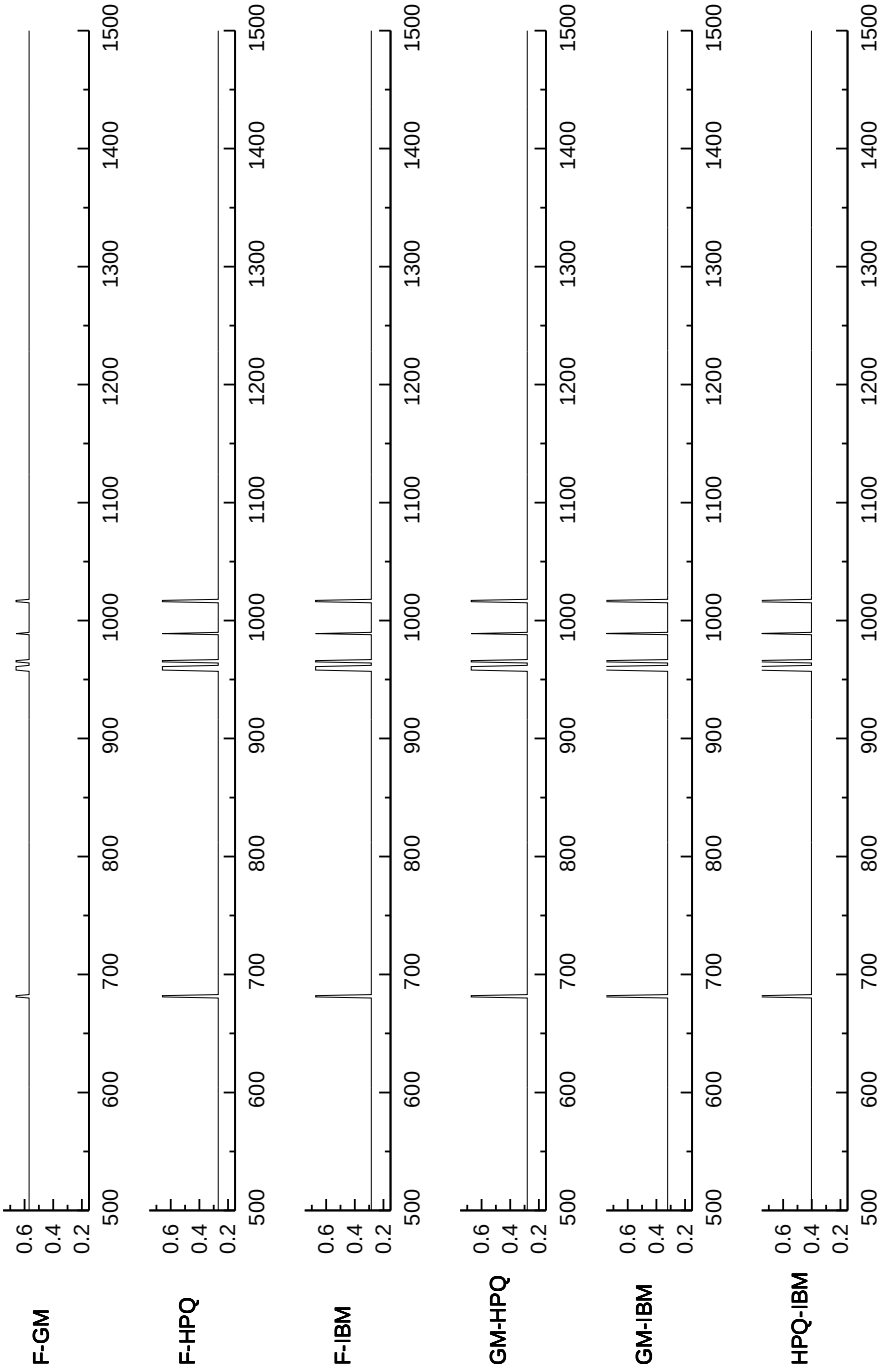


Figure 7: The estimated conditional correlations from the four-variable (F-GM-HPQ-IBM) STCC–model when the transition variable is a lagged S&P 500 return over two days, see Table 10. The estimated location is $\hat{c} = -1.97$. The time period covers the years from 1986 to 1989.

Figure 8: The estimated conditional correlations from the five-variable STCC–model when the transition variable is a lagged S&P 500 return over two days, see Table 10. The estimated location is $\hat{c} = 0.19$. The time period covers the years from 1986 to 1989.

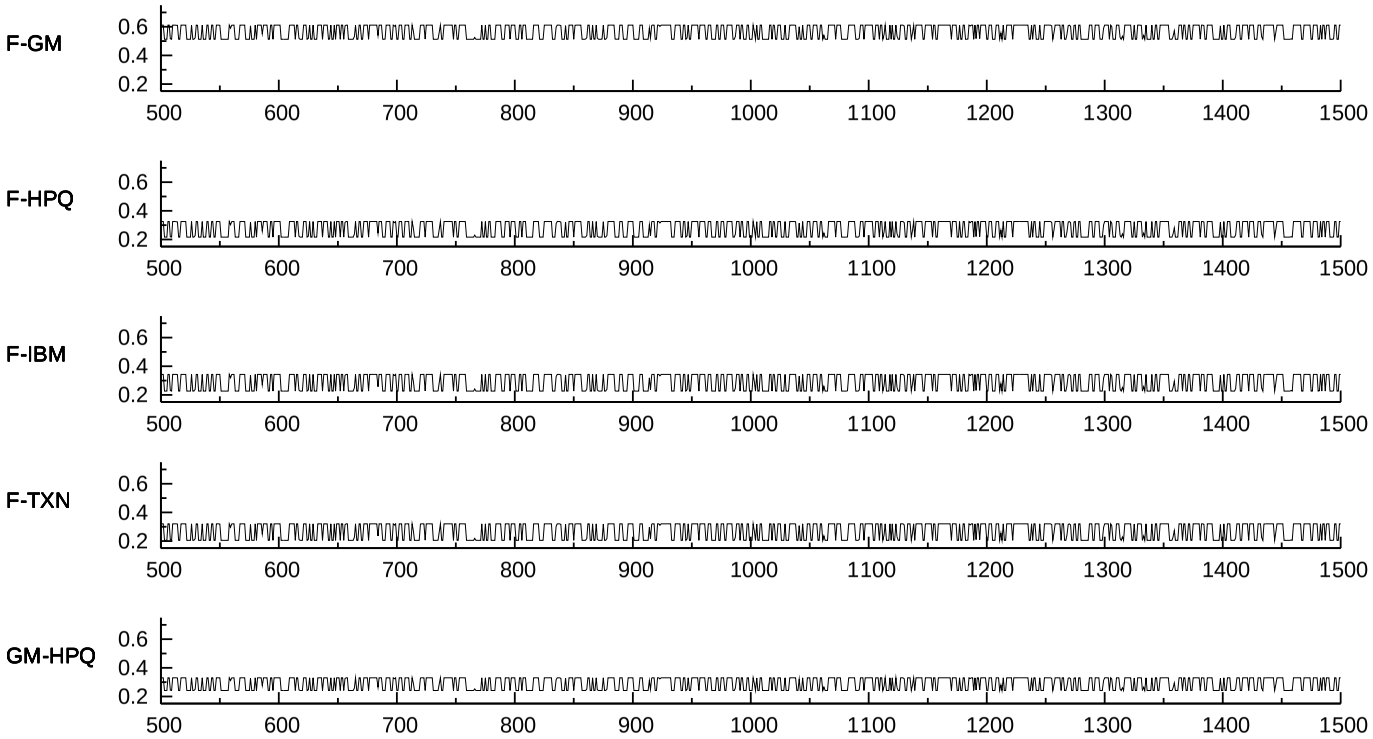


Figure 8 (continued)

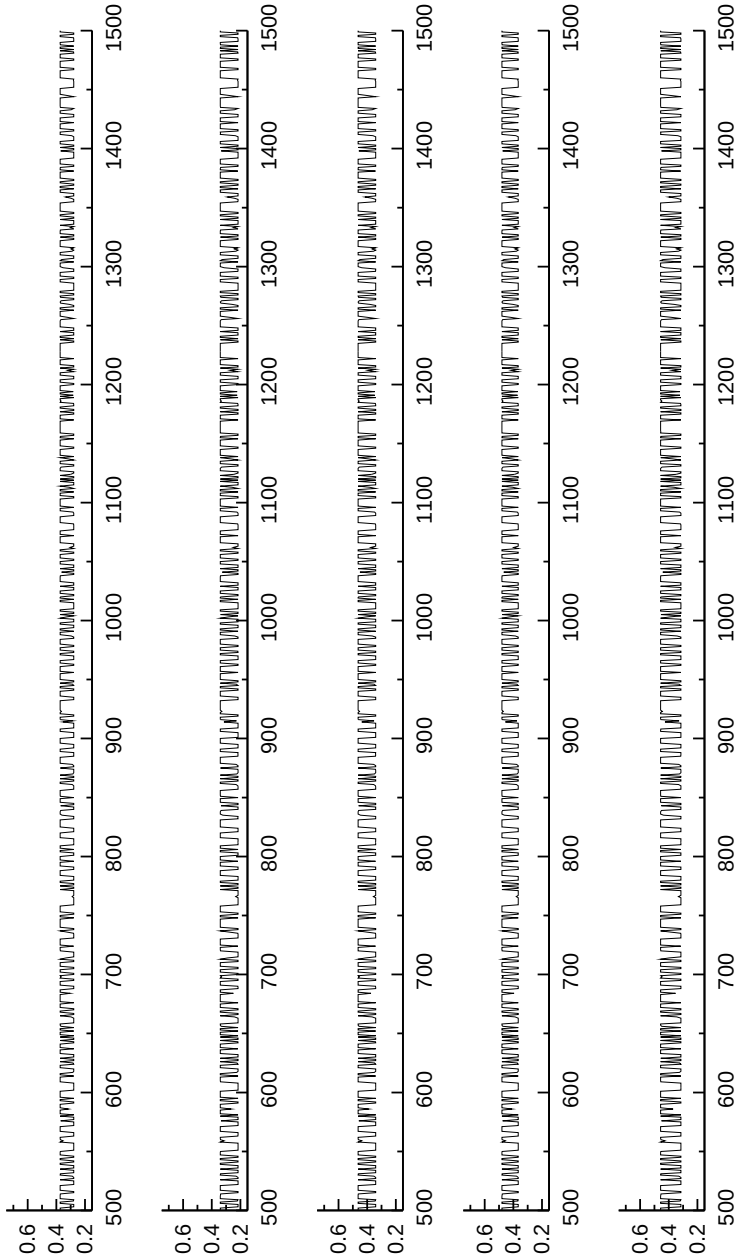
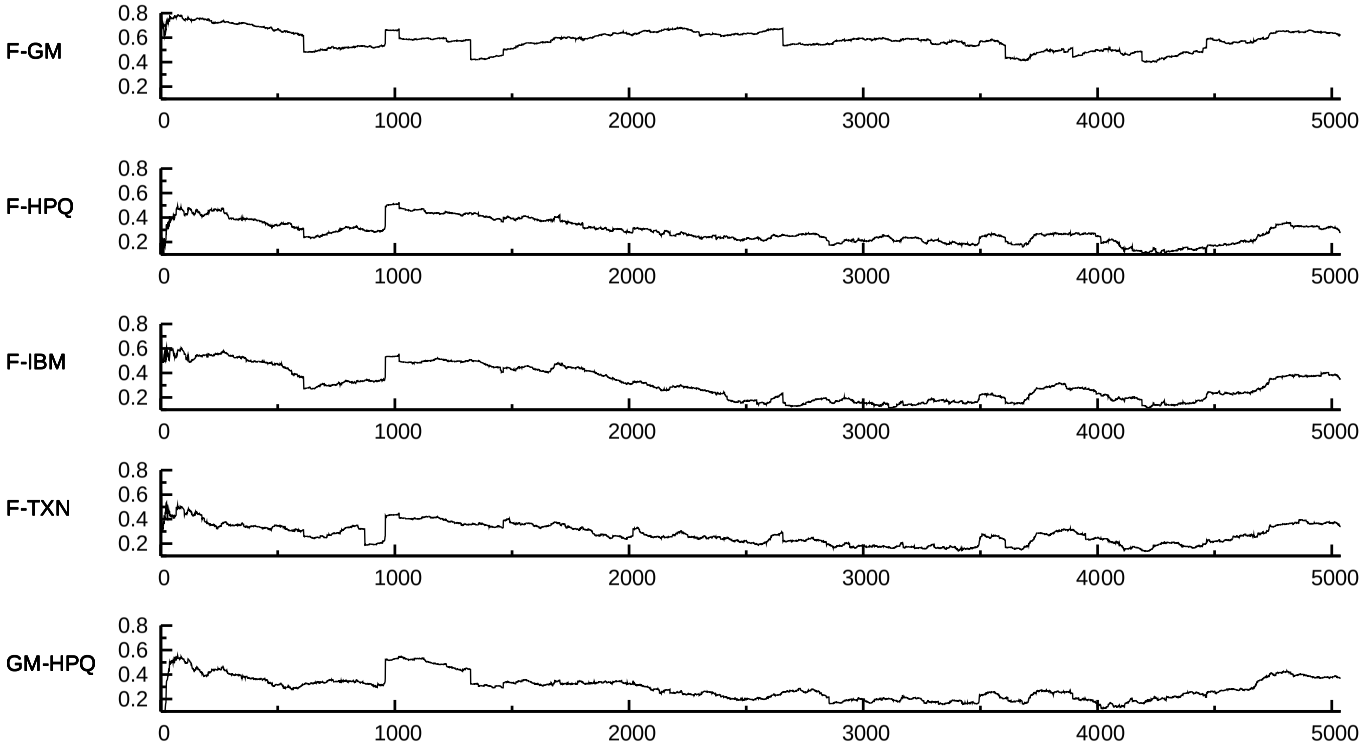


Figure 9: The estimated conditional correlations from the five-variable DCC–model.



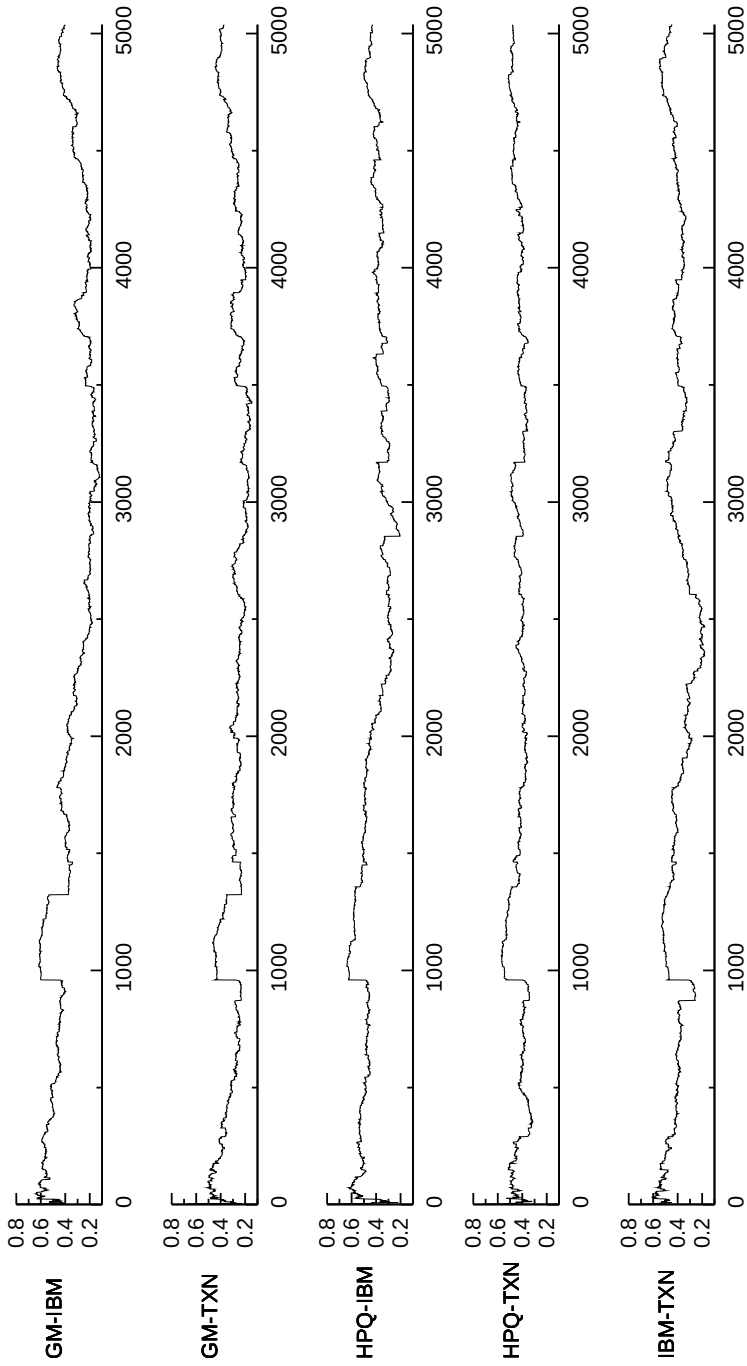


Figure 9 (continued)

	abbr.	min	max	mean (b.t.)	mean (a.t.)	st.dev (b.t.)	st.dev (a.t.)	skewness (b.t.)	skewness (a.t.)	kurtosis (b.t.)	kurtosis (a.t.)	no. of trunc.
Ford	F	−30.44	6.30	−0.0084	0.0062	1.173	0.956	−9.076	−1.162	215.176	18.220	5
General Motors	GM	−30.23	5.93	−0.0028	0.0016	0.950	0.860	−6.720	−0.739	210.671	13.704	2
Hewlett-Packard	HPQ	−29.23	6.93	−0.0053	0.0046	1.317	1.154	−5.841	−0.672	125.323	11.433	3
IBM	IBM	−30.89	5.37	−0.0023	0.0061	1.031	0.854	−10.404	−0.919	306.696	19.480	3
Texas Instruments	TXN	−48.37	9.36	−0.0115	0.0150	1.704	1.308	−9.794	−0.393	230.166	9.379	6
S&P500 Index	–	−9.94	3.78	0.0164	–	0.473	–	−2.012	–	43.194	–	–
Transition variable 1	$s_t^{(1)}$	0	3.39	0.3202	–	0.198	–	4.707	–	53.354	–	–
Transition variable 2	$s_t^{(2)}$	−12.20	6.00	0.0328	–	0.673	–	−1.584	–	31.173	–	–

Table 3: Descriptive statistics of the asset returns. The mean, standard deviation, skewness and kurtosis are reported both before (b.t.) and after (a.t.) removing the extreme negative returns. The cutoff value for truncation corresponds roughly to 10 standard deviations. The transition variables are $s_t^{(1)}$, the lagged absolute S&P 500 index return averaged over past seven days, and $s_t^{(2)}$, the lagged S&P 500 index return over past two days.

	$s_t^{(1)}$		$s_t^{(2)}$	
	LM_{CCC}	p–value	LM_{CCC}	p–value
F – GM	36.65	1×10^{-9}	2.80	0.0943
F – HPQ	9.12	0.0025	20.95	5×10^{-6}
F – IBM	8.21	0.0042	45.59	1×10^{-11}
F – TXN	14.66	0.0001	22.01	3×10^{-6}
GM – HPQ	14.74	0.0001	30.34	4×10^{-8}
GM – IBM	24.80	6×10^{-7}	53.43	3×10^{-13}
GM – TXN	14.12	0.0002	35.35	3×10^{-9}
HPQ – IBM	11.81	0.0008	37.12	1×10^{-9}
HPQ – TXN	7.96	0.0048	21.45	4×10^{-6}
IBM – TXN	13.24	0.0003	48.31	4×10^{-12}
F – GM – HPQ	64.06	8×10^{-14}	30.15	1×10^{-6}
F – GM – IBM	66.57	2×10^{-14}	61.72	3×10^{-13}
F – GM – TXN	75.49	3×10^{-16}	29.71	2×10^{-6}
F – HPQ – IBM	10.98	0.0118	67.54	1×10^{-14}
F – HPQ – TXN	11.20	0.0107	36.88	5×10^{-8}
F – IBM – TXN	16.08	0.0011	77.08	1×10^{-16}
GM – HPQ – IBM	24.47	2×10^{-5}	76.99	1×10^{-16}
GM – HPQ – TXN	14.59	0.0022	50.99	5×10^{-11}
GM – IBM – TXN	26.29	8×10^{-6}	88.13	6×10^{-19}
HPQ – IBM – TXN	11.97	0.0075	63.40	1×10^{-13}
F – GM – HPQ – IBM	77.52	1×10^{-14}	78.67	7×10^{-15}
F – GM – HPQ – TXN	73.64	7×10^{-14}	44.27	7×10^{-8}
F – GM – IBM – TXN	80.47	3×10^{-15}	87.07	1×10^{-16}
F – HPQ – IBM – TXN	8.22	0.2221	83.39	7×10^{-16}
GM – HPQ – IBM – TXN	17.24	0.0085	94.43	4×10^{-18}
F – GM – HPQ – IBM – TXN	81.40	3×10^{-13}	90.24	5×10^{-15}

Table 4: Test of constant conditional correlation against STCC–GARCH model for all combinations of assets. The transition variables are $s_t^{(1)}$, the lagged absolute S&P 500 index returns averaged over seven days, and $s_t^{(2)}$, a lagged S&P 500 index return over two days.

model	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	
F	0.0071	0.0166	0.9760	0.7157	0.5075	$\hat{c} = 0.2395$
	(0.0013)	(0.0020)	(0.0026)	(0.0116)	(0.0127)	(0.0042)
GM	0.0246	0.0380	0.9291			$\hat{\gamma} = 100$
	(0.0059)	(0.0057)	(0.0129)			
F	0.0059	0.0198	0.9744	0.2636	0.4539	$\hat{c} = 0.7221$
	(0.0012)	(0.0025)	(0.0029)	(0.0138)	(0.0378)	(0.0182)
HPQ	0.0111	0.0194	0.9719			$\hat{\gamma} = 100$
	(0.0034)	(0.0039)	(0.0062)			
F	0.0065	0.0220	0.9719	0.2855	0.4332	$\hat{c} = 0.7154$
	(0.0014)	(0.0027)	(0.0032)	(0.0136)	(0.0403)	(0.0325)
IBM	0.0058	0.0703	0.9277			$\hat{\gamma} = 100$
	(0.0017)	(0.0089)	(0.0093)			
F	0.0065	0.0210	0.9727	0.2560	0.3568	$\hat{c} = 0.5076$
	(0.0014)	(0.0027)	(0.0032)	(0.0146)	(0.0297)	(0.0184)
TXN	0.0102	0.0324	0.9612			$\hat{\gamma} = 100$
	(0.0023)	(0.0043)	(0.0050)			
GM	0.0264	0.0474	0.9162	0.2688	0.6000	$\hat{c} = 0.7142$
	(0.0063)	(0.0066)	(0.0144)	(0.0136)	(0.0350)	(0.0120)
HPQ	0.0171	0.0253	0.9614			$\hat{\gamma} = 100$
	(0.0059)	(0.0059)	(0.0101)			
GM	0.0218	0.0440	0.9263	0.3179	0.6823	$\hat{c} = 0.8413$
	(0.0047)	(0.0055)	(0.0110)	(0.0130)	(0.0364)	(0.0173)
IBM	0.0061	0.0706	0.9270			$\hat{\gamma} = 100$
	(0.0017)	(0.0085)	(0.0090)			
GM	0.0259	0.0467	0.9178	0.2687	0.5054	$\hat{c} = 0.6628$
	(0.0058)	(0.0063)	(0.0133)	(0.0138)	(0.0363)	(0.0121)
TXN	0.0125	0.0352	0.9579			$\hat{\gamma} = 100$
	(0.0028)	(0.0047)	(0.0057)			
HPQ	0.0271	0.0351	0.9443	0.3989	0.6503	$\hat{c} = 0.6867$
	(0.0104)	(0.0092)	(0.0166)	(0.0123)	(0.0321)	(0.0178)
IBM	0.0081	0.0762	0.9187			$\hat{\gamma} = 100$
	(0.0020)	(0.0092)	(0.0100)			
HPQ	0.0270	0.0333	0.9458	0.4205	0.6044	$\hat{c} = 0.6827$
	(0.0097)	(0.0085)	(0.0154)	(0.0121)	(0.0334)	(0.0215)
TXN	0.0155	0.0361	0.9548			$\hat{\gamma} = 100$
	(0.0034)	(0.0051)	(0.0065)			
IBM	0.0093	0.0767	0.9162	0.3871	0.6001	$\hat{c} = 0.6966$
	(0.0022)	(0.0094)	(0.0105)	(0.0124)	(0.0351)	(0.0168)
TXN	0.0144	0.0369	0.9550			$\hat{\gamma} = 100$
	(0.0030)	(0.0047)	(0.0058)			

Table 5: Estimation results for all bivariate STCC–GARCH models (standard errors in parentheses) when the transition variable is a lagged absolute S&P 500 index returns averaged over seven days.

model	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$		$\hat{\boldsymbol{P}}_1$		$\hat{\boldsymbol{P}}_2$		
F	0.0071 (0.0012)	0.0169 (0.0020)	0.9758 (0.0026)		F	GM	F	GM	
GM	0.0203 (0.0044)	0.0337 (0.0046)	0.9392 (0.0099)	GM	0.7169 (0.0116)		0.5101 (0.0126)		$\hat{c} = 0.2376$ (0.0043)
IBM	0.0053 (0.0015)	0.0678 (0.0085)	0.9310 (0.0088)	IBM	0.3161^r (0.0235)	0.3037^r (0.0230)	0.2914^r (0.0157)	0.3575^r (0.0152)	$\hat{\gamma} = 100$
F	0.0071 (0.0012)	0.0163 (0.0019)	0.9762 (0.0026)		F	GM	F	GM	
GM	0.0230 (0.0052)	0.0352 (0.0050)	0.9335 (0.0114)	GM	0.7149 (0.0117)		0.5068 (0.0127)		$\hat{c} = 0.2394$ (0.0043)
TXN	0.0116 (0.0016)	0.0327 (0.0044)	0.9608 (0.0053)	TXN	0.2726^r (0.0249)	0.2943^r (0.0235)	0.2755^r (0.0159)	0.2898^r (0.0160)	$\hat{\gamma} = 100$
F	0.0065 (0.0013)	0.0212 (0.0026)	0.9725 (0.0031)		F	IBM	F	IBM	
IBM	0.0069 (0.0018)	0.0688 (0.0089)	0.9267 (0.0095)	IBM	0.2858 (0.0136)		0.4130 (0.0413)		$\hat{c} = 0.7010$ (0.0176)
TXN	0.0130 (0.0027)	0.0354 (0.0043)	0.9572 (0.0053)	TXN	0.2610 (0.0139)	0.3875 (0.0124)	0.3928 (0.0394)	0.5826 (0.0374)	$\hat{\gamma} = 100$
GM	0.0218 (0.0046)	0.0426 (0.0054)	0.9276 (0.0108)		GM	IBM	GM	IBM	
IBM	0.0069 (0.0018)	0.0688 (0.0084)	0.9265 (0.0092)	IBM	0.3157 (0.0132)		0.5811 (0.0381)		$\hat{c} = 0.6874$ (0.0167)
TXN	0.0144 (0.0029)	0.0369 (0.0046)	0.9549 (0.0056)	TXN	0.2716 (0.0138)	0.3863 (0.0124)	0.4954 (0.0385)	0.5851 (0.0376)	$\hat{\gamma} = 100$

Table 6: Estimation results for selected combinations of trivariate STCC–GARCH models (standard errors in parentheses) when the transition variable is a lagged absolute S&P 500 index returns averaged over seven days. The first two models failed to reject the hypothesis of partially constant correlations with respect to the parameters indicated by a superscript r.

model	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$		$\hat{\boldsymbol{P}}_1$				$\hat{\boldsymbol{P}}_2$				
					GM	HPQ	IBM		GM	HPQ	IBM		
GM	0.0225	0.0426	0.9265										
	(0.0049)	(0.0054)	(0.0113)										
HPQ	0.0253	0.0314	0.9489	HPQ	0.2701				0.5841				$\hat{c} = 0.7052$
	(0.0094)	(0.0080)	(0.0149)		(0.0135)				(0.0372)				(0.0141)
IBM	0.0070	0.0684	0.9267	IBM	0.3162	0.4002			0.5819	0.6153			$\hat{\gamma} = 100$
	(0.0018)	(0.0083)	(0.0091)		(0.0131)	(0.0122)			(0.0382)	(0.0368)			
TXN	0.0168	0.0375	0.9525	TXN	0.2736	0.4209	0.3877		0.4748	0.5955	0.5784		
	(0.0033)	(0.0048)	(0.0061)		(0.0136)	(0.0120)	(0.0124)		(0.0399)	(0.0349)	(0.0376)		
					F	GM	HPQ	IBM	F	GM	HPQ	IBM	
F	0.0072	0.0160	0.9765										
	(0.0012)	(0.0019)	(0.0025)										
GM	0.0207	0.0326	0.9398	GM	0.7146				0.5068				
	(0.0046)	(0.0046)	(0.0100)		(0.0117)				(0.0128)				
HPQ	0.0199	0.0244	0.9603	HPQ	0.3399	0.3312			0.2585	0.2826			$\hat{c} = 0.2396$
	(0.0084)	(0.0072)	(0.0133)		(0.0244)	(0.0234)			(0.0157)	(0.0156)			(0.0044)
IBM	0.0064	0.0648	0.9315	IBM	0.3293	0.3178	0.4173		0.2857	0.3523	0.4144		$\hat{\gamma} = 100$
	(0.0017)	(0.0083)	(0.0089)		(0.0228)	(0.0223)	(0.0202)		(0.0159)	(0.0154)	(0.0145)		
TXN	0.0153	0.0335	0.9575	TXN	0.2763	0.2945	0.4352	0.3714	0.2742	0.2890	0.4300	0.4193	
	(0.0030)	(0.0043)	(0.0055)		(0.0246)	(0.0232)	(0.0199)	(0.0201)	(0.0159)	(0.0160)	(0.0140)	(0.0147)	
					F	GM	HPQ	IBM	F	GM	HPQ	IBM	
F	0.0072	0.0161	0.9764										
	(0.0012)	(0.0019)	(0.0025)										
GM	0.0207	0.0326	0.9397	GM	0.7147				0.5068				
	(0.0046)	(0.0046)	(0.0101)		(0.0117)				(0.0127)				
HPQ	0.0193	0.0247	0.9604	HPQ	0.3153	0.2982^R			0.2659	0.2982^R			$\hat{c} = 0.2394$
	(0.0083)	(0.0073)	(0.0134)		(0.0198)	(0.0130)			(0.0148)	(0.0130)			(0.0042)
IBM	0.0065	0.0649	0.9314	IBM	0.3456	0.3409^R	0.4169^R		0.2794	0.3403^R	0.4169^R		$\hat{\gamma} = 100$
	(0.0017)	(0.0083)	(0.0090)		(0.0180)	(0.0126)	(0.0117)		(0.0152)	(0.0126)	(0.0117)		
TXN	0.0153	0.0334	0.9575	TXN	0.2760^R	0.2935^R	0.4346^R	0.4026^R	0.2760^R	0.2935^R	0.4346^R	0.4026^R	
	(0.0030)	(0.0043)	(0.0055)		(0.0130)	(0.0128)	(0.0115)	(0.0119)	(0.0130)	(0.0128)	(0.0115)	(0.0119)	

Table 7: Estimation results for one four-variate and the five-variate STCC–GARCH models (standard errors in parentheses) when the transition variable is a lagged absolute S&P 500 index returns averaged over seven days. The superscript R indicates that the correlation is restricted to be constant.

model	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}_1$	$\hat{\rho}_2$	
F	0.0060	0.0193	0.9747	0.6992	0.2662	$\hat{c} = -1.9766$
	(0.0012)	(0.0024)	(0.0029)	(0.0441)	(0.0134)	(0.0292)
HPQ	0.0104	0.0184	0.9734			$\hat{\gamma} = 100$
	(0.0030)	(0.0034)	(0.0054)			
F	0.0065	0.0209	0.9726	0.7159	0.2829	$\hat{c} = -2.1172$
	(0.0013)	(0.0026)	(0.0031)	(0.0442)	(0.0132)	(0.0582)
IBM	0.0059	0.0707	0.9272			$\hat{\gamma} = 100$
	(0.0017)	(0.0086)	(0.0091)			
F	0.0067	0.0209	0.9726	0.3347	0.2145	$\hat{c} = 0.0139$
	(0.0014)	(0.0027)	(0.0032)	(0.0176)	(0.0189)	(0.0225)
TXN	0.0102	0.0320	0.9625			$\hat{\gamma} = 100$
	(0.0023)	(0.0042)	(0.0050)			
GM	0.0263	0.0463	0.9174	0.7160	0.2777	$\hat{c} = -1.9667$
	(0.0060)	(0.0062)	(0.0136)	(0.0430)	(0.0133)	(0.0329)
HPQ	0.0142	0.0218	0.9670			$\hat{\gamma} = 100$
	(0.0046)	(0.0047)	(0.0079)			
GM	0.0205	0.0414	0.9302	0.8898	0.3048	$\hat{c} = -2.2036$
	(0.0042)	(0.0048)	(0.0098)	(0.0459)	(0.0145)	(0.2217)
IBM	0.0058	0.0694	0.9284			$\hat{\gamma} = 1.92$
	(0.0016)	(0.0082)	(0.0087)			
GM	0.0250	0.0462	0.9199	0.4290	0.1995	$\hat{c} = -0.1817$
	(0.0055)	(0.0061)	(0.0126)	(0.0300)	(0.0268)	(0.1552)
TXN	0.0116	0.0334	0.9602			$\hat{\gamma} = 3.95$
	(0.0026)	(0.0044)	(0.0054)			
HPQ	0.0204	0.0276	0.9563	0.7846	0.4012	$\hat{c} = -1.7490$
	(0.0074)	(0.0067)	(0.0122)	(0.0325)	(0.0121)	(0.0185)
IBM	0.0085	0.0779	0.9166			$\hat{\gamma} = 100$
	(0.0021)	(0.0094)	(0.0103)			
HPQ	0.0280	0.0336	0.9450	0.4786	0.3372	$\hat{c} = 0.2946$
	(0.0113)	(0.0097)	(0.0178)	(0.0128)	(0.0218)	(0.0123)
TXN	0.0149	0.0336	0.9576			$\hat{\gamma} = 100$
	(0.0033)	(0.0048)	(0.0062)			
IBM	0.0098	0.0775	0.9150	0.4468	0.2845	$\hat{c} = 0.3514$
	(0.0023)	(0.0097)	(0.0109)	(0.0131)	(0.0241)	(0.0120)
TXN	0.0136	0.0333	0.9588			$\hat{\gamma} = 100$
	(0.0028)	(0.0043)	(0.0053)			

Table 8: Estimation results for all bivariate STCC–GARCH models (standard errors in parentheses) when the transition variable is a lagged S&P 500 index return over two days.

model	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$		$\hat{\boldsymbol{P}}_1$		$\hat{\boldsymbol{P}}_2$		
F	0.0067	0.0176	0.9755		F	GM	F	GM	
	(0.0012)	(0.0020)	(0.0025)						
GM	0.0184	0.0340	0.9407	GM	0.6118		0.5130		$\hat{c} = 0.1725$
	(0.0036)	(0.0042)	(0.0085)		(0.0108)		(0.0158)		(0.0283)
TXN	0.0113	0.0327	0.9611	TXN	0.3212	0.3409	0.2049	0.2161	$\hat{\gamma} = 100$
	(0.0025)	(0.0043)	(0.0051)		(0.0162)	(0.0160)	(0.0214)	(0.0163)	
F	0.0062	0.0192	0.9745		F	HPQ	F	HPQ	
	(0.0013)	(0.0024)	(0.0029)						
HPQ	0.0127	0.0196	0.9702	HPQ	0.7007		0.2662		$\hat{c} = -1.9711$
	(0.0038)	(0.0039)	(0.0066)		(0.0447)		(0.0134)		(0.0198)
TXN	0.0135	0.0327	0.9591	TXN	0.5253	0.7732	0.2629	0.4228	$\hat{\gamma} = 100$
	(0.0028)	(0.0066)	(0.0055)		(0.0561)	(0.0391)	(0.0134)	(0.0118)	
GM	0.0266	0.0454	0.9177		GM	HPQ	GM	HPQ	
	(0.0059)	(0.0060)	(0.0132)						
HPQ	0.0175	0.0237	0.9623	HPQ	0.7092		0.2785		$\hat{c} = -1.9632$
	(0.0059)	(0.0054)	(0.0097)		(0.0450)		(0.0132)		(0.0240)
TXN	0.0155	0.0346	0.9559	TXN	0.5730	0.7687	0.2785	0.4237	$\hat{\gamma} = 100$
	(0.0032)	(0.0047)	(0.0061)		(0.0536)	(0.0410)	(0.0133)	(0.0117)	
HPQ	0.0337	0.0369	0.9372		HPQ	IBM	HPQ	IBM	
	(0.0168)	(0.0131)	(0.0254)						
IBM	0.0095	0.0757	0.9170	IBM	0.4619		0.3261		$\hat{c} = 0.2559$
	(0.0022)	(0.0095)	(0.0107)		(0.0136)		(0.0216)		(0.0283)
TXN	0.0160	0.0338	0.9566	TXN	0.4811	0.4536	0.3459	0.3003	$\hat{\gamma} = 100$
	(0.0032)	(0.0044)	(0.0058)		(0.0132)	(0.0138)	(0.0219)	(0.0223)	

Table 9: Estimation results for selected combinations of trivariate STCC–GARCH models (standard errors in parentheses) when the transition variable is a lagged S&P 500 index return over two days.

model	$\hat{\alpha}_0$	$\hat{\alpha}$	$\hat{\beta}$		$\hat{\boldsymbol{P}}_1$				$\hat{\boldsymbol{P}}_2$				
					F	GM	HPQ		F	GM	HPQ		
F	0.0062	0.0171	0.9764										
	(0.0011)	(0.0019)	(0.0023)										
GM	0.0172	0.0312	0.9446	GM	0.6576^r				0.5683^r				$\hat{c} = -1.9731$
	(0.0033)	(0.0037)	(0.0077)		(0.0496)				(0.0097)				(0.0205)
HPQ	0.0117	0.0187	0.9718	HPQ	0.6319	0.6475			0.2679	0.2800			$\hat{\gamma} = 100$
	(0.0034)	(0.0036)	(0.0060)		(0.0588)	(0.0532)			(0.0133)	(0.0132)			
IBM	0.0050	0.0607	0.9366	IBM	0.6470	0.7440	0.7379		0.2843	0.3208	0.4026		
	(0.0014)	(0.0075)	(0.0081)		(0.0498)	(0.0394)	(0.0442)		(0.0132)	(0.0129)	(0.0120)		
					F	GM	HPQ	IBM	F	GM	HPQ	IBM	
F	0.0067	0.0172	0.9759										
	(0.0011)	(0.0019)	(0.0024)										
GM	0.0170	0.0322	0.9444	GM	0.6113				0.5133				
	(0.0034)	(0.0040)	(0.0079)		(0.0108)				(0.0159)				
HPQ	0.0154	0.0218	0.9662	HPQ	0.3248	0.3302			0.2155	0.2409			$\hat{c} = 0.1850$
	(0.0059)	(0.0056)	(0.0099)		(0.0161)	(0.0160)			(0.0211)	(0.0207)			(0.0083)
IBM	0.0065	0.0650	0.9312	IBM	0.3430	0.3730	0.4638		0.2264	0.2779	0.3408		$\hat{\gamma} = 100$
	(0.0017)	(0.0084)	(0.0092)		(0.0155)	(0.0151)	(0.0137)		(0.0218)	(0.0210)	(0.0194)		
TXN	0.0144	0.0327	0.9587	TXN	0.3198	0.3399	0.4795	0.4565	0.2041	0.2165	0.3651	0.3152	
	(0.0028)	(0.0040)	(0.0051)		(0.0160)	(0.0157)	(0.0136)	(0.0138)	(0.0217)	(0.0214)	(0.0187)	(0.0203)	

Table 10: Estimation results for one four-variate and the five-variate STCC–GARCH model (standard errors in parentheses) when the transition variable is a lagged S&P 500 index return over two days. When testing the hypothesis of partially constant correlations in the four-variate model, Wald test fails to reject and LM–test barely rejects the constancy of the parameters indicated by a superscript r (p–value of the Wald statistic W_{PCCC} is 0.2646 and of the LM–statistic LM_{PCCC} is 0.0095).

Appendix

Construction of LM(/Wald)–statistic

Let $\boldsymbol{\theta}_0$ be the vector of true parameters. Under suitable assumptions and regularity conditions,

$$\sqrt{T}^{-1}\frac{\partial l\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\xrightarrow{d}N\left(0,\Im\left(\boldsymbol{\theta}_0\right)\right). \tag{13}$$

To derive LM–statistics of the constant conditional correlation hypothesis $\boldsymbol{\rho}_2^*=0$ consider the following quadratic form:

$$T^{-1}\frac{\partial l\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}\Im\left(\boldsymbol{\theta}_0\right)^{-1}\frac{\partial l\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}=T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}\right)\Im\left(\boldsymbol{\theta}_0\right)^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\right)$$

and evaluate it at the maximum likelihood estimators under the restriction $\boldsymbol{\rho}_2^*=0$. The limiting information matrix $\Im(\boldsymbol{\theta}_0)$ is replaced by the consistent estimator

$$\hat{\Im}_T\left(\boldsymbol{\theta}_0\right)=T^{-1}\sum_{t=1}^{T}E\left[\frac{\partial l_t\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}}\frac{\partial l_t\left(\boldsymbol{\theta}_0\right)}{\partial\boldsymbol{\theta}'}\mid\mathcal{F}_{t-1}\right]. \tag{14}$$

The following derivations are straightforward implications of the definitions and elementary rules of matrix algebra. Results in Anderson (2003) and Lütkepohl (1996) are heavily relied upon.

Test of constant conditional correlations against an STCC–GARCH model

To construct the test statistic we introduce some simplifying notation. Let $\boldsymbol{\omega}_i=(\alpha_{i0},\alpha_i,\beta_i)'$, $i=1,\ldots,N$, denote the parameter vectors of the GARCH equations, and $\boldsymbol{\rho}^*=(\boldsymbol{\rho}_1^{*\prime},\boldsymbol{\rho}_2^{*\prime})'$, where $\boldsymbol{\rho}_1^*=vecl\boldsymbol{P}_1^*$ and $\boldsymbol{\rho}_2^*=vecl\boldsymbol{P}_2^*$ are the vectors holding all the unique off-diagonal elements in the two matrices $\boldsymbol{P}_1^*$ and $\boldsymbol{P}_2^*$, respectively. The notation $vecl\boldsymbol{P}$ is used to denote the vec-operator applied to the strictly lower triangular part of the matrix $\boldsymbol{P}$. Let $\boldsymbol{\theta}=(\boldsymbol{\omega}_1',\ldots,\boldsymbol{\omega}_N',\boldsymbol{\rho}^{*\prime})'$ be the full parameter vector and $\boldsymbol{\theta}_0$ the corresponding vector of true parameters under the null. The linearized time-varying correlation matrix is $\boldsymbol{P}_t^*=\boldsymbol{P}_1^*-s_t\boldsymbol{P}_2^*$ as defined in (9). Furthermore, let $\boldsymbol{v}_{it}=(1,y_{it}^2,h_{it})'$, $i=1,\ldots,N$, and $\boldsymbol{v}_{\boldsymbol{\rho}^*t}=(1,-s_t)'$. Symbols $\otimes$ and $\odot$ represent the Kronecker and Hadamard products of two matrices, respectively. Let 1_i be a $N\times 1$ vector of zeros with ith element equal to one and $\mathbf{1}_n$ be a $n\times n$ matrix of ones. The identity matrix $\boldsymbol{I}$ is of size N unless otherwise indicated by a subscript.

Consider the log-likelihood function for observation t as defined in (7) with linearized time-varying correlation matrix:

$$l_t\left(\boldsymbol{\theta}\right)=-\frac{N}{2}\log\left(2\pi\right)-\frac{1}{2}\sum_{i=1}^{N}\log\left(h_{it}\right)-\frac{1}{2}\log\left|\boldsymbol{P}_t^*\right|-\frac{1}{2}\boldsymbol{z}_t'\boldsymbol{P}_t^{*-1}\boldsymbol{z}_t.$$

The first order derivatives of the log-likelihood function with respect to the GARCH and correlation parameters are

$$\begin{aligned}\frac{\partial l_t\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\omega}_i}&=-\frac{1}{2h_{it}}\frac{\partial h_{it}}{\partial\boldsymbol{\omega}_i}\left\{1-z_{it}1_i'\boldsymbol{P}_t^{*-1}\boldsymbol{z}_t\right\},\quad i=1,\ldots,N,\\ \frac{\partial l_t\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\rho}^*}&=-\frac{1}{2}\frac{\partial\left(vec\boldsymbol{P}_t^*\right)'}{\partial\boldsymbol{\rho}^*}\left\{vec\boldsymbol{P}_t^{*-1}-\left(\boldsymbol{P}_t^{*-1}\otimes\boldsymbol{P}_t^{*-1}\right)\left(\boldsymbol{z}_t\otimes\boldsymbol{z}_t\right)\right\},\end{aligned}$$

where

$$\begin{aligned}\frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i} &= \boldsymbol{v}_{i,t-1} + \beta_i \frac{\partial h_{i,t-1}}{\partial \boldsymbol{\omega}_i}, \quad i = 1, \ldots, N, \\ \frac{\partial \left(vec\boldsymbol{P}_t^*\right)'}{\partial \boldsymbol{\rho}^*} &= \boldsymbol{v}_{\boldsymbol{\rho}^* t} \otimes \boldsymbol{U}'.\end{aligned}$$

The matrix $\boldsymbol{U}$ is an $N^2 \times \frac{N(N-1)}{2}$ matrix of zeros and ones, whose columns are defined as

$$\left[vec\left(1_i 1_j' + 1_j 1_i'\right)\right]_{i=1,\ldots N-1, j=i+1,\ldots,N}$$

and the columns appear in the same order from left to right as the indices in $vecl\boldsymbol{P}_t$. Under the null hypothesis $\boldsymbol{\rho}_2^* = 0$, and thus the derivatives at the true parameter values under the null can be written as

$$\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} = -\frac{1}{2h_{it}} \frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \left\{1 - z_{it} 1_i' \boldsymbol{P}_1^{*-1} \boldsymbol{z}_t\right\}, \quad i = 1, \ldots, N, \tag{15}$$

$$\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^*} = -\frac{1}{2} \frac{\partial \left(vec\boldsymbol{P}_t^*(\boldsymbol{\theta}_0)\right)'}{\partial \boldsymbol{\rho}^*} \left\{vec\boldsymbol{P}_1^{*-1} - \left(\boldsymbol{P}_1^{*-1} \otimes \boldsymbol{P}_1^{*-1}\right)(\boldsymbol{z}_t \otimes \boldsymbol{z}_t)\right\}. \tag{16}$$

Taking conditional expectations of the cross products of (15) and (16) yields, for $i, j = 1, \ldots, N$,

$$\begin{aligned}
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i'}\right] &= \frac{1}{4h_{it}^2} \frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i'} \left(1 + 1_i' \boldsymbol{P}_1^{*-1} 1_i\right), \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_j'}\right] &= \frac{1}{4h_{it}h_{jt}} \frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \frac{\partial h_{jt}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_j'} \left(\rho_{1,ij}^* 1_i' \boldsymbol{P}_1^{*-1} 1_j\right), \quad i \neq j \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^*} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}\right] &= \frac{1}{4} \frac{\partial (vec\boldsymbol{P}_t^*(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}^*} \left(\boldsymbol{P}_1^{*-1} \otimes \boldsymbol{P}_1^{*-1} + \left(\boldsymbol{P}_1^{*-1} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_1^{*-1} \otimes \boldsymbol{I}\right)\right) \frac{\partial vec\boldsymbol{P}_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}, \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}\right] &= \frac{1}{4h_{it}} \frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} \left(1_i' \boldsymbol{P}_1^{*-1} \otimes 1_i' + 1_i' \otimes 1_i' \boldsymbol{P}_1^{*-1}\right) \frac{\partial vec\boldsymbol{P}_t^*(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}},
\end{aligned} \tag{17}$$

where

$$\boldsymbol{K} = \begin{bmatrix} 1_1 1_1' & \cdots & 1_N 1_1' \\ \vdots & \ddots & \vdots \\ 1_1 1_N' & \cdots & 1_N 1_N' \end{bmatrix}. \tag{18}$$

Expressions (17) for the conditional expectations follow from the fact that for a model with general correlation matrix $\boldsymbol{P}_t$,

$$\begin{aligned}
E_{t-1}\left[\boldsymbol{z}_t \boldsymbol{z}_t' \otimes \boldsymbol{z}_t \boldsymbol{z}_t'\right] &= \left(\boldsymbol{I}_{N^2} + \boldsymbol{K}\right)\left(\boldsymbol{P}_t \otimes \boldsymbol{P}_t\right) + vec\boldsymbol{P}_t \left(vec\boldsymbol{P}_t\right)' \\
&= \left(\boldsymbol{P}_t \otimes \boldsymbol{P}_t\right) + \left(\boldsymbol{I} \otimes \boldsymbol{P}_t\right) \boldsymbol{K} \left(\boldsymbol{I} \otimes \boldsymbol{P}_t\right) + vec\boldsymbol{P}_t \left(vec\boldsymbol{P}_t\right)'
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
E_{t-1}\left[z_{it} \boldsymbol{z}_t' \otimes \boldsymbol{z}_t \boldsymbol{z}_t'\right] &= E_{t-1}\left[\boldsymbol{z}_t \boldsymbol{z}_t' \otimes \boldsymbol{z}_t \boldsymbol{z}_t'\right]\left(1_i \otimes \boldsymbol{I}\right), \\
E_{t-1}\left[z_{it} z_{jt}' \otimes \boldsymbol{z}_t \boldsymbol{z}_t'\right] &= \left(1_i' \otimes \boldsymbol{I}\right) E_{t-1}\left[\boldsymbol{z}_t \boldsymbol{z}_t' \otimes \boldsymbol{z}_t \boldsymbol{z}_t'\right]\left(1_j \otimes \boldsymbol{I}\right),
\end{aligned}$$

see Anderson (2003). In the present case $\boldsymbol{P}_t$ is replaced with $\boldsymbol{P}_1^*$.

The estimator for the information matrix is obtained by making use of the submatrices in (17). For a more compact expression, let $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}', \ldots, \boldsymbol{x}_{Nt}')'$ where $\boldsymbol{x}_{it} = -\frac{1}{2h_{it}} \frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i}$, and let $\boldsymbol{x}_{\boldsymbol{\rho}^* t} = -\frac{1}{2} \boldsymbol{v}_{\boldsymbol{\rho}^* t} \otimes \boldsymbol{U}'$, and let $\boldsymbol{x}_{it}^0$, $i = 1, \ldots, N, \boldsymbol{\rho}^*$, denote the corresponding expressions

evaluated at the true values under the null hypothesis. Setting

$$\begin{aligned}
\boldsymbol{M}_1 &= T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_t^0\boldsymbol{x}_t^{0\prime}\odot\left(\left(\boldsymbol{I}+\boldsymbol{P}_1^*\odot\boldsymbol{P}_1^{*-1}\right)\otimes\mathbf{1}_3\right),\\
\boldsymbol{M}_2 &= T^{-1}\sum_{t=1}^{T}\begin{bmatrix}\boldsymbol{x}_{1t}^0 & & 0\\ & \ddots & \\ 0 & & \boldsymbol{x}_{Nt}^0\end{bmatrix}\begin{bmatrix}1_1'\boldsymbol{P}_1^{*-1}\otimes 1_1'+1_1'\otimes 1_1'\boldsymbol{P}_1^{*-1}\\ \vdots\\ 1_N'\boldsymbol{P}_1^{*-1}\otimes 1_N'+1_N'\otimes 1_N'\boldsymbol{P}_1^{*-1}\end{bmatrix}\boldsymbol{x}_{\boldsymbol{\rho}^*t}^{0\prime},\\
\boldsymbol{M}_3 &= T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{\boldsymbol{\rho}^*t}^0\left(\boldsymbol{P}_1^{*-1}\otimes\boldsymbol{P}_1^{*-1}+\left(\boldsymbol{P}_1^{*-1}\otimes\boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}_1^{*-1}\otimes\boldsymbol{I}\right)\right)\boldsymbol{x}_{\boldsymbol{\rho}^*t}^{0\prime},
\end{aligned}$$

the information matrix $\mathfrak{I}(\boldsymbol{\theta}_0)$ is approximated by

$$\begin{aligned}
\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0) &= T^{-1}\sum_{t=1}^{T}E\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\mid\mathcal{F}_{t-1}\right]\\
&= \begin{bmatrix}\boldsymbol{M}_1 & \boldsymbol{M}_2\\ \boldsymbol{M}_2' & \boldsymbol{M}_3\end{bmatrix}
\end{aligned}$$

The block corresponding to the correlation parameters of the inverse of $\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0)$ can be calculated as

$$\left(\boldsymbol{M}_3-\boldsymbol{M}_2'\boldsymbol{M}_1^{-1}\boldsymbol{M}_2\right)^{-1}$$

from where the south-east $\frac{N(N-1)}{2}\times\frac{N(N-1)}{2}$ block corresponding to $\boldsymbol{\rho}_2^*$ can be extracted. Replacing the true unknown values with maximum likelihood estimators, the test statistic simplifies to

$$LM = T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\rho}_2^{*\prime}}\right)\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_2^*,\boldsymbol{\rho}_2^*)}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\rho}_2^*}\right) \tag{20}$$

where $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho}_2^*,\boldsymbol{\rho}_2^*)}$ is the block of the inverse of $\hat{\mathfrak{I}}_T$ corresponding to those correlation parameters that are set to zero under the null. It follows from (13) and consistency and asymptotic normality of ML estimators that the statistic (20) has an asymptotic $\chi^2_{\frac{N(N-1)}{2}}$ distribution when the null hypothesis is valid.

Test of constant conditional correlations against partially constant STCC–GARCH model

In this case the null model is a CCC–GARCH model, and the alternative model is partially constant STCC–GARCH model. Let there be k pairs of variables with constant correlations in the alternative model. The test is as above, but with the following changes to definitions and notations. The linearized time-varying correlation matrix is $\boldsymbol{P}_t^* = \boldsymbol{P}_1^* - s_t\boldsymbol{P}_2^{P*}$, where $\boldsymbol{P}_2^{P*}$ is as $\boldsymbol{P}_2^*$ but with the elements corresponding to the constant correlations under the alternative set to zero. The vector of correlation parameters is $\boldsymbol{\rho}^* = (\boldsymbol{\rho}_1^{*\prime},\boldsymbol{\rho}_2^{P*\prime})'$, where $\boldsymbol{\rho}_2^{P*} = vecl\boldsymbol{P}_2^{P*}$ but with the elements corresponding to the constant correlations under the alternative being deleted. Furthermore,

$$\frac{\partial\left(vec\boldsymbol{P}_t^*\right)'}{\partial\boldsymbol{\rho}^*}$$

is as above, but with k rows deleted so that the remaining rows are corresponding to the elements in $\boldsymbol{\rho}^* = (\boldsymbol{\rho}_1^{*\prime},\boldsymbol{\rho}_2^{P*\prime})'$. The same rows are also deleted from $\boldsymbol{x}_{\boldsymbol{\rho}^*t}$. With these modifications the test statistic is as in (20) above, and the asymptotic distribution under the null hypothesis is χ^2_R where R is the number of restrictions to be tested.

Test of partially constant correlations against an STCC–GARCH model

To construct the test statistic we introduce some simplifying notation. Let $\boldsymbol{\omega}_i = (\alpha_{i0}, \alpha_i, \beta_i)'$, $i = 1, \ldots, N$, denote the parameter vectors of the GARCH equations, $\boldsymbol{\rho} = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2')'$, where $\boldsymbol{\rho}_1 = vecl\boldsymbol{P}_1$ and $\boldsymbol{\rho}_2 = vecl\boldsymbol{P}_2$, and $\boldsymbol{\varphi} = (c, \gamma)'$. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}_1', \ldots, \boldsymbol{\omega}_N', \boldsymbol{\rho}', \boldsymbol{\varphi}')'$ be the full parameter vector and $\boldsymbol{\theta}_0$ the corresponding vector of true parameters under the null. The time-varying correlation matrix is $\boldsymbol{P}_t = (1 - G_t)\boldsymbol{P}_1 + G_t\boldsymbol{P}_2$ as defined in (5) and (6). Furthermore, let $\boldsymbol{v}_{it} = (1, y_{it}^2, h_{it})'$, $i = 1, \ldots, N$, $\boldsymbol{v}_{\boldsymbol{\rho}t} = (1 - G_t, G_t)'$ and $\boldsymbol{v}_{\boldsymbol{\varphi}t} = (-\gamma, s_t - c)'$. Symbols $\otimes$ and $\odot$ represent the Kronecker and Hadamard products of two matrices, respectively. Let 1_i be a $1 \times N$ vector of zeros with ith element equal to one and $\mathbf{1}_n$ be a $n \times n$ matrix of ones. The identity matrix $\boldsymbol{I}$ is of size N unless otherwise indicated by a subscript.

Consider the log-likelihood function for observation t as defined in (7):

$$l_t(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\log(h_{it}) - \frac{1}{2}\log|\boldsymbol{P}_t| - \frac{1}{2}\boldsymbol{z}_t'\boldsymbol{P}_t^{-1}\boldsymbol{z}_t.$$

The elements of the score for observation t are

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\omega}_i} &= -\frac{1}{2h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta})}{\partial \boldsymbol{\omega}_i}\left\{1 - z_{it}1_i'\boldsymbol{P}_t^{-1}\boldsymbol{z}_t\right\}, \quad i = 1, \ldots, N, \\
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\rho}} &= -\frac{1}{2}\frac{\partial(vec\boldsymbol{P}_t(\boldsymbol{\theta}))'}{\partial \boldsymbol{\rho}}\left\{vec\boldsymbol{P}_t^{-1} - \left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{P}_t^{-1}\right)(\boldsymbol{z}_t\otimes\boldsymbol{z}_t)\right\}, \\
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} &= -\frac{1}{2}\frac{\partial(vec\boldsymbol{P}_t(\boldsymbol{\theta}))'}{\partial \boldsymbol{\varphi}}\left\{vec\boldsymbol{P}_t^{-1} - \left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{P}_t^{-1}\right)(\boldsymbol{z}_t\otimes\boldsymbol{z}_t)\right\},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i} &= \boldsymbol{v}_{i,t-1} + \beta_i\frac{\partial h_{i,t-1}}{\partial \boldsymbol{\omega}_i}, \quad i = 1, \ldots, N, \\
\frac{\partial(vec\boldsymbol{P}_t)'}{\partial \boldsymbol{\rho}} &= \boldsymbol{v}_{\boldsymbol{\rho}t}\otimes\boldsymbol{U}', \\
\frac{\partial(vec\boldsymbol{P}_t)'}{\partial \boldsymbol{\varphi}} &= \boldsymbol{v}_{\boldsymbol{\varphi}t}(1 - G_t)G_t vec(\boldsymbol{P}_1 - \boldsymbol{P}_2)'.
\end{aligned}$$

Next we evaluate the score at the true parameters under the null. Taking conditional expectations of the outer product of the score using (19) gives

$$\begin{aligned}
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i'}\right] &= \frac{1}{4h_{it}^2}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i'}\left(1 + 1_i'\boldsymbol{P}_t^{-1}1_i\right), \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_j'}\right] &= \frac{1}{4h_{it}h_{jt}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial h_{jt}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_j'}\left(\rho_{t,ij}1_i'\boldsymbol{P}_t^{-1}1_j\right), \quad i \neq j \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}'}\right] &= \frac{1}{4h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\left(1_i'\boldsymbol{P}_t^{-1}\otimes 1_i' + 1_i'\otimes 1_i'\boldsymbol{P}_t^{-1}\right)\frac{\partial vec\boldsymbol{P}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}'}, \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{4h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\left(1_i'\boldsymbol{P}_t^{-1}\otimes 1_i' + 1_i'\otimes 1_i'\boldsymbol{P}_t^{-1}\right)\frac{\partial vec\boldsymbol{P}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}, \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}'}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}}\left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{P}_t^{-1} + \left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}'}, \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}}\left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{P}_t^{-1} + \left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}, \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\varphi}}\left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{P}_t^{-1} + \left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}_t^{-1}\otimes\boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}, \qquad (21)
\end{aligned}$$

where $\boldsymbol{K}$ is defined as in (18).

The estimator for the infomation matrix is obtained by using the submatrices in (21). To derive a more compact expression for the information matrix, let $\boldsymbol{x}_t = (\boldsymbol{x}'_{1t}, \ldots, \boldsymbol{x}'_{Nt})'$ where $\boldsymbol{x}_{it} = -\frac{1}{2h_{it}} \frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i}$, let $\boldsymbol{x}_{\boldsymbol{\rho}t} = -\frac{1}{2}\boldsymbol{v}_{\boldsymbol{\rho}t} \otimes \boldsymbol{U}'$, and $\boldsymbol{x}_{\boldsymbol{\varphi}t} = -\frac{1}{2}\boldsymbol{v}_{\boldsymbol{\varphi}t}(1-G_t)G_t vec(\boldsymbol{P}_1 - \boldsymbol{P}_2)'$ and let $\boldsymbol{x}^0_{it}$, $i = 1, \ldots, N, \boldsymbol{\rho}, \boldsymbol{\varphi}$, denote the corresponding expressions evaluated at the true values under the null hypothesis. Setting

$$
\begin{aligned}
\boldsymbol{M}_1 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}^0_t \boldsymbol{x}^{0\prime}_t \odot \left(\left(\boldsymbol{I} + \boldsymbol{P}_t \odot \boldsymbol{P}_t^{-1}\right) \otimes \boldsymbol{1}_3\right), \\
\boldsymbol{M}_2 &= T^{-1}\sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{x}^0_{1t} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{x}^0_{Nt} \end{bmatrix} \begin{bmatrix} 1'_1 \boldsymbol{P}_t^{-1} \otimes 1'_1 + 1'_1 \otimes 1'_1 \boldsymbol{P}_t^{-1} \\ \vdots \\ 1'_N \boldsymbol{P}_t^{-1} \otimes 1'_N + 1'_N \otimes 1'_N \boldsymbol{P}_t^{-1} \end{bmatrix} \boldsymbol{x}^{0\prime}_{\boldsymbol{\rho}t}, \\
\boldsymbol{M}_3 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}^0_{\boldsymbol{\rho}t} \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{P}_t^{-1} + \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{I}\right)\right) \boldsymbol{x}^{0\prime}_{\boldsymbol{\rho}t}, \\
\boldsymbol{M}_4 &= T^{-1}\sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{x}^0_{1t} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{x}^0_{Nt} \end{bmatrix} \begin{bmatrix} 1'_1 \boldsymbol{P}_t^{-1} \otimes 1'_1 + 1'_1 \otimes 1'_1 \boldsymbol{P}_t^{-1} \\ \vdots \\ 1'_N \boldsymbol{P}_t^{-1} \otimes 1'_N + 1'_N \otimes 1'_N \boldsymbol{P}_t^{-1} \end{bmatrix} \boldsymbol{x}^{0\prime}_{\boldsymbol{\varphi}t}, \\
\boldsymbol{M}_5 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}^0_{\boldsymbol{\rho}t} \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{P}_t^{-1} + \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{I}\right)\right) \boldsymbol{x}^{0\prime}_{\boldsymbol{\varphi}t}, \\
\boldsymbol{M}_6 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}^0_{\boldsymbol{\varphi}t} \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{P}_t^{-1} + \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_t^{-1} \otimes \boldsymbol{I}\right)\right) \boldsymbol{x}^{0\prime}_{\boldsymbol{\varphi}t},
\end{aligned}
$$

the information matrix $\mathfrak{I}(\boldsymbol{\theta}_0)$ is approximated by

$$
\begin{aligned}
\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0) &= T^{-1}\sum_{t=1}^{T} E\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mid \mathcal{F}_{t-1}\right] \\
&= \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 & \boldsymbol{M}_4 \\ \boldsymbol{M}'_2 & \boldsymbol{M}_3 & \boldsymbol{M}_5 \\ \boldsymbol{M}'_4 & \boldsymbol{M}'_5 & \boldsymbol{M}_6 \end{bmatrix}.
\end{aligned}
$$

The block of the inverse of $\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0)$ corresponding to the correlation and transition parameters is given by

$$
\left(\begin{bmatrix} \boldsymbol{M}_3 & \boldsymbol{M}_5 \\ \boldsymbol{M}'_5 & \boldsymbol{M}_6 \end{bmatrix} - \begin{bmatrix} \boldsymbol{M}'_2 \\ \boldsymbol{M}'_4 \end{bmatrix} \boldsymbol{M}_1^{-1} \begin{bmatrix} \boldsymbol{M}_2 & \boldsymbol{M}_4 \end{bmatrix}\right)^{-1}
$$

from which the south-east $N(N-1) \times N(N-1)$ block $[\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0)]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})}$ corresponding to the correlation parameters can be extracted. The test statistic evaluated at the restricted maximum likelihood estimates is then

$$
LM_{PCCC} = T^{-1}\boldsymbol{q}(\hat{\boldsymbol{\theta}})' \left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})} \boldsymbol{q}(\hat{\boldsymbol{\theta}}). \tag{22}
$$

In (22), $\boldsymbol{q}(\hat{\boldsymbol{\theta}})$ is the $N(N-1) \times 1$ block of the score vector corresponding to the correlation parameters whose elements equal $\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \rho_{ij}}$ if the correlation between assets i and j is constrained to be constant and zero otherwise. Under the null hypothesis, the LM–statistic has an asymptotic χ^2_R distribution where R is the number of restrictions to be tested.

The Wald statistic is

$$
W_{PCCC} = T\boldsymbol{a}(\hat{\boldsymbol{\theta}})' \left(\boldsymbol{A}\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})} \boldsymbol{A}'\right)^{-1} \boldsymbol{a}(\hat{\boldsymbol{\theta}}) \tag{23}
$$

where $\hat{\boldsymbol{\theta}}$ is the vector of maximum likelihood estimates of the full STCC–GARCH model, $\boldsymbol{a}$ is the $R \times 1$ vector of constraints, where R is the number of those constraints, more specifically

$$\boldsymbol{a} = \boldsymbol{V}' vecl\left(\boldsymbol{P}_1 - \boldsymbol{P}_2\right),$$

where matrix $\boldsymbol{V}$ is an $\frac{N(N-1)}{2} \times R$ matrix of zeros and ones, whose columns are defined as

$$\left[vecl\left(1_i 1_j'\right)\right]_{i>j}$$

where i and j correspond to the indices of the assets whose correlation is restricted equal under the null model and the columns appear in the same order from left to right as the indices in $vecl(\boldsymbol{P}_t)$, $\boldsymbol{A} = \frac{\partial \boldsymbol{a}}{\partial \boldsymbol{\rho}'}$, and $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho},\boldsymbol{\rho})}$ is the $N(N-1) \times N(N-1)$ block corresponding the correlation parameters of the inverse of $\hat{\mathfrak{I}}_T$. Under the null hypothesis, the Wald statistic is also asymptotically χ^2_R distributed where R is the number of restrictions to be tested.

Test of partially constant correlations against a less restricted STCC–GARCH model

In this final case, the alternative model is partially constant STCC–GARCH model. Let there be k pairs of variables with constant correlations in the alternative model. The test is as above but with the following changes to definitions and notations. The vector of correlation parameters is $\boldsymbol{\rho} = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2^{P\prime})'$, where the $\frac{N(N-1)}{2} - k \times 1$ vector $\boldsymbol{\rho}_2^P$ is $vecl\boldsymbol{P}_2$ without the constant elements. Then the partial derivatives are as above, with the modification that

$$\frac{\partial\left(vec\boldsymbol{P}_t\right)'}{\partial \boldsymbol{\rho}}$$

is as above, but with k rows deleted so that the remaining rows are corresponding to the elements in $\boldsymbol{\rho} = (\boldsymbol{\rho}_1', \boldsymbol{\rho}_2^{P\prime})'$, and of the first $\frac{N(N-1)}{2}$ rows, the k rows corresponding to the constant correlations in the alternative model are multiplied with 1 instead of $1 - G_t$. With this modification the calculation of the statistics is straightforward, and again the asymptotic distribution under the null hypothesis is χ^2_R where R is the number of restrictions to be tested.

References

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001): "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association*, 96, 42–55.

Anderson, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 3rd edn.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79–109.

Bera, A. K., and S. Kim (2002): "Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns," *Journal of Empirical Finance*, 9, 171–195.

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

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

Bookstaber, R. (1997): "Global risk management: Are we missing the point?," *Journal of Portfolio Management*, 23, 102–107.

Chesnay, F., and E. Jondeau (2001): "Does correlation between stock returns really increase during turbulent periods?," *Economic Notes*, 30, 53–80.

de Santis, G., and B. Gerard (1997): "International asset pricing and portfolio diversification with time-varying risk," *Journal of Finance*, 52, 1881–1912.

Doornik, J. A. (2002): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, 3rd edn., see also www.doornik.com.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1006.

——— (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. F., and K. Sheppard (2001): "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," NBER Working Paper 8554.

He, C., and T. Teräsvirta (2004): "An extended constant conditional correlation GARCH model and its fourth-moment structure," *Econometric Theory*, 20, 904–926.

Jeantheau, T. (1998): "Strong consistency of estimators for multivariate ARCH models," *Econometric Theory*, 14, 70–86.

King, M. A., and S. Wadhwani (1990): "Transmission of volatility between stock markets," *The Review of Financial Studies*, 3, 5–33.

Kwan, C. K., W. K. Li, and K. Ng (2005): "A multivariate threshold GARCH model with time-varying correlations," Preprint series 2005–12, National University of Singapore, Institution of Mathematical Sciences.

Lanne, M., and P. Saikkonen (2005): "Nonlinear GARCH models for highly persistent volatility," *Econometrics Journal*, 8, 251–276.

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

Ling, S., and M. McAleer (2003): "Asymptotic theory for a vector ARMA–GARCH model," *Econometric Theory*, 19, 280–310.

Longin, F., and B. Solnik (2001): "Extreme correlation of international equity markets," *Journal of Finance*, 56, 649–676.

Lütkepohl, H. (1996): *Handbook of Matrices*. John Wiley & Sons, Chichester.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

Pelletier, D. (2006): "Regime switching for dynamic correlations," *Journal of Econometrics*, 131, 445–473.

Silvennoinen, A., and T. Teräsvirta (2007): "Multivariate GARCH models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch. Springer, New York.

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

Tse, Y. K., and K. C. Tsui (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model

Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model

Abstract

In this paper we propose a multivariate GARCH model with a time-varying conditional correlation structure. The new Double Smooth Transition Conditional Correlation GARCH model extends the Smooth Transition Conditional Correlation GARCH model of Silvennoinen and Teräsvirta (2005) by including another variable according to which the correlations change smoothly between states of constant correlations. A Lagrange multiplier test is derived to test the constancy of correlations against the DSTCC–GARCH model, and another one to test for another transition in the STCC–GARCH framework. In addition, other specification tests, with the aim of aiding the model building procedure, are considered. Analytical expressions for the test statistics and the required derivatives are provided. The model is applied to a selection of world stock indices, and it is found that time is an important factor affecting correlations between them.

This paper is joint work with Timo Teräsvirta.

We thank Pierre Giot for helpful comments and suggestions. Participants of the conference Econometric Society Australasian Meeting, Alice Springs, July 2006, also provided useful remarks. The responsibility for any errors and shortcomings in this paper remains ours.

1 Introduction

Multivariate financial time series have been subject to many modelling proposals incorporating conditional heteroskedasticity, originally introduced by Engle (1982) in a univariate context. For a review, the reader is referred to a recent survey on multivariate GARCH models by Bauwens, Laurent, and Rombouts (2006). One may model the time-varying covariances directly. Examples of this are VEC and BEKK models, as well factor GARCH ones, all discussed in Bauwens, Laurent, and Rombouts (2006). Alternatively, one may model the conditional correlations. The simplest approach is to assume that the correlations are time-invariant. Although the Constant Conditional Correlation (CCC) GARCH model of Bollerslev (1990) is attractive to the practitioner due to its interpretable parameters and easy estimation, its fundamental assumption that correlations remain constant over time has often been found unrealistic. In order to remedy this problem, Tse and Tsui (2002) and Engle (2002) introduced models with dynamic conditional correlations called the VC–GARCH and the DCC–GARCH model, respectively, that impose GARCH-type structure on the correlations. By construction, these models have the property that the variation in correlations is mainly due to the size and the sign of the shock of the previous time period.

An interesting model combining aspects from both the CCC–GARCH and the DCC–GARCH has been suggested by Pelletier (2006). The author introduces a regime switching correlation structure driven by an unobserved state variable following a first-order Markov chain. The regime switching model asserts that the correlations remain constant in each regime and the change between the states is abrupt and governed by transition probabilities. Thus the factors affecting the correlations remain latent and are not observed.

In a recent paper, Silvennoinen and Teräsvirta (2005) introduced the Smooth Transition Conditional Correlation (STCC) GARCH model.[1] In this model the correlations vary smoothly between two extreme states of constant correlations and the dynamics are driven by an observable transition variable. The transition variable can be chosen by the modeller, and the model combined with tests of constant correlations constitutes a useful tool for modellers interested in characterizing the dynamic structure of the correlations. This paper extends the STCC–GARCH model into one that allows variation in conditional correlations to be controlled by two observable transition variables instead of only one. This makes it possible, for example, to nest the Berben and Jansen (2005a) model with time as the transition variable in this general double-transition model.

It has become a widely accepted feature of financial data that volatile periods in financial markets are related to an increase in correlations among assets. However, as pointed out by Boyer, Gibson, and Loretan (1999) and Longin and Solnik (2001), in many studies this hypothesis is not investigated properly and the reported results may be misleading. In fact, the latter authors report evidence that in international markets correlations are not related to market volatility as measured in large absolute

[1] A bivariate special case of the STCC–GARCH model was coincidentally introduced in Berben and Jansen (2005a).

returns, but only to large negative returns, or to the market trend. Our modelling framework allows the researcher to easily explore such possibilities by first testing the relevance of a model with a transition variable corresponding to the hypothesis to be tested and, in case of rejection, estimating the model to find out the direction of change in correlations controlled by that variable; see Silvennoinen and Teräsvirta (2005) for an example.

The paper is organized as follows. In Section 2 the new DSTCC–GARCH model is introduced and its estimation discussed. Section 3 gives the testing procedures and Section 4 reports simulation experiments on the tests. In Section 5 we apply our model to a set of four international stock market indices, namely French CAC 40, German DAX, FTSE 100 from UK, and Hong Kong Hang Seng, from December 1990 until present. Finally, Section 6 concludes. The detailed derivations of the tests can be found in the Appendix.

2 The Double Smooth Transition Conditional Correlation GARCH model

2.1 The general multivariate GARCH model

Consider the following stochastic N-dimensional vector process with the standard representation

$$\boldsymbol{y}_t = E\left[\boldsymbol{y}_t \mid \mathcal{F}_{t-1}\right] + \boldsymbol{\varepsilon}_t \qquad t = 1, 2, \ldots, T \tag{1}$$

where $\mathcal{F}_{t-1}$ is the sigma-field generated by all the information until time $t-1$. Each of the univariate error processes has the specification

$$\varepsilon_{it} = h_{it}^{1/2} z_{it}$$

where the errors z_{it} form a sequence of independent random variables with mean zero and variance one, for each $i = 1, \ldots, N$. The conditional variance h_{it} follows a univariate GARCH process, for example that of Bollerslev (1986)

$$h_{it} = \alpha_{i0} + \sum_{j=1}^{q} \alpha_{ij} \varepsilon_{i,t-j}^2 + \sum_{j=1}^{p} \beta_{ij} h_{i,t-j} \tag{2}$$

with the non-negativity and stationarity restrictions imposed. The results in this paper are derived using (2) with $p = q = 1$ to account for the conditional heteroskedasticity. It is straightforward to modify them to allow for a higher-order or some other type of GARCH process. The conditional covariances of the vector $\boldsymbol{z}_t$ are given by

$$E\left[\boldsymbol{z}_t \boldsymbol{z}_t' \mid \mathcal{F}_{t-1}\right] \quad = \quad \boldsymbol{P}_t. \tag{3}$$

Furthermore, the standardized errors $\boldsymbol{\eta}_t = \boldsymbol{P}_t^{-1/2} \boldsymbol{z}_t \sim iid(\boldsymbol{0}, \boldsymbol{I}_N)$. Since z_{it} has unit variance for all i, $\boldsymbol{P}_t = [\rho_{ij,t}]$ is the conditional correlation matrix for the $\boldsymbol{\varepsilon}_t$ whose

elements $\rho_{ij,t}$ are allowed to be time-varying for $i \neq j$. It will, however, be assumed that $\boldsymbol{P}_t \in \mathcal{F}_{t-1}$.

The conditional covariance matrix $\boldsymbol{H}_t = \boldsymbol{S}_t \boldsymbol{P}_t \boldsymbol{S}_t$, where $\boldsymbol{P}_t$ is the conditional correlation matrix as in equation (3), and $\boldsymbol{S}_t = diag(h_{1t}^{1/2}, .., h_{Nt}^{1/2})$ with elements defined in (2), is positive definite whenever the correlation matrix $\boldsymbol{P}_t$ is positive definite.

2.2 Smooth transitions in conditional correlations

The idea of introducing smooth transition in the conditional correlations is discussed in detail in Silvennoinen and Teräsvirta (2005) where a simple structure with one type of transition between two states of constant correlations is introduced. Specifically, the STCC–GARCH model defines the time-varying correlation structure as

$$\boldsymbol{P}_t = (1 - G_t)\boldsymbol{P}_{(1)} + G_t \boldsymbol{P}_{(2)} \tag{4}$$

where the transition function $G_t = G(s_t; \gamma, c)$ is the logistic function

$$G_t = \left(1 + e^{-\gamma(s_t - c)}\right)^{-1}, \quad \gamma > 0 \tag{5}$$

that is bounded between zero and one. Furthermore, $\boldsymbol{P}_{(1)}$ and $\boldsymbol{P}_{(2)}$ represent the two extreme states of correlations between which the conditional correlations can vary over time according to the transition variable s_t. The two parameters in (5), γ and c, define the speed and location of the transition. When the transition variable has values less than c, the correlations are closer to the state defined by $\boldsymbol{P}_{(1)}$ than the one defined by $\boldsymbol{P}_{(2)}$. For $s_t > c$, the situation is the opposite. The parameter γ controls the smoothness of the transition between the two states. The closer γ is to zero the slower the transition. As $\gamma \to \infty$, the transition function eventually becomes a step function. The positive definiteness of $\boldsymbol{P}_t$ in each point in time is ensured by the requirement that the two correlation matrices $\boldsymbol{P}_{(1)}$ and $\boldsymbol{P}_{(2)}$ are positive definite. As a special case, by defining the transition variable to be the calendar time, $s_t = t/T$, one arrives at the Time-Varying Conditional Correlation (TVCC) GARCH model. A bivariate version of this model was introduced by Berben and Jansen (2005a).

We extend the original STCC–GARCH model by allowing the conditional correlations to vary according to two transition variables. The time-varying correlation structure in the Double Smooth Transition Conditional Correlation (DSTCC) GARCH model is imposed through the following equations:

$$\begin{aligned} \boldsymbol{P}_t &= (1 - G_{1t})\boldsymbol{P}_{(1)t} + G_{1t}\boldsymbol{P}_{(2)t} \\ \boldsymbol{P}_{(i)t} &= (1 - G_{2t})\boldsymbol{P}_{(i1)} + G_{2t}\boldsymbol{P}_{(i2)}, \quad i = 1, 2 \end{aligned} \tag{6}$$

where the transition functions are the logistic functions

$$G_{it} = \left(1 + e^{-\gamma_i(s_{it} - c_i)}\right)^{-1}, \quad \gamma_i > 0, \quad i = 1, 2 \tag{7}$$

and s_{it}, $i = 1, 2$, are transition variables that can be either stochastic or deterministic. The correlation matrix $\boldsymbol{P}_t$ is thus a convex combination of four positive definite matrices, $\boldsymbol{P}_{(11)}$, $\boldsymbol{P}_{(12)}$, $\boldsymbol{P}_{(21)}$, and $\boldsymbol{P}_{(22)}$, each of which defines an extreme state of constant correlation. The positive definiteness of $\boldsymbol{P}_t$ at each point in time follows again from the positive definiteness of these four matrices. In (7) the parameters γ_i and c_i determine the speed and the location of the transition i, $i = 1, 2$. The transition variables are chosen by the modeller. As in the STCC–GARCH model, the values of these variables are assumed to be known at time t. Possible choices are for instance functions of lagged elements of $\boldsymbol{y}_t$, or exogenous variables. When applying the model to stock return series one could consider functions of market indices or business cycle indicators, or simply time. If one of the transition variables is time, say $s_{2t} = t/T$, the model with correlation dynamics (6) is the Time-Varying Smooth Transition Conditional Correlation (TVSTCC) GARCH model. In this case it may be illustrative to write (6) as

$$\boldsymbol{P}_t \;=\; (1-G_{2t})\left((1-G_{1t})\boldsymbol{P}_{(11)} + G_{1t}\boldsymbol{P}_{(21)}\right) + G_{2t}\left((1-G_{1t})\boldsymbol{P}_{(12)} + G_{1t}\boldsymbol{P}_{(22)}\right). \quad (8)$$

The role of the correlation matrices describing the constant states is easily seen from (8). At the beginning of the sample the correlations vary smoothly between the states defined by $\boldsymbol{P}_{(11)}$ and $\boldsymbol{P}_{(21)}$: when $s_{1t} < c_1$, the correlations are closer to the state in $\boldsymbol{P}_{(11)}$ than $\boldsymbol{P}_{(21)}$ whereas when $s_{1t} > c_1$, the situation is the opposite. As time evolves the correlations in $\boldsymbol{P}_{(11)}$ and $\boldsymbol{P}_{(21)}$ transform smoothly to the ones in $\boldsymbol{P}_{(12)}$ and $\boldsymbol{P}_{(22)}$, respectively. Therefore, at the end of the sample, s_{1t} shifts the correlations between these two matrices.

The specification (6) describes the correlation structre of the DSTCC–GARCH model in its fully general form. Imposing certain restrictions on the correlations give rise to numerous special cases; those will be discussed in Section 3. One restricted version, however, is worth discussing in detail. The effect of the two transition variables can be independent in a sense that the time-variation of the correlations due to one of the transition variables do not depend on the value of the other transition variable. This condition can be expressed as

$$\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} \;=\; \boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)} \quad (9a)$$

or, equivalently,

$$\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(21)} \;=\; \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(22)}. \quad (9b)$$

In terms of equation (8) these conditions imply that on the right-hand side this equation the matrices with coefficients $\pm G_{1t}G_{2t}$ are eliminated. Furthermore, from (9a) and (9b) it follows that the difference between the extreme states described by one of the transition variables remains constant accross all values of the other transition variable. In this case the dynamic conditional correlations of the DSTCC–GARCH model become

$$\boldsymbol{P}_t \;=\; (1 - G_{1t} - G_{2t})\boldsymbol{P}_{(11)} + G_{1t}\boldsymbol{P}_{(21)} + G_{2t}\boldsymbol{P}_{(12)}. \quad (10)$$

This parsimonious specification may prove useful when dealing with large systems because one only has to estimate three correlation matrices instead of four in an unrestricted DSTCC–GARCH model.

2.3 Estimation of the DSTCC–GARCH model

For maximum likelihood estimation of parameters we assume joint conditional normality of the errors:

$$\boldsymbol{z}_t \mid \mathcal{F}_{t-1} \sim N\left(\mathbf{0}, \boldsymbol{P}_t\right).$$

Denoting by $\boldsymbol{\theta}$ the vector of all the parameters in the model, the log-likelihood for observation t is

$$l_t\left(\boldsymbol{\theta}\right) = -\frac{N}{2}\log\left(2\pi\right) - \frac{1}{2}\sum_{i=1}^{N}\log h_{it} - \frac{1}{2}\log\left|\boldsymbol{P}_t\right| - \frac{1}{2}\boldsymbol{z}_t'\boldsymbol{P}_t^{-1}\boldsymbol{z}_t, \quad t = 1, \ldots, T \tag{11}$$

and maximizing $\sum_{t=1}^{T} l_t(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ yields the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T$.

For inference we assume that the asymptotic distribution of the ML-estimator is normal, that is,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathfrak{I}^{-1}(\boldsymbol{\theta}_0)\right)$$

where $\boldsymbol{\theta}_0$ is the true parameter and $\mathfrak{I}(\boldsymbol{\theta}_0)$ the population information matrix evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Asymptotic properties of the estimators are yet to be explored, see the discussion in Silvennoinen and Teräsvirta (2005). To increase efficiency of the estimation, maximization of the log-likelihood is carried out iteratively by concentrating the likelihood, in each round by splitting the parameters into three sets: GARCH, correlation, and transition function parameters. The log-likelihood is maximized with respect to one set at the time keeping the other parameters fixed at their previously estimated values. The convergence is reached once the estimated values cannot be improved upon when compared with the ones obtained from previous iteration. As mentioned in Section 2.2, the transition between the extreme states becomes more rapid and the transition function eventually becomes a step function as $\gamma \to \infty$. When γ has reached a value large enough, no increment will change the shape of the transition function. The likelihood function becomes flat with respect to that parameter and numerical optimizers have difficulties in converging. Therefore, one may want to fix an upper limit for γ, whose value naturally depends on the transition variable in question. Plotting a graph of the transition function can be useful in deciding such a limit. It should be noted that, if the upper limit is reached, the resulting estimates for the rest of the parameters are conditional on this value of γ. Furthermore, it should be pointed out, that estimation requires care. The log-likelihood may have several local maxima, so estimation should be initiated from a set of different starting-values, and the maxima thus obtained compared before settling for final estimates. All computations in this paper have been performed using Ox, version 3.30, see Doornik (2002), and our own source code.

Before estimating an STCC–GARCH or a DSTCC–GARCH model, however, it is necessary to test the hypothesis that the conditional correlations are constant. The reason for this is that some of the parameters of the alternative model are not identified if the true model has constant conditional correlations. Estimating an STCC–GARCH or a DSTCC–GARCH model without first testing the constancy hypothesis could thus

lead to inconsistent parameter estimates. The same is true if one wishes to increase the number of transitions in an already estimated STCC–GARCH model. Testing procedures will be discussed in the next section.

3 Hypothesis testing

3.1 Testing for smooth transitions in conditional correlations

Parametric modelling of the dynamic behaviour of the conditional correlations should begin with testing constancy of the correlations. Neglected variation in parameters leads to a misspecified likelihood and thus to invalid asymptotic inference. Tse (2000), Bera and Kim (2002), Engle and Sheppard (2001), and Silvennoinen and Teräsvirta (2005) have proposed tests for this purpose. Tse (2000) derives a Lagrange multiplier (LM) test where the alternative model imposes ARCH-type dynamics on the conditional correlations. Bera and Kim (2002) discuss testing the hypothesis of no parameter variation using the information matrix test of White (1982). The test of Engle and Sheppard (2001) is based on the fact that the standardized residuals $\boldsymbol{\eta}_t$ should be iid both in time and accross the series if the model is correctly specified. This test, however, is not only a test of constant correlations but a general misspecification test as it cannot distinguish between misspecified conditional correlations and conditional heteroskedasticity in the univariate residual series.

The approach of Silvennoinen and Teräsvirta (2005) differs from the others in that the test is conditioned on a particular transition variable and in effect tests whether that particular factor affects conditional correlations between the variables. A failure to reject the constancy of correlations is thus interpreted as evidence that this transition variable is not informative about possible time-variation of the correlations. A non-rejection thus does not indicate that the correlations are constant, but the test may of course be carried out for a sequence of different transition variables. But then, a rejection of the null hypothesis does provide evidence of nonconstancy of the conditional correlations and may be taken to imply that the transition variable in question carries information about the time-varying structure of the correlations.

After fitting an STCC–GARCH model to the data one may wish to see whether or not the transition variable of the model is the only factor that affects the conditional correlations over time. In the present framework this means that there may be another factor whose effect on correlations cannot be ignored. For instance, the unconditional correlations may vary as a function of time, in which case a second transition depending directly on time together with the previous one would provide a better description of the correlation dynamics than the STCC–GARCH model does. A linear function of time would indicate a monotonic relationship between calendar time and correlations, whereas introducing higher-order polynomials or nonlinear functions would allow that structure to capture more complicated patterns in time-varying correlations.

An indication of the importance of a second transition variable can be obtained by testing the constancy of correlations against an STCC–GARCH model in which the correlations are functions of to this particular transition variable. The next step

is to estimate the STCC–GARCH model with the transition variable against which the strongest rejection of constancy is obtained, and proceed by testing this model against the DSTCC–GARCH one.

The null hypothesis in testing for another transition is $\gamma_2 = 0$ in (6) and (7). The problem of unidentified parameters under the null hypothesis is circumvented by linearizing the form of dynamic correlations under the alternative model. This is done by a Taylor approximation of the second transition function, G_{2t}, around $\gamma_2 = 0$; see Luukkonen, Saikkonen, and Teräsvirta (1988) for this idea. Replacing the transition function in (6) by the approximation, the dynamic conditional correlations become

$$\boldsymbol{P}_t^* \quad = \quad (1 - G_{1t})\boldsymbol{P}_{(1)}^* + G_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^* + \boldsymbol{R} \tag{12}$$

where the remainder $\boldsymbol{R}$ is the error due to the linearization. Note that under the null hypothesis $\boldsymbol{R} = \mathbf{0}_{N\times N}$, so the remainder does not affect the asymptotic null distribution of the LM–test statistic. Note also that when $G_{1t} \equiv 0$ so that under the null the correlations are constant, the test collapses into the correlation constancy test in Silvennoinen and Teräsvirta (2005). For details of the linearization and the transformed dynamic correlations in (12), see the Appendix. The auxiliary null hypothesis can now be stated as $vecl\boldsymbol{P}_{(3)}^* = \mathbf{0}_{N(N-1)/2\times 1}$, where $vecl(\cdot)$ is an operator that stacks the columns of the strict lower triangular part of its argument square matrix. Under the null hypothesis,

$$\boldsymbol{P}_t^* \quad = \quad (1 - G_{1t})\boldsymbol{P}_{(1)}^* + G_{1t}\boldsymbol{P}_{(2)}^*. \tag{13}$$

Constructing the Lagrange multiplier test yields the statistic and its asymptotic null distribution in the usual way. The test statistic is

$$T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\rho}_{(3)}^{*\prime}}\right)\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_{(3)}^*,\boldsymbol{\rho}_{(3)}^*)}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\rho}_{(3)}^{*}}\right) \sim \chi^2_{N(N-1)/2}. \tag{14}$$

The detailed form of (14) can be found in the Appendix.

It should be pointed out that even if constancy is rejected against an STCC–GARCH model for both s_{1t} and s_{2t}, the test for another transition after estimating this model for one of the two may not be able to reject the null hypothesis. This may be the case when both variables contain similar information about the correlations, whereby adding a second transition will not improve the model. Because estimation of an STCC–GARCH model can sometimes be a computationally demanding task, some idea of suitable transition variables may be obtained by testing constancy of correlations directly against the DSTCC–GARCH model. The null hypothesis is $\gamma_1 = \gamma_2 = 0$ in (6) and (7). To circumvent the problem with unidentified parameters under the null, both transition functions, G_{1t} and G_{2t}, in (6) are linearized around $\gamma_1 = 0$ and $\gamma_2 = 0$, respectively, as discussed above. The linearized dynamic correlations then become

$$\boldsymbol{P}_t^* \quad = \quad \boldsymbol{P}_{(1)}^* + s_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^* + s_{1t}s_{2t}\boldsymbol{P}_{(4)}^* + \boldsymbol{R} \tag{15}$$

where $\boldsymbol{R}$ again holds all approximation error. The auxiliary null hypothesis based on the transformed dynamic correlations is now $vecl\boldsymbol{P}_{(2)}^* = vecl\boldsymbol{P}_{(3)}^* = vecl\boldsymbol{P}_{(4)}^* =$

$\mathbf{0}_{N(N-1)/2\times 1}$ under which the conditional correlations are constant, $\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)}^*$. The Lagrange multiplier test statistic and its asymptotic null distribution are constructed as usual:

$$T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial(\boldsymbol{\rho}_{(2)}^{*\prime},\boldsymbol{\rho}_{(3)}^{*\prime},\boldsymbol{\rho}_{(4)}^{*\prime})}\right)\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_{(2-4)}^{*},\boldsymbol{\rho}_{(2-4)}^{*})}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial(\boldsymbol{\rho}_{(2)}^{*\prime},\boldsymbol{\rho}_{(3)}^{*\prime},\boldsymbol{\rho}_{(4)}^{*\prime})^{\prime}}\right) \sim \chi^2_{3N(N-1)/2}. \tag{16}$$

The detailed form of (16) can be found in the Appendix.

As discussed in Section 2.2, the full DSTCC–GARCH model is simplified when the effects of the two transition variables are independent. This restricted version can be used as a more parsimonious alternative than the DSTCC–GARCH model when testing constancy. For instance, if one of the transition variables, say s_{2t}, is time, one can test constancy against an alternative where the variation controlled by s_{1t} does not depend on time, that is, the differences between the two extremes states are equal during the whole sample period. The constancy of correlations is tested by testing $\gamma_1 = \gamma_2 = 0$ in (10) and now the linearized equation is simply a special case of (15) such that $\boldsymbol{P}_{(4)}^* = \mathbf{0}$. The auxiliary null hypothesis is $vecl\boldsymbol{P}_{(2)}^* = vecl\boldsymbol{P}_{(3)}^* = \mathbf{0}_{N(N-1)/2\times 1}$ under which the conditional correlations are constant: $\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)}^*$. The Lagrange multiplier test statistic and its asymptotic null distribution are the following:

$$T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial(\boldsymbol{\rho}_{(2)}^{*\prime},\boldsymbol{\rho}_{(3)}^{*\prime})}\right)\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_{(2-3)}^{*},\boldsymbol{\rho}_{(2-3)}^{*})}\left(\sum_{t=1}^{T}\frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial(\boldsymbol{\rho}_{(2)}^{*\prime},\boldsymbol{\rho}_{(3)}^{*\prime})^{\prime}}\right) \sim \chi^2_{N(N-1)}. \tag{17}$$

Statistic (17) is a special case of (16) and its detailed form can be found in the Appendix.

If the tests fail to reject the null hypothesis or the rejection is not particularly strong, the reason for this can be that some correlations are constant. If sufficiently many but not all correlations are constant according to one of the transition variables or both, the tests may not be sufficiently powerful to reject the null hypothesis. In these cases the power of the tests can be increased by modifying the alternative model. For instance, one may want to test constancy of correlations against a DSTCC–GARCH model in which some correlations are constant with respect to one of the transition variables, or both. Similarly, an STCC–GARCH model containing constant correlations can be tested against a DSTCC–GARCH model with constancy restrictions. These tests are straightforward extensions of the tests already discussed, see the Appendix for details. A test of constant correlations against an STCC–GARCH model containing constant correlations is discussed in Silvennoinen and Teräsvirta (2005).

When it comes to 'fine-tuning' of the model, i.e., when the model under the null hypothesis has the same number of transitions as under the alternative and the modeller is focused on potential constancy of some of the correlations controlled by one of the transition variables or both, tests of partial constancy can also be built on the Wald principle. This is quite practical because after estimating the alternative model,

several restrictions can be tested at the same time without re-estimation. In our experience, when it comes to restricting correlations to be constant, the specification search beginning with a general model and restricting correlations generally yields the same final model as would a bottoms-up approach beginning with a restricted model and testing for additional time-varying correlations. This conclusion has been reached using both simulated series and observed returns. The only difference between these two approaches seems to be that the final model is obtained faster with the former than with the latter. Restricting some of the correlations to be constant decreases the number of parameters to be estimated, which is convenient especially in large models.

4 Size simulations

Empirical size of the LM–type test of STCC–GARCH models against DSTCC–GARCH ones are investigated by simulation. The observations are generated from a bivariate first-order STCC–GARCH model. The transition variable is generated from an exogenous process: $s_t = h_{et}^{1/2} z_{et}$, where h_{et} has a GARCH(1,1) structure and $z_{et} \sim nid(0,1)$. The STCC–GARCH model is tested against a DSTCC–GARCH model where the correlations vary also as function of time, i.e. the alternative model is the TVSTCC–GARCH model. The parameter values in each of the individual GARCH equations are chosen such that they resemble results often found in fitting GARCH(1,1) models to financial return series. Thus,

$$\begin{aligned} h_{1t} &= 0.01 + 0.04\varepsilon_{1,t-1}^2 + 0.94 h_{1,t-1} \\ h_{2t} &= 0.03 + 0.05\varepsilon_{2,t-1}^2 + 0.92 h_{2,t-1} \\ h_{et} &= 0.005 + 0.03 s_{t-1}^2 + 0.96 h_{et,t-1}. \end{aligned}$$

We conduct three experiments where $\rho_{(1)} = 0$, and $\rho_{(2)} = 1/3, 1/2, 2/3$. The location parameter $c = 0$. We consider two choices for the value of the slope parameter γ. The first one represents a rather slow transition, $\gamma = 5$, in which case about 75% of the correlations lie genuinely between $\rho_{(1)}$ and $\rho_{(2)}$, and the remaining 25% take one of the extreme values. The other choice is $\gamma = 20$, and the ratios are now interchanged: only 25% of the correlations are different from $\rho_{(1)}$ or $\rho_{(2)}$. The sample sizes are $T = 1000$ and $T = 2500$. Considering longer time series was found unnecessary because the results suggested that the empirical size is close to the nominal one at these sample sizes already. The results in Table 1 are based on 5000 replications.

We carry out another size simulation experiment in which we test the CCC–GARCH model directly against the TVSTCC–GARCH model. In the latter model, one transition is controlled by an exogenous GARCH(1,1) process and the other one by time. Specifically, the model under the null is a bivariate CCC–GARCH(1,1) model where the GARCH processes are h_{1t} and h_{2t} from the previous study. For the constant correlation between the series we use four values: 0, 1/3, 1/2, and 2/3. These four experiments are performed using samples of sizes $T = 1000$ and $T = 2500$, with 5000 replications. The results in Table 2 indicate that the test does not suffer

		$\rho_1 = 0,\ \rho_2 = 1/3$		$\rho_1 = 0,\ \rho_2 = 1/2$		$\rho_1 = 0,\ \rho_2 = 2/3$	
	nominal size	$\gamma = 5$	$\gamma = 20$	$\gamma = 5$	$\gamma = 20$	$\gamma = 5$	$\gamma = 20$
	1%	0.0104	0.0082	0.0082	0.0098	0.0072	0.0100
$T = 1000$	5%	0.0482	0.0418	0.0472	0.0500	0.0432	0.0502
	10%	0.0944	0.0884	0.0970	0.0996	0.0912	0.1016
	1%	0.0098	0.0106	0.0106	0.0120	0.0074	0.0108
$T = 2500$	5%	0.0518	0.0512	0.0498	0.0472	0.0456	0.0462
	10%	0.0958	0.1020	0.1018	0.1010	0.0964	0.1012

Table 1: Size of the test of STCC–GARCH model against an STCC–GARCH model with an additional transition for sample sizes 1000 and 2500 and for three choices of correlations for the extreme states; 5000 replications.

	nominal size	$\rho = 0$	$\rho = 1/3$	$\rho = 1/2$	$\rho = 2/3$
	1%	0.0116	0.0122	0.0138	0.0144
$T = 1000$	5%	0.0522	0.0538	0.0542	0.0616
	10%	0.1016	0.1054	0.1050	0.1156
	1%	0.0102	0.0136	0.0146	0.0146
$T = 2500$	5%	0.0466	0.0550	0.0588	0.0538
	10%	0.0950	0.1068	0.1132	0.1072

Table 2: Size of the test of CCC–GARCH model against an STCC–GARCH model with two transitions for sample sizes 1000 and 2500 and for three choices of correlations for the extreme states; 5000 replications.

from size distortions and can thus be applied without further adjustments. This was also the case for the test of constant correlations against the STCC–GARCH model, reported in Silvennoinen and Teräsvirta (2005).

5 Correlations between world market indices

Correlations are especially relevant to risk management and finding efficient hedging positions for portfolios. Inaccurate estimates of correlations put the performance of hedging operations at risk. One often recommended hedging strategy is to diversify the portfolio internationally. However, due to the globalization the financial markets around the world are increasingly integrated which can weaken the protection of the portfolio against local or national crises.

	min	max	mean	st.dev	skewness	kurtosis
CAC	-12.13	11.03	0.1459	2.7320	-0.1272	4.1330
DAX	-14.08	12.89	0.1776	2.9699	-0.2894	5.1952
FTSE	-8.86	10.07	0.1282	2.0667	-0.0944	4.8674
HSI	-19.92	13.92	0.2147	3.4701	-0.4302	5.9599

Table 3: Descriptive statistics of the return series.

In this section we investigate the correlation dynamics among world stock indices. The interest lies in revealing potential risks to internationally diversified portfolios, posed by increasing integration of the world markets. The four major indices considered are the French CAC 40, German DAX, FTSE 100 from UK, and Hong Kong Hang Seng (HSI). We use weekly observations recorded as the closing price of the current week from the beginning of December 1990 to the end of April 2006, 804 observations in all. Weekly observations are preferred to daily ones because the aggregation over time is likely to weaken the effect of different opening hours of the markets around the world. Martens and Poon (2001) discussed the problem of distinguishing contemporaneous correlation from a spillover effect and provided evidence of downward bias in estimated correlations in the presence of nonsynchronous markets. The returns are calculated as differenced log prices. Descriptive statistics of the return series are reported in Table 3.

It is often found, see for instance Lin, Engle, and Ito (1994), de Santis and Gerard (1997), Longin and Solnik (2001), Chesnay and Jondeau (2001), and Cappiello, Engle, and Sheppard (2003), that the correlations behave differently in times of distress from what they do during periods of tranquillity. It is therefore of interest to study how the general level of uncertainty or market turbulence affects the correlation dynamics between the stock indices. In order to do this, we choose our first transition variable to be the one-week lag of the CBOE volatility index (VIX) that represents the market expectation of 30-day volatility. It is constructed using the implied volatilities of a wide range of S&P 500 index options. The VIX is a commonly used measure of market risk and is for this reason often referred to as the 'investor fear gauge'. The values of the index exceeding 30 are generally associated with a large amount of volatility, due to investor fear or uncertainty, whereas the index falling below 20 indicates less stressful, even complacent, times in the markets. But then, the level of unconditional correlations can also change over time. In order to allow for this effect we use time, rescaled between zero and one, as our second transition variable in our DSTCC–GARCH model.

Tests of constant correlations against smooth transition over time as well as against correlations that vary according to the lagged VIX result in rejecting the null hypothesis in the full four-variate model. When testing constancy against the DSTCC–GARCH model, the rejection of the null model is very strong (the p–value equals 2×10^{-27}). It appears that both time itself and the volatility index convey informa-

Estimated model Transition variable in the test	CCC t/T	CCC VIX_{t-1}	CCC VIX_{t-1} and t/T	TVCC VIX_{t-1}
CAC–DAX	1×10^{-24}	0.0005	3×10^{-23}	0.0020
CAC–FTSE	1×10^{-17}	0.0020	2×10^{-16}	0.4136
CAC–HSI	0.0009	0.0220	0.0022	0.2756
DAX–FTSE	3×10^{-11}	0.0156	9×10^{-10}	0.6879
DAX–HSI	0.0047	0.1770	0.0087	0.4827
FTSE–HSI	0.0030	0.0043	0.0030	0.1234

Table 4: Results (p–values) from bivariate tests of constant correlations against an STCC–GARCH model with single or double transition, and bivariate tests of another transition in the TVCC–GARCH model.

model	$\hat{\rho}_{(1)}$	$\hat{\rho}_{(2)}$	$\hat{c}$	$\hat{\gamma}$
CAC–DAX	0.5457 (0.0844)	0.9553 (0.0288)	0.48 (0.08)	6.11 (2.37)
CAC–FTSE	0.6047 (0.0317)	0.8935 (0.0137)	0.60 (0.03)	15.07 (2.35)
CAC–HSI	0.2948 (0.0448)	0.5285 (0.0464)	0.54 (0.05)	43.50 (14.10)
DAX–FTSE	0.5032 (0.0817)	0.8270 (0.0337)	0.52 (0.10)	8.18 (3.72)
DAX–HSI	0.3229 (0.0425)	0.5377 (0.0352)	0.51 (0.00)	500 (—)
FTSE–HSI	0.2554 (0.1258)	0.5358 (0.0452)	0.32 (0.12)	8.66 (2.22)

Table 5: Estimation results for each of the bivariate TVCC–GARCH models. The standard errors are given in parentheses.

tion about the process causing the conditional correlations to fluctuate over time. As Silvennoinen and Teräsvirta (2005) showed, valuable information of the behaviour of the correlations can be extracted from studying submodels. For this reason we study bivariate models of stock returns before considering the full four-variate model.

Table 4 contains p–values of the tests of constant correlations against the TVCC–GARCH model, the STCC–GARCH model where the transition variable is the lagged VIX, and the TVSTCC–GARCH model. Time clearly appears to be an indicator of change in correlations: the null hypothesis is rejected for every bivariate combination of the indices. The volatility index VIX seems to be a substantially weaker indicator than time. When VIX is the transition variable, the test rejects at the 1% significance level only in three of the six cases, although five of the six p–values do remain below 0.05. When constancy is tested directly against TVSTCC–GARCH the rejections are very strong, see the fourth column of Table 4. Because all tests reject constancy of correlations in favour of variation in time, we first estimate the TVCC–GARCH model for each pair of series and then test for another transition. Note that in one model the parameter estimate of γ reaches its upper bound (500) and thus defines the transition as being nearly a break.

The resulting p–values are given in the fifth column of Table 4. For all models except the CAC–DAX one the null hypothesis is not rejected when the alternative model is the complete DSTCC–GARCH model. This indicates that some of the correlations in the complete TVSTCC–GARCH model may be constant in the VIX dimension either at the beginning or the end of the sample, or both. These scenarios can be tested by testing for another transition when the null model is the complete TVCC–GARCH model and a partially constant TVSTCC–GARCH model forms the alternative.

Because the model restricts the location of the smooth transition over time to be the same for all indices in the full four-variate model, we have to check whether imposing that type of restriction is plausible. This is done by comparing the estimated bivariate TVCC–GARCH models in Table 5. Their time-varying bivariate conditional correlations are plotted in Figure 1. All correlations increase during the sample time period. The transitions between Hang Seng and both CAC and DAX are quite abrupt whereas the ones between CAC and DAX, FTSE and DAX, and FTSE and Hang Seng are rather smooth and not completed (the transition function does not reach values zero or one) during the observation period. Although some of the locations seem to be significantly different from each other, the transitions still occur around the turn of the century, and we shall consider complete four-variate models.

Within this framework the rejection of constancy of correlations when using time as the transition variable is much stronger (the p–value equals 2×10^{-32}) than when the lagged VIX is used (the p–value equals 0.0111). Consequently, we proceed to first estimate a TVCC–GARCH model and then test for another transition. This test rejects the null model (the p–value equals 0.0045), and we estimate the full four-variate DSTCC–GARCH model. As the bivariate tests already suggest, some of the estimates of the correlations do not differ significantly at the beginning of the sample between the states in $\boldsymbol{P}_{(11)}$ and $\boldsymbol{P}_{(21)}$, or at the end of the sample between the states in $\boldsymbol{P}_{(12)}$ and $\boldsymbol{P}_{(22)}$.

We test the TVCC–GARCH model against DSTCC–GARCH models that are partially constant in the VIX dimension. As alternative models we consider different combinations of pairs of correlations that are restricted constant according to VIX (for conciseness we do not report the partial tests). There is a clear indication that the correlations between Hang Seng and the other indices do not vary according to VIX at any point in time during our observation period. Our conclusion is that those correlations are only controlled by time. We continue testing for time-variation according to VIX in the remaining correlations and are able to reject the TVCC–GARCH model only against the DSTCC–GARCH model in which VIX acts as an indicator of time-varying correlations between the European indices.

The final model and the estimated parameters can be found in Table 6. To give a visual idea of how the correlations vary over time, the estimated correlations are plotted in Figure 2. The correlations seem to have increased at the turn of the century. The estimated midpoint of the transition, 0.54, points at the spring of 1999 and agrees well with the results from bivariate models. This is in agreement with Cappiello, Engle, and Sheppard (2003). They document a structural break that implies an increase in the correlations from their previous unconditional level around

January 1999. This coincides with the introduction of Euro, which affected markets both within and outside the Euro-area.

Before the 1990's the correlations have been reported to shift to higher levels during periods of distress than they have during calm periods, a phenomenon that has become global with the increase of financial market integration, see for instance Lin, Engle, and Ito (1994) and de Santis and Gerard (1997). Our observation period starts in December 1990, and the estimation results suggest that the behaviour of the correlations may have changed when it comes to calm and turbulent periods. The correlations that actually respond to VIX behave in an opposite way to the one documented before 1990. The estimated transition due to VIX is rather abrupt, and one may thus speak about high and low volatility regimes. During the former, i.e., when the volatility index exceeds the estimated location, $\hat{c}_1 = 23.31$, which constitutes 21% of the sample, the correlations are lower than during calm periods. That is, the uncertainty of the investors shows as a decrease in correlations. Similar behaviour was found in Silvennoinen and Teräsvirta (2005) for daily returns of a pair of stocks in the S&P 500 stock index, although for a majority of them financial distress increased the correlations between returns.

6 Conclusions

In this paper we extend the Smooth Transition Conditional Correlation (STCC) GARCH model of Silvennoinen and Teräsvirta (2005). The new model, the Double Smooth Transition Conditional Correlation (DSTCC) GARCH model allows time-variation in the conditional correlations to be controlled by two transition variables instead of only one. A useful choice for one of the transition variables is simply time, in which case the model also accounts for a change in unconditional correlations over time. This is a very appealing property because in applications of GARCH models the number of observations is often quite large. The time series may be, for example, daily returns and consist of several years of data. It is not reasonable to simply assume that the correlations remain constant over years and in fact, as shown in the empirical application, and also in that of Berben and Jansen (2005a, b), this is generally not the case.

We also complement the battery of specification and misspecification tests in Silvennoinen and Teräsvirta (2005). We derive LM–tests for testing constancy of correlations against the DSTCC–GARCH model and testing whether another transition is required, i.e. testing STCC–GARCH model against DSTCC–GARCH model. We also discuss the implementation of partial constancy restrictions into the tests above. This becomes especially relevant when the number of variables in the model is large, because the tests offer an opportunity to reduce the otherwise large number of parameters to be estimated.

We apply the DSTCC–GARCH model to a set of world stock indices from Europe and Asia. As discussed in Longin and Solnik (2001), the market trend affects correlations more than volatility. We use the CBOE volatility index as one transition

variable to account for both uncertainty and volatility on the markets. The other transition variable has been time which allows the level of unconditional correlations to change over time. We find a clear upward shift in the level of unconditional correlations around the turn of the century. This change is significant both within and accross the two geographical areas, Europe and Asia. The volatility index seems to carry some information about the time-varying correlations in Europe, especially towards the end of the observation period, and estimation results suggest that the correlations tend to decrease whenever the markets grow uncertain. It is clear from this application that the extension of the original STCC–GARCH model proves useful because of its ability to separate the effects of different transition variables, in our application those being the market uncertainty and time, on conditional correlations.

Figure 1: Estimated time-varying correlations in the bivariate TVCC–GARCH models.

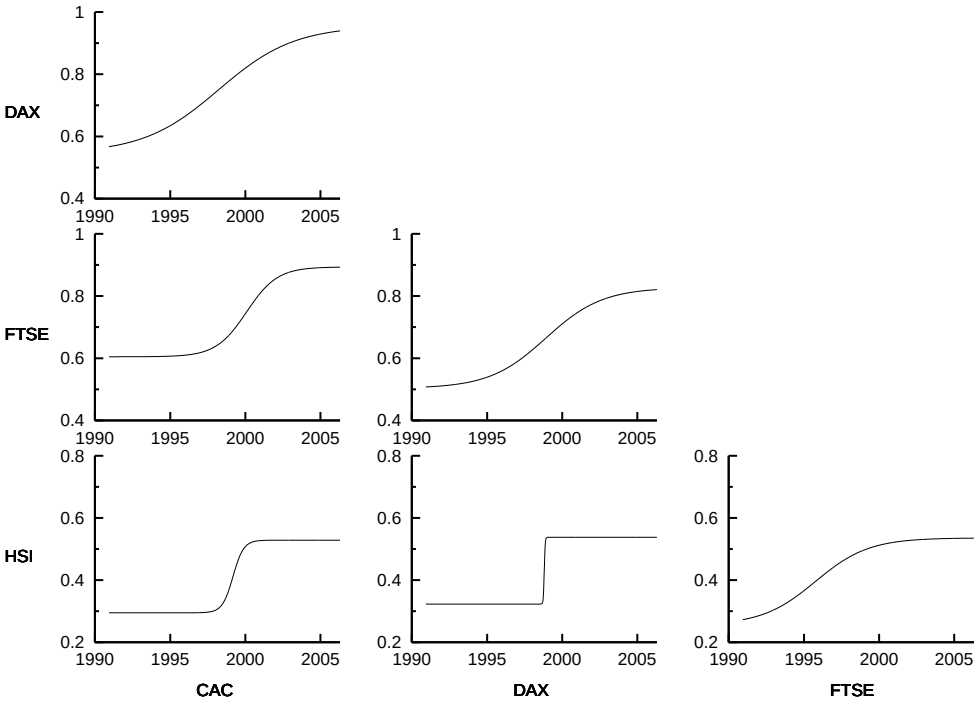
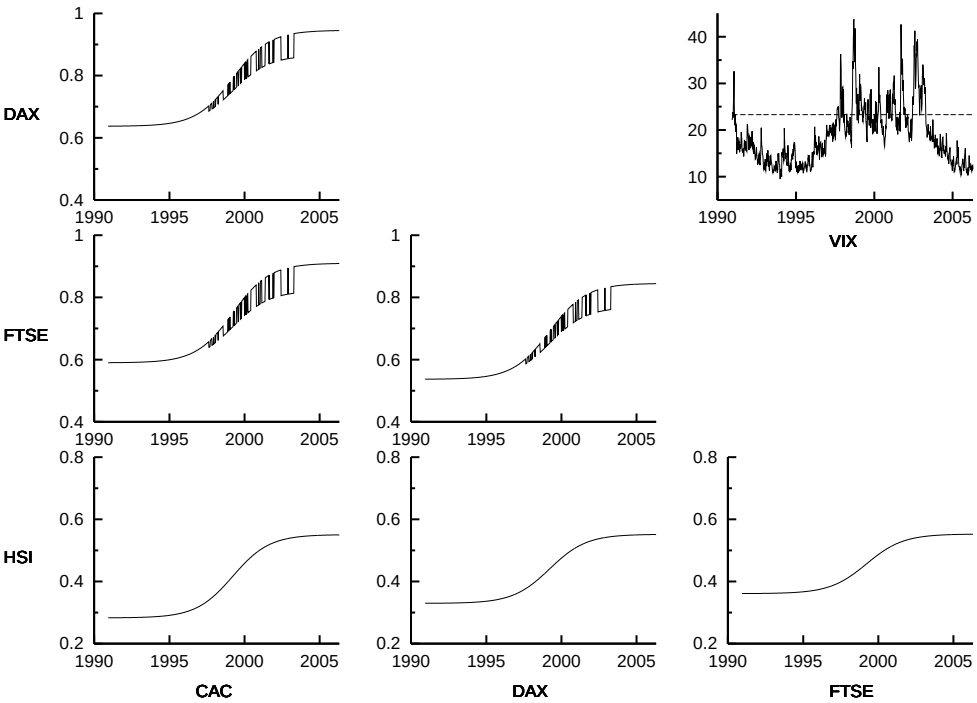


Figure 2: Estimated time-varying correlations in the four-variate TVSTCC–GARCH model. In the top right corner is the volatility index VIX and the estimated location $\hat{c}_1$ of the transition.



GARCH-parameters:

	CAC	DAX	FTSE	HSI
$\hat{\alpha}_0$	0.1252 (0.0364)	0.1134 (0.0402)	0.0880 (0.0358)	0.1641 (0.0716)
$\hat{\alpha}_1$	0.0410 (0.0078)	0.0425 (0.0092)	0.0413 (0.0104)	0.0727 (0.0147)
$\hat{\beta}_1$	0.9398 (0.0104)	0.9430 (0.0119)	0.9361 (0.0164)	0.9138 (0.0157)

s_{1t}

Correlation parameters:

$\hat{\boldsymbol{P}}_{(21)}$

	CAC	DAX	FTSE
DAX	0.6371^{r_1} (0.0308)		
FTSE	0.5898^{r_1} (0.0325)	0.5368^{r_1} (0.0370)	
HSI	0.2828^{r_1} (0.0463)	0.3296^{r_1} (0.0467)	0.3610^{r_1} (0.0437)

$\hat{\boldsymbol{P}}_{(22)}$

	CAC	DAX	FTSE
DAX	0.8650 (0.0237)		
FTSE	0.8214 (0.0328)	0.7690 (0.0362)	
HSI	0.5508^{r_1} (0.0403)	0.5518^{r_1} (0.0395)	0.5527^{r_1} (0.0400)

$\hat{\boldsymbol{P}}_{(11)}$

	CAC	DAX	FTSE
DAX	0.6371^{r_1} (0.0308)		
FTSE	0.5898^{r_1} (0.0325)	0.5368^{r_1} (0.0370)	
HSI	0.2828^{r_1} (0.0463)	0.3296^{r_1} (0.0467)	0.3610^{r_1} (0.0437)

$\hat{\boldsymbol{P}}_{(12)}$

	CAC	DAX	FTSE
DAX	0.9457 (0.0078)		
FTSE	0.9100 (0.0117)	0.8451 (0.0189)	
HSI	0.5508^{r_1} (0.0403)	0.5518^{r_1} (0.0395)	0.5527^{r_1} (0.0400)

s_{2t}

Transition parameters:

	G_{1t}
$\hat{c}_1$	23.31 (0.01)
$\hat{\gamma}_1$	500 (—)

	G_{2t}
$\hat{c}_2$	0.54 (0.02)
$\hat{\gamma}_2$	12.62 (2.16)

Table 6: Estimation results for the TVSTCC–GARCH model. The transition variable s_{1t} is one week lag of the volatility index (VIX) and s_{2t} is time in percentage. The correlation matrices $\boldsymbol{P}_{(11)}$ and $\boldsymbol{P}_{(21)}$ refer to the extreme states according to the VIX at the beginning of the sample, and similarly matrices $\boldsymbol{P}_{(12)}$ and $\boldsymbol{P}_{(22)}$ refer to those at the end of the sample. The parameters indicated by a superscript r_1 are restricted constant with respect to the transition variable s_{1t}. The transition functions G_{1t} and G_{2t} are functions of s_{1t} and s_{2t}, respectively. The standard errors are given in parentheses.

Appendix

Construction of the auxiliary null hypothesis 1: Test for another transition

The null hypothesis for the test for another transition is $\gamma_2 = 0$ in (6). When the null hypothesis is true, some of the parameters in the model cannot be identified. This problem is circumvented following Luukkonen, Saikkonen, and Teräsvirta (1988). Linearizing the transition function G_{2t} by a first-order Taylor approximation around $\gamma_2 = 0$ yields

$$G_{2t} \quad \tilde{=} \quad 1/2 + 1/4\gamma_2(s_{2t} - c_2) + \boldsymbol{R} \tag{18}$$

where $\boldsymbol{R}$ is the remainder that equals zero when the null hypothesis is valid. Thus, ignoring $\boldsymbol{R}$ and inserting (18) into (6) the dynamic correlations become

$$\begin{aligned} \boldsymbol{P}_t^* \quad = \quad & (1 - G_{1t})(1/2 - 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(11)} + (1 - G_{1t})(1/2 + 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(12)} \\ & + G_{1t}(1/2 - 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(21)} + G_{1t}(1/2 + 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(21)}. \end{aligned}$$

Rearranging the terms yields

$$\boldsymbol{P}_t^* \quad = \quad (1 - G_{1t})\boldsymbol{P}_{(1)}^* + G_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^* \tag{19}$$

where

$$\begin{aligned} \boldsymbol{P}_{(1)}^* \quad &= \quad 1/2(\boldsymbol{P}_{(11)} + \boldsymbol{P}_{(12)}) + 1/4c_2\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)}) \\ \boldsymbol{P}_{(2)}^* \quad &= \quad 1/2(\boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) + 1/4c_2\gamma_2(\boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}) \\ \boldsymbol{P}_{(3)}^* \quad &= \quad -1/4(1 - G_{1t})\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)}) - 1/4G_{1t}\gamma_2(\boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}). \end{aligned}$$

Under H_0, $\gamma_2 = 0$ and hence

$$\begin{aligned} \boldsymbol{P}_{(1)}^* \quad &= \quad 1/2(\boldsymbol{P}_{(11)} + \boldsymbol{P}_{(12)}) \\ \boldsymbol{P}_{(2)}^* \quad &= \quad 1/2(\boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) \\ \boldsymbol{P}_{(3)}^* \quad &= \quad \boldsymbol{0}_{N\times N} \end{aligned}$$

and the model collapses to the STCC–GARCH model with correlations varying according to s_{1t}. The auxiliary null hypothesis is therefore

$$H_0^{aux} \quad : \quad vecl\boldsymbol{P}_{(3)}^* = \boldsymbol{0}_{N(N-1)/2\times 1}$$

in (19).[2] The LM–test of this auxiliary null hypothesis is carried out in the usual way and the test statistic is χ^2 distributed with $N(N-1)/2$ degrees of freedom. The construction of the LM–test is discussed in a later subsection.

Construction of the auxiliary null hypothesis 2: Test of constant correlations against an STCC–GARCH model with two transitions

The constancy of correlations hypothesis is equivalent to $\gamma_1 = \gamma_2 = 0$ in (6). The problem with unidentified parameteres under the null is avoided by linearizing both transition functions, G_{1t} and G_{2t}, by first-order Taylor expansions around $\gamma_1 = 0$ and $\gamma_2 = 0$, respectively. This yields

$$G_{it} \quad \tilde{=} \quad 1/2 + 1/4\gamma_i(s_{it} - c_i) + \boldsymbol{R}_i, \qquad i = 1, 2. \tag{20}$$

[2]The notation $vecl\boldsymbol{P}$ is used to denote the vec-operator applied to the strictly lower triangular part of the square matrix $\boldsymbol{P}$.

G_{it} does not carry any approximation error under the null. Replacing the transition functions in (6) with the equations (20) gives

$$\begin{aligned}\boldsymbol{P}_t^* &= (1/2 - 1/4\gamma_1(s_{1t} - c_1))(1/2 - 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(11)} \\ &\quad +(1/2 - 1/4\gamma_1(s_{1t} - c_1))(1/2 + 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(12)} \\ &\quad +(1/2 + 1/4\gamma_1(s_{1t} - c_1))(1/2 - 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(21)} \\ &\quad +(1/2 + 1/4\gamma_1(s_{1t} - c_1))(1/2 + 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(22)}.\end{aligned}$$

Rearranging the terms gives

$$\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)}^* + s_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^* + s_{1t}s_{2t}\boldsymbol{P}_{(4)}^* \tag{21}$$

where

$$\begin{aligned}\boldsymbol{P}_{(1)}^* &= 1/4(\boldsymbol{P}_{(11)} + \boldsymbol{P}_{(12)} + \boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) \\ &\quad +1/8c_1\gamma_1(\boldsymbol{P}_{(11)} + \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}) \\ &\quad +1/8c_2\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} + \boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}) \\ &\quad +1/16c_1c_2\gamma_1\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) \\ \boldsymbol{P}_{(2)}^* &= -1/8\gamma_1(\boldsymbol{P}_{(11)} + \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}) \\ &\quad -1/16c_2\gamma_1\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) \\ \boldsymbol{P}_{(3)}^* &= -1/8\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} + \boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}) \\ &\quad -1/16c_1\gamma_1\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) \\ \boldsymbol{P}_{(4)}^* &= 1/16\gamma_1\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} - \boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}).\end{aligned}$$

Under H_0, $\gamma_1 = \gamma_2 = 0$ and hence

$$\begin{aligned}\boldsymbol{P}_{(1)}^* &= 1/4(\boldsymbol{P}_{(11)} + \boldsymbol{P}_{(12)} + \boldsymbol{P}_{(21)} + \boldsymbol{P}_{(22)}) \\ \boldsymbol{P}_{(2)}^* &= \boldsymbol{0}_{N\times N} \\ \boldsymbol{P}_{(3)}^* &= \boldsymbol{0}_{N\times N} \\ \boldsymbol{P}_{(4)}^* &= \boldsymbol{0}_{N\times N}.\end{aligned}$$

Therefore, the auxiliary null hypothesis is stated as

$$H_0^{aux} \quad : \quad vecl\boldsymbol{P}_{(2)}^* = vecl\boldsymbol{P}_{(3)}^* = vecl\boldsymbol{P}_{(4)}^* = \boldsymbol{0}_{N(N-1)/2\times 1}$$

in (21). The test statistic for the LM–test for this auxiliary null hypothesis is χ^2 distributed with $3N(N-1)/2$ degrees of freedom. The details of the construction of the LM–test are discussed in a later subsection.

Construction of the auxiliary null hypothesis 3: Test of constant correlations against an STCC–GARCH model with two transitions, transition variables are independent

The constancy of correlations hypothesis is imposed by setting $\gamma_1 = \gamma_2 = 0$ in (10). The problem with unidentified parameters under the null is avoided by linearizing both transition functions, G_{1t} and G_{2t}, by first-order Taylor expansions around $\gamma_1 = 0$ and $\gamma_2 = 0$, respectively. Replacing the transition functions in (10) by the linearized ones (20) gives

$$\begin{aligned}\boldsymbol{P}_t^* &= (-1/4\gamma_1(s_{1t} - c_1) - 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(11)} \\ &\quad +(1/2 + 1/4\gamma_1(s_{1t} - c_1))\boldsymbol{P}_{(21)} \\ &\quad +(1/2 + 1/4\gamma_2(s_{2t} - c_2))\boldsymbol{P}_{(12)}.\end{aligned}$$

Rearranging the terms gives

$$\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)}^* + s_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^* \tag{22}$$

where

$$\begin{aligned}
\boldsymbol{P}_{(1)}^* &= 1/2(\boldsymbol{P}_{(12)} + \boldsymbol{P}_{(21)}) + 1/4c_1\gamma_1(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(21)}) + 1/4c_2\gamma_2(\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)}) \\
\boldsymbol{P}_{(2)}^* &= 1/4\gamma_1(\boldsymbol{P}_{(21)} - \boldsymbol{P}_{(11)}) \\
\boldsymbol{P}_{(3)}^* &= 1/4\gamma_2(\boldsymbol{P}_{(12)} - \boldsymbol{P}_{(11)}).
\end{aligned}$$

Under H_0, $\gamma_1 = \gamma_2 = 0$ and hence

$$\begin{aligned}
\boldsymbol{P}_{(1)}^* &= 1/2(\boldsymbol{P}_{(12)} + \boldsymbol{P}_{(21)}) \\
\boldsymbol{P}_{(2)}^* &= \mathbf{0}_{N\times N} \\
\boldsymbol{P}_{(3)}^* &= \mathbf{0}_{N\times N}.
\end{aligned}$$

Therefore, the auxiliary null hypothesis is stated as

$$H_0^{aux} \quad : \quad vecl\boldsymbol{P}_{(2)}^* = vecl\boldsymbol{P}_{(3)}^* = \mathbf{0}_{N(N-1)/2\times 1}$$

in (22). The test statistic for the LM–test for this auxiliary null hypothesis is χ^2 distributed with $N(N-1)$ degrees of freedom. The details of the construction of the LM–test are discussed in a later subsection.

Construction of LM(/Wald)–statistic

Let $\boldsymbol{\theta}_0$ be the vector of true parameters. Under suitable assumptions and regularity conditions,

$$\sqrt{T}^{-1}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{d} N(0, \mathfrak{I}(\boldsymbol{\theta}_0)). \tag{23}$$

To derive LM–statistics of the null hypothesis consider the following quadratic form:

$$T^{-1}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\mathfrak{I}(\boldsymbol{\theta}_0)^{-1}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = T^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right)\mathfrak{I}(\boldsymbol{\theta}_0)^{-1}\left(\sum_{t=1}^{T}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)$$

and evaluate it at the maximum likelihood estimators under the restriction of the null hypothesis. The limiting information matrix $\mathfrak{I}(\boldsymbol{\theta}_0)$ is replaced by the consistent estimator

$$\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0) = T^{-1}\sum_{t=1}^{T} E\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mid \mathcal{F}_{t-1}\right]. \tag{24}$$

The following derivations are straightforward implications of the definitions and elementary rules of matrix algebra. Results in Anderson (2003) and Lütkepohl (1996) are heavily relied upon.

Test of constant conditional correlations against a DSTCC–GARCH model

The model under the null is the CCC–GARCH model. The alternative model is an STCC–GARCH model with two transitions where the correlations are controlled by the transition variables s_{1t} and s_{2t}. Under the null, the linearized time-varying correlation matrix is $\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)}^* + s_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^* + s_{1t}s_{2t}\boldsymbol{P}_{(4)}^*$ as defined in (21). To construct the test statistic we introduce some simplifying notation. Let $\boldsymbol{\omega}_i = (\alpha_{i0}, \alpha_i, \beta_i)'$, $i = 1, \ldots, N$, denote the parameter vectors of the GARCH equations, and $\boldsymbol{\rho}^* = (\boldsymbol{\rho}_{(1)}^{*\prime}, \ldots, \boldsymbol{\rho}_{(4)}^{*\prime})'$, where

$\boldsymbol{\rho}^*_{(j)} = vecl\boldsymbol{P}^*_{(j)}$, $j = 1, \ldots, 4$, are the vectors holding all unique off-diagonal elements in the four matrices $\boldsymbol{P}^*_{(1)}, \ldots, \boldsymbol{P}^*_{(4)}$, respectively. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}'_1, \ldots, \boldsymbol{\omega}'_N, \boldsymbol{\rho}^{*\prime})'$ be the full parameter vector and $\boldsymbol{\theta}_0$ the corresponding vector of true parameters under the null. Furthermore, let $\boldsymbol{v}_{it} = (1, \varepsilon^2_{it}, h_{it})'$, $i = 1, \ldots, N$, and $\boldsymbol{v}_{\boldsymbol{\rho}^* t} = (1, s_{1t}, s_{2t}, s_{1t}s_{2t})'$. Symbols $\otimes$ and $\odot$ represent the Kronecker and Hadamard products of two matrices, respectively. Let 1_i be an $N \times 1$ vector of zeros with ith element equal to one and $\mathbf{1}_n$ be an $n \times n$ matrix of ones. The identity matrix $\boldsymbol{I}$ is of size N unless otherwise indicated by a subscript.

Consider the log-likelihood function for observation t as defined in (11) with linearized time-varying correlation matrix:

$$l_t(\boldsymbol{\theta}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\log(h_{it}) - \frac{1}{2}\log|\boldsymbol{P}^*_t| - \frac{1}{2}\boldsymbol{z}'_t\boldsymbol{P}^{*-1}_t\boldsymbol{z}_t.$$

The first order derivatives of the log-likelihood function with respect to the GARCH and correlation parameters are

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\omega}_i} &= -\frac{1}{2h_{it}}\frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i}\left\{1 - z_{it}1'_i\boldsymbol{P}^{*-1}_t\boldsymbol{z}_t\right\}, \quad i = 1, \ldots, N \\
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\rho}^*} &= -\frac{1}{2}\frac{\partial(vec\boldsymbol{P}^*_t)'}{\partial \boldsymbol{\rho}^*}\left\{vec\boldsymbol{P}^{*-1}_t - \left(\boldsymbol{P}^{*-1}_t \otimes \boldsymbol{P}^{*-1}_t\right)(\boldsymbol{z}_t \otimes \boldsymbol{z}_t)\right\}
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i} &= \boldsymbol{v}_{i,t-1} + \beta_i\frac{\partial h_{i,t-1}}{\partial \boldsymbol{\omega}_i}, \quad i = 1, \ldots, N \\
\frac{\partial(vec\boldsymbol{P}^*_t)'}{\partial \boldsymbol{\rho}^*} &= \boldsymbol{v}_{\boldsymbol{\rho}^* t} \otimes \boldsymbol{U}'.
\end{aligned}$$

The matrix $\boldsymbol{U}$ is an $N^2 \times \frac{N(N-1)}{2}$ matrix of zeros and ones, whose columns are defined as

$$\left[vec\left(1_i1'_j + 1_j1'_i\right)\right]_{i=1,\ldots N-1, j=i+1,\ldots,N}$$

and the columns appear in the same order from left to right as the indices in $vecl\boldsymbol{P}_t$. Under the null hypothesis $\boldsymbol{\rho}^*_{(2)} = \boldsymbol{\rho}^*_{(3)} = \boldsymbol{\rho}^*_{(4)} = 0$, and thus the derivatives at the true parameter values under the null can be written as

$$\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i} = -\frac{1}{2h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\left\{1 - z_{it}1'_i\boldsymbol{P}^{*-1}_{(1)}\boldsymbol{z}_t\right\}, \quad i = 1, \ldots, N \tag{25}$$

$$\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^*} = -\frac{1}{2}\frac{\partial(vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}^*}\left\{vec\boldsymbol{P}^{*-1}_{(1)} - \left(\boldsymbol{P}^{*-1}_{(1)} \otimes \boldsymbol{P}^{*-1}_{(1)}\right)(\boldsymbol{z}_t \otimes \boldsymbol{z}_t)\right\}. \tag{26}$$

Taking conditional expectations of the cross products of (25) and (26) yields, for $i, j = 1, \ldots, N$,

$$\begin{aligned}
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_i}\right] &= \frac{1}{4h^2_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_i}\left(1 + 1'_i\boldsymbol{P}^{*-1}_{(1)}1_i\right) \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_j}\right] &= \frac{1}{4h_{it}h_{jt}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial h_{jt}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_j}\left(\rho^*_{1,ij}1'_i\boldsymbol{P}^{*-1}_{(1)}1_j\right), \quad i \neq j \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^*}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}^*}\left(\boldsymbol{P}^{*-1}_{(1)} \otimes \boldsymbol{P}^{*-1}_{(1)} + \left(\boldsymbol{P}^{*-1}_{(1)} \otimes \boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}^{*-1}_{(1)} \otimes \boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}} \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}\right] &= \frac{1}{4h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\left(1'_i\boldsymbol{P}^{*-1}_{(1)} \otimes 1'_i + 1'_i \otimes 1'_i\boldsymbol{P}^{*-1}_{(1)}\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}
\end{aligned} \tag{27}$$

where

$$\boldsymbol{K} = \begin{bmatrix} 1_1 1_1' & \cdots & 1_N 1_1' \\ \vdots & \ddots & \vdots \\ 1_1 1_N' & \cdots & 1_N 1_N' \end{bmatrix}. \tag{28}$$

For the derivation of the expressions (27), see Silvennoinen and Teräsvirta (2005).

The estimator of the information matrix is obtained by making use of the submatrices in (27). For a more compact expression, let $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}', \ldots, \boldsymbol{x}_{Nt}')'$ where $\boldsymbol{x}_{it} = -\frac{1}{2h_{it}} \frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i}$, and let $\boldsymbol{x}_{\boldsymbol{\rho}^* t} = -\frac{1}{2} \frac{\partial (vec \boldsymbol{P}_t^*)'}{\partial \boldsymbol{\rho}^*}$, and let $\boldsymbol{x}_{it}^0$, $i = 1, \ldots, N, \boldsymbol{\rho}^*$, denote the corresponding expressions evaluated at the true values under the null hypothesis. Setting

$$\begin{aligned}
\boldsymbol{M}_1 &= T^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t^0 \boldsymbol{x}_t^{0\prime} \odot \left(\left(\boldsymbol{I} + \boldsymbol{P}_{(1)}^* \odot \boldsymbol{P}_{(1)}^{*-1} \right) \otimes \boldsymbol{1}_3 \right) \\
\boldsymbol{M}_2 &= T^{-1} \sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{x}_{1t}^0 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{x}_{Nt}^0 \end{bmatrix} \begin{bmatrix} 1_1' \boldsymbol{P}_{(1)}^{*-1} \otimes 1_1' + 1_1' \otimes 1_1' \boldsymbol{P}_{(1)}^{*-1} \\ \vdots \\ 1_N' \boldsymbol{P}_{(1)}^{*-1} \otimes 1_N' + 1_N' \otimes 1_N' \boldsymbol{P}_{(1)}^{*-1} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{\rho}^* t}^{0\prime} \\
\boldsymbol{M}_3 &= T^{-1} \sum_{t=1}^{T} \boldsymbol{x}_{\boldsymbol{\rho}^* t}^0 \left(\boldsymbol{P}_{(1)}^{*-1} \otimes \boldsymbol{P}_{(1)}^{*-1} + \left(\boldsymbol{P}_{(1)}^{*-1} \otimes \boldsymbol{I} \right) \boldsymbol{K} \left(\boldsymbol{P}_{(1)}^{*-1} \otimes \boldsymbol{I} \right) \right) \boldsymbol{x}_{\boldsymbol{\rho}^* t}^{0\prime}
\end{aligned}$$

the information matrix $\mathfrak{I}(\boldsymbol{\theta}_0)$ is approximated by

$$\begin{aligned}
\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0) &= T^{-1} \sum_{t=1}^{T} E \left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mid \mathcal{F}_{t-1} \right] \\
&= \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{M}_2' & \boldsymbol{M}_3 \end{bmatrix}.
\end{aligned}$$

The block corresponding to the correlation parameters of the inverse of $\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0)$ can be calculated as

$$\left(\boldsymbol{M}_3 - \boldsymbol{M}_2' \boldsymbol{M}_1^{-1} \boldsymbol{M}_2 \right)^{-1}$$

from where the south-east $\frac{3N(N-1)}{2} \times \frac{3N(N-1)}{2}$ block corresponding to $\boldsymbol{\rho}_{(2)}^*$, $\boldsymbol{\rho}_{(3)}^*$, and $\boldsymbol{\rho}_{(4)}^*$ can be extracted. Replacing the true unknown values with maximum likelihood estimators, the test statistic simplifies to

$$T^{-1} \left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial (\boldsymbol{\rho}_{(2)}^{*\prime}, \boldsymbol{\rho}_{(3)}^{*\prime}, \boldsymbol{\rho}_{(4)}^{*\prime})} \right) \left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}}) \right]^{-1}_{(\boldsymbol{\rho}_{(2-4)}^*, \boldsymbol{\rho}_{(2-4)}^*)} \left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial (\boldsymbol{\rho}_{(2)}^{*\prime}, \boldsymbol{\rho}_{(3)}^{*\prime}, \boldsymbol{\rho}_{(4)}^{*\prime})'} \right) \tag{29}$$

where $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho}_{(2-4)}^*, \boldsymbol{\rho}_{(2-4)}^*)}$ is the block of the inverse of $\hat{\mathfrak{I}}_T$ corresponding to those correlation parameters that are set to zero under the null. It follows from (23) and consistency and asymptotic normality of ML estimators that the statistic (29) has an asymptotic $\chi^2_{\frac{3N(N-1)}{2}}$ distribution when the null hypothesis is valid.

Test of constant conditional correlations against a partially constant DSTCC–GARCH model

The test of constant correlations of previous subsection is not affected unless one or more of the parameters are restricted to be constant according to one of the transition variables in both extreme states described by the other transition variable. In those cases certain

parameters in the linearized time-varying correlation matrix $\boldsymbol{P}_t^*$ in (21) are set to zero. Let there be k pairs of correlation parameters that, under the alternative hypothesis, are restricted to be constant with respect to the transition variable s_{t1} in both extreme states described by s_{t2}. That is, there are k pairs of restrictions as follows:

$$\rho_{(11)ij} = \rho_{(21)ij} \quad \text{and} \quad \rho_{(12)ij} = \rho_{(22)ij}, \quad i > j$$

where $\rho_{(mn)ij}$ is the ij-element of the correlation matrix $\boldsymbol{P}_{(mn)}$ in (6). In the linearized correlation matrix $\boldsymbol{P}_t^*$, the k elements in each of the matrices $\boldsymbol{P}_{(2)}^*$ and $\boldsymbol{P}_{(4)}^*$ corresponding to these restrictions are set to zero. Similarly, let there be l pairs of correlation parameters that, under the alternative hypothesis, are restricted to be constant with respect to s_{t2} in both extreme states described by s_{t1}. The pairs of restrictions are then

$$\rho_{(11)ij} = \rho_{(12)ij} \quad \text{and} \quad \rho_{(21)ij} = \rho_{(22)ij}, \quad i > j.$$

In the linearized correlation matrix $\boldsymbol{P}_t^*$ the l elements in each of the matrices $\boldsymbol{P}_{(3)}^*$ and $\boldsymbol{P}_{(4)}^*$ corresponding to these restrictions are set to zero. The vector of correlation parameters $\boldsymbol{\rho}^*$ is formed as before but the elements corresponding to the restrictions, i.e., the elements that were set to zero, are excluded. Furthermore,

$$\frac{\partial (vec\boldsymbol{P}_t^*)'}{\partial \boldsymbol{\rho}^*}$$

is defined as before, but with m (m equals $2k + 2l$ less the number of possibly overlapping restrictions) rows deleted so that the remaining rows correspond to the elements in $\boldsymbol{\rho}^*$. The same rows are also deleted from $\boldsymbol{x}_{\rho^* t}$. With these modifications the test statistic is as in (29) above, and its asymptotic distribution under the null hypothesis is $\chi^2_{\frac{3N(N-1)}{2}-m}$.

Test of constant conditional correlations against a DSTCC–GARCH model whose transition variables are independent

The model under the null is the CCC–GARCH model. The alternative model is an STCC–GARCH model with two transitions where the correlations are varying according to the the transition variables s_{1t} and s_{2t} and the restriction $\boldsymbol{P}_{(11)} - \boldsymbol{P}_{(12)} = \boldsymbol{P}_{(21)} - \boldsymbol{P}_{(22)}$ holds. Under the null, the linearized time-varying correlation matrix is $\boldsymbol{P}_t^* = \boldsymbol{P}_{(1)}^* + s_{1t}\boldsymbol{P}_{(2)}^* + s_{2t}\boldsymbol{P}_{(3)}^*$ as defined in (22). The statistic is constructed as in the case of testing constancy of correlations against DSTCC–GARCH model but with following modifications: Let $\boldsymbol{\rho}^* = (\boldsymbol{\rho}_{(1)}^{*\prime}, \boldsymbol{\rho}_{(2)}^{*\prime}, \boldsymbol{\rho}_{(3)}^{*\prime})'$, where $\boldsymbol{\rho}_{(j)}^* = vecl\boldsymbol{P}_{(j)}^*$, $j = 1, 2, 3$, are the vectors holding all the unique off-diagonal elements in the four matrices $\boldsymbol{P}_{(1)}^*, \boldsymbol{P}_{(2)}^*, \boldsymbol{P}_{(3)}^*$, respectively. Furthermore, define $\boldsymbol{v}_{\rho^* t} = (1, s_{1t}, s_{2t})'$. With these changes the test statistic is as constructed as before, and the block corresponding to the correlation parameters of the inverse of $\hat{\mathfrak{I}}_T(\boldsymbol{\theta}_0)$ can be calculated as

$$\left(\boldsymbol{M}_3 - \boldsymbol{M}_2'\boldsymbol{M}_1^{-1}\boldsymbol{M}_2\right)^{-1}$$

from where the south-east $N(N-1) \times N(N-1)$ block corresponding to $\boldsymbol{\rho}_{(2)}^*$ and $\boldsymbol{\rho}_{(3)}^*$ can be extracted. Replacing the true unknown values with maximum likelihood estimators, the test statistic simplifies to

$$T^{-1}\left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial(\boldsymbol{\rho}_{(2)}^{*\prime}, \boldsymbol{\rho}_{(3)}^{*\prime})}\right)\left[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_{(2-3)}^*, \boldsymbol{\rho}_{(2-3)}^*)}\left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial(\boldsymbol{\rho}_{(2)}^{*\prime}, \boldsymbol{\rho}_{(3)}^{*\prime})'}\right) \tag{30}$$

where $[\hat{\mathfrak{I}}_T(\hat{\boldsymbol{\theta}})]^{-1}_{(\boldsymbol{\rho}^*_{(2-3)},\boldsymbol{\rho}^*_{(2-3)})}$ is the block of the inverse of $\hat{\mathfrak{I}}_T$ corresponding to those correlation parameters that are set to zero under the null. It follows from (23) and consistency and asymptotic normality of ML estimators that the statistic (30) has an asymptotic $\chi^2_{N(N-1)}$ distribution when the null hypothesis is valid.

Testing for the additional transition in the STCC–GARCH model

The model under the null is an STCC–GARCH model where the correlations are varying according to the transition variable s_{1t}. The transition that we wish to test for is a function of s_{2t}. Under the null, the linearized time-varying correlation matrix is $\boldsymbol{P}^*_t = (1-G_{1t})\boldsymbol{P}^*_{(1)} + G_{1t}\boldsymbol{P}^*_{(2)} + s_{2t}\boldsymbol{P}^*_{(3)}$ as defined in (19). The notation is as in the previous subsection with the following modifications: Let $\boldsymbol{\rho}^* = (\boldsymbol{\rho}^{*\prime}_{(1)}, \boldsymbol{\rho}^{*\prime}_{(2)}, \boldsymbol{\rho}^{*\prime}_{(3)})'$ where $\boldsymbol{\rho}^*_{(j)} = vecl\boldsymbol{P}^*_{(j)}$, $i = 1,\ldots,3$, and $\boldsymbol{\varphi} = (c_1, \gamma_1)'$. Let $\boldsymbol{\theta} = (\boldsymbol{\omega}'_1, \ldots, \boldsymbol{\omega}'_N, \boldsymbol{\rho}^{*\prime}, \boldsymbol{\varphi}')'$ be the full parameter vector, and $\boldsymbol{\theta}_0$ the corresponding vector of the true parameters under the null. Let $\boldsymbol{v}_{\boldsymbol{\rho}^* t} = (1-G_{1t}, G_{1t}, s_{2t})'$, and let $\boldsymbol{v}_{\boldsymbol{\varphi} t} = (\gamma_1, c_1 - s_{1t})'$.

The first order derivatives of the log-likelihood function with respect to the GARCH, correlation, and transition parameters are

$$\begin{aligned}
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\omega}_i} &= -\frac{1}{2h_{it}}\frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i}\left\{1 - z_{it}1'_i\boldsymbol{P}^{*-1}_t\boldsymbol{z}_t\right\}, \quad i = 1,\ldots,N \\
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\rho}^*} &= -\frac{1}{2}\frac{\partial\,(vec\boldsymbol{P}^*_t)'}{\partial \boldsymbol{\rho}^*}\left\{vec\boldsymbol{P}^{*-1}_t - \left(\boldsymbol{P}^{*-1}_t \otimes \boldsymbol{P}^{*-1}_t\right)(\boldsymbol{z}_t \otimes \boldsymbol{z}_t)\right\} \\
\frac{\partial l_t(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} &= -\frac{1}{2}\frac{\partial\,(vec\boldsymbol{P}^*_t)'}{\partial \boldsymbol{\varphi}}\left\{vec\boldsymbol{P}^{*-1}_t - \left(\boldsymbol{P}^{*-1}_t \otimes \boldsymbol{P}^{*-1}_t\right)(\boldsymbol{z}_t \otimes \boldsymbol{z}_t)\right\}
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial h_{it}}{\partial \boldsymbol{\omega}_i} &= \boldsymbol{v}_{i,t-1} + \beta_i\frac{\partial h_{i,t-1}}{\partial \boldsymbol{\omega}_i}, \quad i = 1,\ldots,N \\
\frac{\partial\,(vec\boldsymbol{P}^*_t)'}{\partial \boldsymbol{\rho}^*} &= \boldsymbol{v}_{\boldsymbol{\rho}^* t} \otimes \boldsymbol{U}' \\
\frac{\partial\,(vec\boldsymbol{P}^*_t)'}{\partial \boldsymbol{\varphi}} &= \boldsymbol{v}_{\boldsymbol{\varphi} t}(1-G_{1t})G_{1t}vec(\boldsymbol{P}^*_{(1)} - \boldsymbol{P}^*_{(2)})'.
\end{aligned}$$

Evaluating the score at the true parameters under the null and taking conditional expectations of the cross products of the first-order derivatives gives

$$\begin{aligned}
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_i}\right] &= \frac{1}{4h^2_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_i}\left(1 + 1'_i\boldsymbol{P}^{*0^{-1}}_t 1_i\right) \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_j}\right] &= \frac{1}{4h_{it}h_{jt}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial h_{jt}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}'_j}\left(\rho^{*0}_{t,ij}1'_i\boldsymbol{P}^{*0^{-1}}_t 1_j\right), \quad i \neq j \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}\right] &= \frac{1}{4h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\left(1'_i\boldsymbol{P}^{*0^{-1}}_t \otimes 1'_i + 1'_i \otimes 1'_i\boldsymbol{P}^{*0^{-1}}_t\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}} \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{4h_{it}}\frac{\partial h_{it}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\omega}_i}\left(1'_i\boldsymbol{P}^{*0^{-1}}_t \otimes 1'_i + 1'_i \otimes 1'_i\boldsymbol{P}^{*0^{-1}}_t\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'} \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^*}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}^*}\left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{P}^{*0^{-1}}_t + \left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^{*\prime}} \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\rho}^*}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\rho}^*}\left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{P}^{*0^{-1}}_t + \left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'} \\
E_{t-1}\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}}\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right] &= \frac{1}{4}\frac{\partial(vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0))'}{\partial \boldsymbol{\varphi}}\left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{P}^{*0^{-1}}_t + \left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{I}\right)\boldsymbol{K}\left(\boldsymbol{P}^{*0^{-1}}_t \otimes \boldsymbol{I}\right)\right)\frac{\partial vec\boldsymbol{P}^*_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}
\end{aligned} \tag{31}$$

where $\boldsymbol{K}$ is defined as before and $\boldsymbol{P}_t^{*0}$ is $\boldsymbol{P}_t^*$ evaluated at the true parameters under the null.

The estimator of the information matrix is obtained by using the submatrices in (31). To make the expression more compact, let $\boldsymbol{x}_t = (\boldsymbol{x}_{1t}', \ldots, \boldsymbol{x}_{Nt}')'$ where $\boldsymbol{x}_{it} = -\frac{1}{2h_{it}}\frac{\partial h_{it}}{\partial \omega_i}$. Furthermore, let $\boldsymbol{x}_{\boldsymbol{\rho}^* t} = -\frac{1}{2}\frac{\partial (vec\boldsymbol{P}_t^*)'}{\partial \boldsymbol{\rho}^*}$, and $\boldsymbol{x}_{\boldsymbol{\varphi} t} = -\frac{1}{2}\frac{\partial (vec\boldsymbol{P}_t^*)'}{\partial \boldsymbol{\varphi}}$. Finally, let $\boldsymbol{x}_{it}^0$, $i = 1, \ldots, N, \boldsymbol{\rho}, \boldsymbol{\varphi}$, denote the corresponding expressions evaluated at the true parameters under the null. Setting

$$
\begin{aligned}
\boldsymbol{M}_1 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_t^0 \boldsymbol{x}_t^{0\prime} \odot \left(\left(\boldsymbol{I} + \boldsymbol{P}_t^{*0} \odot \boldsymbol{P}_t^{*0^{-1}}\right) \otimes \boldsymbol{1}_3\right) \\
\boldsymbol{M}_2 &= T^{-1}\sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{x}_{1t}^0 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{x}_{Nt}^0 \end{bmatrix} \begin{bmatrix} 1_1' \boldsymbol{P}_t^{*0^{-1}} \otimes 1_1' + 1_1' \otimes 1_1' \boldsymbol{P}_t^{*0^{-1}} \\ \vdots \\ 1_N' \boldsymbol{P}_t^{*0^{-1}} \otimes 1_N' + 1_N' \otimes 1_N' \boldsymbol{P}_t^{*0^{-1}} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{\rho}^* t}^{0\prime} \\
\boldsymbol{M}_3 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_{\boldsymbol{\rho}^* t}^0 \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{P}_t^{*0^{-1}} + \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{I}\right)\right) \boldsymbol{x}_{\boldsymbol{\rho}^* t}^{0\prime} \\
\boldsymbol{M}_4 &= T^{-1}\sum_{t=1}^{T} \begin{bmatrix} \boldsymbol{x}_{1t}^0 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{x}_{Nt}^0 \end{bmatrix} \begin{bmatrix} 1_1' \boldsymbol{P}_t^{*0^{-1}} \otimes 1_1' + 1_1' \otimes 1_1' \boldsymbol{P}_t^{*0^{-1}} \\ \vdots \\ 1_N' \boldsymbol{P}_t^{*0^{-1}} \otimes 1_N' + 1_N' \otimes 1_N' \boldsymbol{P}_t^{*0^{-1}} \end{bmatrix} \boldsymbol{x}_{\boldsymbol{\varphi} t}^{0\prime} \\
\boldsymbol{M}_5 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_{\boldsymbol{\rho}^* t}^0 \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{P}_t^{*0^{-1}} + \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{I}\right)\right) \boldsymbol{x}_{\boldsymbol{\varphi} t}^{0\prime} \\
\boldsymbol{M}_6 &= T^{-1}\sum_{t=1}^{T} \boldsymbol{x}_{\boldsymbol{\varphi} t}^0 \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{P}_t^{*0^{-1}} + \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{I}\right) \boldsymbol{K} \left(\boldsymbol{P}_t^{*0^{-1}} \otimes \boldsymbol{I}\right)\right) \boldsymbol{x}_{\boldsymbol{\varphi} t}^{0\prime}
\end{aligned}
$$

the information matrix $\Im(\boldsymbol{\theta}_0)$ is approximated by

$$
\begin{aligned}
\hat{\Im}_T(\boldsymbol{\theta}_0) &= T^{-1}\sum_{t=1}^{T} E\left[\frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mid \mathcal{F}_{t-1}\right] \\
&= \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 & \boldsymbol{M}_4 \\ \boldsymbol{M}_2' & \boldsymbol{M}_3 & \boldsymbol{M}_5 \\ \boldsymbol{M}_4' & \boldsymbol{M}_5' & \boldsymbol{M}_6 \end{bmatrix}.
\end{aligned}
$$

The block of the inverse of $\hat{\Im}_T(\boldsymbol{\theta}_0)$ corresponding to the correlation and transition parameters is given by

$$
\left(\begin{bmatrix} \boldsymbol{M}_3 & \boldsymbol{M}_5 \\ \boldsymbol{M}_5' & \boldsymbol{M}_6 \end{bmatrix} - \begin{bmatrix} \boldsymbol{M}_2' \\ \boldsymbol{M}_4' \end{bmatrix} \boldsymbol{M}_1^{-1} \begin{bmatrix} \boldsymbol{M}_2 & \boldsymbol{M}_4 \end{bmatrix}\right)^{-1}
$$

from where the $\frac{N(N-1)}{2} \times \frac{N(N-1)}{2}$ block corresponding to $\boldsymbol{\rho}_{(3)}^*$ can be extracted. Replacing the true unknown parameter values with their maximum likelihood estimators, the test statistic simplifies to

$$
T^{-1}\left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\rho}_{(3)}^{*\prime}}\right) \left[\hat{\Im}_T(\hat{\boldsymbol{\theta}})\right]^{-1}_{(\boldsymbol{\rho}_{(3)}^*, \boldsymbol{\rho}_{(3)}^*)} \left(\sum_{t=1}^{T} \frac{\partial l_t(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\rho}_{(3)}^*}\right) \tag{32}
$$

where $[\hat{\Im}_T(\boldsymbol{\theta}_0)]^{-1}_{(\boldsymbol{\rho}_{(3)}^*, \boldsymbol{\rho}_{(3)}^*)}$ is the block of the inverse of $\hat{\Im}_T$ corresponding to those correlation parameters that are set to zero under the null. (32) has an asymptotic $\chi^2_{\frac{N(N-1)}{2}}$ distribution when the null is true.

Test of the partially constant STCC–GARCH model against a partially constant DSTCC–GARCH model

When testing the hypothesis that some of the correlation parameters are constant according to the transition variable s_{2t} in both extreme states described by variable s_{1t}, the following modifications need to be done to the testing procedure of the previous subsection: Let there be k pairs of correlation restrictions in the alternative hypothesis of the form

$$\rho_{(11)ij} = \rho_{(12)ij} \quad \text{and} \quad \rho_{(21)ij} = \rho_{(22)ij}, \quad i > j.$$

In the linearized correlation matrix $\boldsymbol{P}_t^*$ the k elements in matrix $\boldsymbol{P}_{(3)}^*$ corresponding to these restrictions are set to zero, and the vector $\boldsymbol{\rho}^*$ is defined as before but excluding the elements that have been set to zero. Furthermore,

$$\frac{\partial (vec\boldsymbol{P}_t^*)'}{\partial \boldsymbol{\rho}^*}$$

is defined as before, but with the corresponding k rows deleted. The same rows are also deleted from $\boldsymbol{x}_{\rho^* t}$.

When restricting some of the correlation parameters constant according to the transition variable s_{1t} in both extreme states described by variable s_{2t} the test is as defined in the previous subsection with the following modifications: Let there be l pairs of correlations of the form

$$\rho_{(11)ij} = \rho_{(21)ij} \quad \text{and} \quad \rho_{(12)ij} = \rho_{(22)ij}, \quad i > j$$

in both null and alternative hypothesis. In the linearized correlation matrix $\boldsymbol{P}_t^*$ the l elements in matrix $\boldsymbol{P}_{(2)}^*$ corresponding to these restrictions are set to zero, and the vector $\boldsymbol{\rho}^*$ is defined as before but excluding the elements that have been set to zero. Furthermore, when forming

$$\frac{\partial (vec\boldsymbol{P}_t^*)'}{\partial \boldsymbol{\rho}^*}$$

the first $\frac{N(N-1)}{2}$ rows are multiplied by 1 instead of $1 - G_{1t}$, and from the next $\frac{N(N-1)}{2}$ rows, l rows corresponding to the restricted correlations are deleted. The same rows are also deleted from $\boldsymbol{x}_{\rho^* t}$.

These two specifications for partial constancy can be combined, and the test statistic is as defined in the previous subsection with the modifications described above. The asymptotic distribution needs to be adjusted for degrees of freedom to equal the number of restrictions that are tested.

References

Anderson, T. W. (2003): *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 3rd edn.

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79–109.

Bera, A. K., and S. Kim (2002): "Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns," *Journal of Empirical Finance*, 9, 171–195.

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

——— (2005b): "Bond market and stock market integration in Europe," DNB Working Paper No. 60.

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

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

Boyer, B. H., M. S. Gibson, and M. Loretan (1999): "Pitfalls in tests for changes in correlations," Board of Governors of the Federal Reserve System, International Finance Discussion Paper No. 597.

Cappiello, L., R. F. Engle, and K. Sheppard (2003): "Asymmetric dynamics in the correlations of global equity and bond returns," ECB Working paper No. 204.

Chesnay, F., and E. Jondeau (2001): "Does correlation between stock returns really increase during turbulent periods?," *Economic Notes*, 30, 53–80.

de Santis, G., and B. Gerard (1997): "International asset pricing and portfolio diversification with time-varying risk," *Journal of Finance*, 52, 1881–1912.

Doornik, J. A. (2002): *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press, 3rd edn., see also www.doornik.com.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1006.

——— (2002): "Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models," *Journal of Business and Economic Statistics*, 20, 339–350.

Engle, R. F., and K. Sheppard (2001): "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH," NBER Working Paper 8554.

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

Longin, F., and B. Solnik (2001): "Extreme correlation of international equity markets," *Journal of Finance*, 56, 649–676.

Lütkepohl, H. (1996): *Handbook of Matrices*. John Wiley & Sons, Chichester.

Luukkonen, R., P. Saikkonen, and T. Teräsvirta (1988): "Testing linearity against smooth transition autoregressive models," *Biometrika*, 75, 491–499.

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

Pelletier, D. (2006): "Regime switching for dynamic correlations," *Journal of Econometrics*, 131, 445–473.

Silvennoinen, A., and T. Teräsvirta (2005): "Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations," SSE/EFI Working Paper Series in Economics and Finance No. 577.

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

Tse, Y. K., and K. C. Tsui (2002): "A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations," *Journal of Business and Economic Statistics*, 20, 351–362.

White, H. (1982): "Maximum likelihood estimation of misspecified models," *Econometrica*, 50, 1–25.

Numerical aspects of the estimation of multivariate GARCH models

Numerical aspects of the estimation of multivariate GARCH models

Abstract

This paper discusses general issues related to estimation of multivariate GARCH models. The estimation is a numerical procedure and therefore understanding the model at hand and the algorithms used, as well as recognizing potential numerical problems encountered along the way, are of tantamount importance in order to produce reliable model estimates. Special emphasis is put on the computational aspects, such as numerical methods, software programs, use of object oriented programming, and efficiency of computer code.

The author wishes to thank Mika Meitz, Tomoaki Nakatani, Timo Teräsvirta, and Glen Wade for helpful discussions and encouragement. The research was done while the author was visiting the School of Finance and Economics, University of Technology, Sydney, whose kind hospitality is gratefully acknowledged. Very special thanks go to Tony Hall for making the visit possible. The author is responsible for any errors and shortcomings in this paper.

1 Introduction

While there is little predictability in asset returns themselves, the time-varying volatility of an individual asset has direct implications on its riskiness. The autoregressive conditional heteroskedasticity (ARCH) model introduced by Engle (1982) provides a way of capturing volatility behaviour present in financial data. Modelling volatility has ever since been an object of a large body of research resulting in numerous model specifications within a class of generalized ARCH (GARCH) models. For references of developments in the area of univariate volatility modelling, see for instance Bollerslev, Engle, and Nelson (1994), Engle (1995), Palm (1996), Shephard (1996), and Teräsvirta (2007).

Many financial decisions, however, do not depend on the behaviour of a single asset only. Comovement between assets in, say, a portfolio is an important factor when it comes to practical finance. Asset pricing and hedging are examples in which a major role is played by the covariance of the assets in a portfolio. Another area of application is volatility and correlation transmission in studies of contagion. Extending a univariate GARCH model to a multivariate setting has been in focus for many years. Multivariate GARCH (MGARCH) models have problems that do not appear in the univariate context. For example, the number of parameters may increase rapidly with the dimension of the model. Another problem is that ensuring positive definiteness of the conditional covariance matrix can be troublesome. The statistical properties of a general MGARCH model are difficult to establish. There have been several proposals for MGARCH models that try to solve or at least alleviate these problems. For reviews of MGARCH models, see Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2007).

In this paper we discuss problems related to estimation of MGARCH models. The existing literature focuses largely on comparison of ready-to-use packages, see for example Brooks, Burke, and Persand (2003). The study conducted in their paper seems to indicate that these packages can deliver results incompatible with each other. The question of the reliability of the model estimates is left open as different packages provide different results. Without being able to see the implementation of the algorithm used in the package there is no way of knowing the reason behind this outcome. We endeavour to open the 'black box' of the estimation functions provided by the available software programs. To do this requires programming, so we analyze some popular software and consider their support for programming. We also emphasize the need to understand issues related to optimizing algorithms and numerical problems.

The paper is organized as follows. In Section 2 we list important issues related to estimation of MGARCH models. In Section 3 we review the most popular numerical algorithms included in software programs and address numerical problems and possible resolutions. Programming and the relevant features of some of the popular software programs are discussed in Section 4, and Section 5 concludes.

2 Estimation of MGARCH models

In principle, (quasi) maximum likelihood estimation of parametric MGARCH models is straightforward. The objective is to maximize a function that is given in closed form, with respect to a set of parameters. From a mathematical point of view, this is a trivial task, even if the parameter space is constrained by equality or inequality restrictions. In practice, the task often turns out to be computationally highly non-trivial. A major problem with estimation is that the empirical likelihood has several local maxima. Most algorithms used in the numerical optimization of the likelihood function are local optimizers that detect a local maximum but are not designed to look further for alternative solutions. Using global optimizers improve the situation but still cannot guarantee an optimal solution. The problem is exacerbated by the fact that the likelihood surface is often ill-behaved, for example having very flat areas. Another issue is that whenever the parameter space has restrictions, numerical optimizers tend to get stuck on the boundary. This problem arises especially when the core algorithm is meant to solve an unconstrained optimization problem and the user has to impose the restrictions separately. Even algorithms that provide a way of including such restrictions cannot always escape the boundary either. The conditions to ensure positive definiteness of the covariance matrix can in some cases be expressed in terms of parameter restrictions. However, increasing the number of restrictions in the parameter space increases the risk of getting stuck on the boundary. This also complicates the optimization task because the algorithms may have problems in finding search directions that lie within the constrained parameter space.

Estimation is a time-consuming procedure that can, in the first instance, be altered by an appropriate choice of an optimizing algorithm and the set of starting values. Another route is to look, at a technical level, at the computer code and how the estimation procedure is written. Programming skills and knowledge of properties of different software programs are essential in attempts to create efficient and well-written code and to manage modifications and improvements in existing ones. For example, some languages support object oriented programming, which is very useful in data handling and simulations. These issues become even more important when the goal is to produce a program that makes it possible to estimate several MGARCH models, and, furthermore, is fast, reliable, and easy to use, so the code can be shared with and understood by others. Furthermore, these aspects are important in modelling financial time series. Estimation is costly in the MGARCH framework in a sense of time-consumption, because the number of observations is large when compared to many other estimation problems in econometrics, and the structure of the models is highly nonlinear. Careful consideration of numerical methods and software programs can make a difference in obtaining reliable results in a reasonable amount of time.

3 Numerical optimization

The estimators for parametric multivariate GARCH models are generally obtained by maximizing a log-likelihood function. An analytical solution to the optimization

problem, however, is highly intractable and usually impossible to find. Therefore numerical methods are called for. To set notation, let $l(\boldsymbol{\theta})$ be the log-likelihood function that depends on the parameter vector $\boldsymbol{\theta}$ belonging to the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^p$, and let $\hat{\boldsymbol{\theta}}$ be the value that maximizes the likelihood. The numerical optimization procedure is iterative, that is, commencing from a starting value $\boldsymbol{\theta}_1$, the vector of parameters are updated from $\boldsymbol{\theta}_i$ to $\boldsymbol{\theta}_{i+1}$ in each iteration i, often preferably such that $l(\boldsymbol{\theta}_{i+1}) > l(\boldsymbol{\theta}_i)$, until a stopping rule is encountered. This rule is a combination of convergence criteria and a maximum number of iterations allowed. If the iterative process converges, the value from the last iteration $\boldsymbol{\theta}_k$ serves as an approximation to $\hat{\boldsymbol{\theta}}$ whose goodness is controlled by the stopping rule. In case of divergence, the stopping rule ensures that the procedure is only continued up to a certain number of iterations, whereafter which a failure in convergence is reported.

Thus, estimating an MGARCH model of choice requires decisions regarding the starting values, the algorithm, and the convergence criteria. These choices are not independent of each other. The central and active, but also most delicate, part of the estimation procedure is the algorithm that updates the vector of parameters. Therefore one needs to understand the essential build of the algorithm in order to find out how the choices of starting values and stopping rules can affect the results. The issues here are, for instance, the sensitivity to misleading or non-trivial likelihood contours (such as local maxima, ridges, or nearly flat surfaces), the speed of convergence, and whether the convergence criteria should be placed on the parameter values themselves, on derivatives of the likelihood, or on search directions.

3.1 Numerical optimization – methods

Gradient methods

Generally, the updating formula for the parameters is of the form

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + s_i \boldsymbol{d}_i \tag{1}$$

where $s_i \in \mathbb{R}^+$ is step size and $\boldsymbol{d}_i \in \mathbb{R}^p$ determines the search direction. First-order Taylor expansion of $l(\boldsymbol{\theta}_i + \boldsymbol{d}_i)$ yields the linear approximation $l(\boldsymbol{\theta}_i) + \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'} \boldsymbol{d}_i$ and hence the direction of search that increases the value of the log-likelihood can be characterized as any vector $\boldsymbol{d}_i$ that satisfies $\frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'} \boldsymbol{d}_i > 0$. Setting $\boldsymbol{d}_i = \boldsymbol{Q}_i \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}}$ where $\boldsymbol{Q}_i$ is any symmetric, positive definite $p \times p$ matrix trivially satisfies the condition. The updating formula (1) thus becomes

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + s_i \boldsymbol{Q}_i \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}} \tag{2}$$

and it suffices to find a suitable step size to follow an increasing direction when maximizing the objective function. The algorithms following (2) are called gradient methods and they differ mainly on the choice of $\boldsymbol{Q}_i$. The scaling factor s_i is used to increase the speed of convergence but also to ensure that the iterative search converges

to a maximum. For instance, the step size can be chosen by univariate line search procedure by maximizing $l(\boldsymbol{\theta}_i + s\boldsymbol{Q}_i \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}})$ with respect to s.

The Newton-Raphson algorithm, see e.g. Gill, Murray, and Wright (1981), is a direct application of the Newton's method to maximum likelihood estimation. The second-order Taylor expansion of $l(\boldsymbol{\theta}_i + \boldsymbol{d}_i)$ gives the quadratic approximation $l(\boldsymbol{\theta}_i) + \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'}\boldsymbol{d}_i + \frac{1}{2}\boldsymbol{d}_i' \frac{\partial^2 l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\boldsymbol{d}_i$. The optimal direction must satisfy first-order conditions, that is, $\frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'} + \boldsymbol{d}_i' \frac{\partial^2 l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \boldsymbol{0}$, which yields $\boldsymbol{d}_i = -(\frac{\partial^2 l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'})^{-1} \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}}$. The iteration formula becomes

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - s_i(\sum_{t=1}^{T} \frac{\partial^2 l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'})^{-1} \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}}.$$

Whenever the log-likelihood function is concave and the Hessian negative definite in the area around the maximum and the starting values are not too far from the true optimum, the Newton-Raphson algorithm in principle works well. As $\boldsymbol{\theta}_i \to \hat{\boldsymbol{\theta}}$, $-\frac{\partial^2 l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$ converges to $-\frac{\partial^2 l(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}$, which is, up to a scalar, an estimator of the information matrix.

The BHHH algorithm of Berndt, Hall, Hall, and Hausman (1974) makes use of the equivalent representation of the information matrix by using another (up to a scalar factor) estimator $\sum_{t=1}^{T} \frac{\partial l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'}$. The iteration scheme in this case is the following:

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + s_i(\sum_{t=1}^{T} \frac{\partial l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}} \frac{\partial l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'})^{-1} \frac{\partial l(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}}.$$

One advantage of this approach compared to the Newton-Raphson algorithm is that the matrix premultiplying the score vector is always positive definite. Another is that it requires the use of first-order derivatives only. If one wants to use analytical forms of derivatives, the expressions for second-order derivatives can be quite tedious to work out. However, derivatives can also be computed numerically. From the theoretical point of view, differences in the performance of the Newton-Raphson and BHHH algorithms can be accredited to information matrix equality failing to hold. This can be an indication of the model misspecification.

The Newton-Raphson and BHHH algorithms may be costly as the dimension of the multivariate problem increases due to the large number of matrix inversions they require. Quasi-Newton methods replace the inverted Hessian by approximations that evolve with the iterations. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm follows (2) with

$$\begin{aligned} \boldsymbol{Q}_i &= \boldsymbol{Q}_{i-1} + (1 + \frac{\boldsymbol{q}_i'\boldsymbol{Q}_{i-1}\boldsymbol{q}_i}{\boldsymbol{q}_i'(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})})\frac{(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})'}{\boldsymbol{q}_i'(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})} \\ &\quad - \frac{(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})\boldsymbol{q}_i'\boldsymbol{Q}_{i-1} + \boldsymbol{Q}_{i-1}\boldsymbol{q}_i(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})'}{\boldsymbol{q}_i'(\boldsymbol{\theta}_i - \boldsymbol{\theta}_{i-1})} \end{aligned}$$

where $\boldsymbol{q}_i = \frac{\partial l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}} - \frac{\partial l_t(\boldsymbol{\theta}_{i-1})}{\partial \boldsymbol{\theta}}$; see for instance Gill, Murray, and Wright (1981). This procedure relies on the assumption that the objective function can be approximated by

a quadratic function around the maximum. It is faster than the previous algorithms and only uses the first-order derivatives.

Thus far the methods discussed deal with unconstrained optimization. MGARCH models often involve parameter restrictions to ensure existence of moments, stationarity, positivity of certain parameters, and positive definiteness of the conditional covariance matrix. Although sometimes it is possible to transform the parameters in the model in such a way that incorporates the information in the constraints, generally that is either inconvenient, burdensome, or even impossible to implement. Nonlinear programming subject to equality or inequality constraints is therefore useful. Sequential quadratic programming (SQP) is designed to solve a nonlinear optimization problem subject to nonlinear constraints, stated for instance as

$$\max l(\boldsymbol{\theta}) \quad \text{s.t.} \quad g(\boldsymbol{\theta}) > 0$$

where both $l(\boldsymbol{\theta})$ and $g(\boldsymbol{\theta})$ are nonlinear. For a review on SQP algorithms and theory, see Boggs and Tolle (1995). In its basic form SQP solves the problem by using a sequence of quadratic programming approximations obtained by replacing the nonlinear constraints by first-order Taylor approximations and the nonlinear objective function by second-order Taylor approximation that is augmented by the second order information from the constraints. More specifically, let $\mathcal{L}(\boldsymbol{\theta}_i, \boldsymbol{\lambda}_i) = l(\boldsymbol{\theta}_i) - \boldsymbol{\lambda}_i' g(\boldsymbol{\theta}_i)$ be the associated Lagrangian with nonnegative vector of Lagrange multipliers $\boldsymbol{\lambda}_i$. The search direction $\boldsymbol{d}_i$ then solves iteratively the quadratic problem with linear constraints:

$$\min\{\frac{\partial l_t(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'}\boldsymbol{d}_i + \frac{1}{2}\boldsymbol{d}_i'\boldsymbol{H}_i\boldsymbol{d}_i\} \quad \text{s.t.} \quad g(\boldsymbol{\theta}_i) + \frac{\partial g(\boldsymbol{\theta}_i)}{\partial \boldsymbol{\theta}'}\boldsymbol{d}_i > 0$$

where $\boldsymbol{H}_i$ approximates the Hessian of the Lagrangian function $\mathcal{L}(\boldsymbol{\theta}_i, \boldsymbol{\lambda}_i)$. Once the optimal search direction is found, the Lagrangian is optimized, resulting in the next iterate $\boldsymbol{\theta}_{i+1}$, and the Hessian is updated using, for instance, the BFGS algorithm. After this a new iteration is commenced to find a new search direction. Note however, that the new direction of search may not be feasible, and several proposals have been made on how to tilt the direction into a feasible set. Recently, Lawrence and Tits (2001) proposed a feasible sequential quadratic programming (SQPF) algorithm. The main advantage in their approach is that it reduces the amount of computations per iterate when searching for the new search direction. This ensures faster convergence which is a major issue in nonlinear optimization.

Simulated annealing

All of the algorithms discussed above implicitly rely on some underlying presumptions on the objective function, such as approximately quadratic behaviour near optimum, continuity, or first- or second-order differentiability. Furthermore, all of them impose the condition $l(\boldsymbol{\theta}_{i+1}) > l(\boldsymbol{\theta}_i)$ during the course of iterations. Even in cases where asymptotically likelihood functions have a single global maximum, sample behaviour is quite different and likelihood functions often exhibit multiple local maxima. Updating the parameter vector in such a way that increases the likelihood at every step

will, unless getting stuck on the edge of the parameter space, converge to a nearby maximum. These algorithms are so called local optimizers, and their major drawback is indeed that they only find the local maximum, which is usually nearest the starting point of the algorithm. Commencing the algorithm from a range of different starting values may lead to different optima which can be then compared in terms of the function values at those points. When using local optimizers one should consider fairly extensive array of starting values in order to increase the probability that the chosen optimal point is at least close to the global maximum. The global optimization algorithms which evidently try to escape local maxima include, for instance, adaptive random search method, genetic algorithms and simulated annealing. Simulated annealing has proven to be the most reliable one in systems with a large number of variables and robust to likelihood surfaces that exhibit several maxima or other non-trivial surface shapes. The simulated annealing algorithm of Corana, Marchesi, Martini, and Ridella (1987), modified by Goffe, Ferrier, and Rogers (1994) and Brooks and Morgan (1995), relates to a technique used in thermodynamics to slowly cool metal to a global low energy state. Using a large step size the algorithm first explores the functions on the entire surface, moving both up- and downhill, and employing random moves in the directions of each of the parameters in turn. After finding the most promising area for the global maximum the algorithm refines the search, still allowing for steps that may decrease the likelihood value, which helps to avoid local maxima. For an overview of the theoretical aspects, implementations, and practical behaviour, see Brooks and Morgan (1995).

The algorithm starts at an initial temperature $T = T_0$, the vector of step-lengths $\boldsymbol{s} = \boldsymbol{s}_0$, and the parameter values $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The initial optimum point is recorded as $\boldsymbol{\theta}_{\text{opt}} = \boldsymbol{\theta}_0$ with optimal value $l_{\text{opt}} = l(\boldsymbol{\theta}_0)$. A new candidate point $\boldsymbol{\theta}'$ is chosen by varying one the elements of $\boldsymbol{\theta}$ at a time:

$$\theta'_j = \theta_j + r s_j$$

where θ_j and s_j are the jth elements of $\boldsymbol{\theta}$ and $\boldsymbol{s}$, respectively, and, for each j, r is randomly drawn from $U[-1, 1]$. The new point is accepted if $l(\boldsymbol{\theta}') > l(\boldsymbol{\theta})$ and stored as a new optimum if $l(\boldsymbol{\theta}') > l_{\text{opt}}$. Otherwise the decision is made based on the so-called Metropolis criterion: the point is accepted if

$$e^{\frac{l(\boldsymbol{\theta}') - l(\boldsymbol{\theta})}{T}} > u$$

where u is drawn from $U[0, 1]$. This implies that, for a fixed temperature T, the smaller the decrement in the likelihood value is, the larger the probability of accepting such a move. After trying a new point in the direction of each of the parameters, the procedure is started over and repeated N^s times. After those N^s steps the vector of step lengths $\boldsymbol{s}$ is adjusted in each of its elements such that about half of all moves are accepted. This ensures that if only a small portion of moves are accepted in the direction of parameter θ_j, the corresponding step length is decreased, which increases the probability of moves being accepted, and vice versa. The procedure is started over again, and after the loops described above are carried through N^T times, the temperature is reduced to vT, where $v \in (0, 1)$ is a fixed proportion. Choosing v

close to one makes the temperature reduction process slow and hence increases the probability of escaping local maxima.

The algorithm is restarted from the current optimum point, but now the reduced temperature will decrease the number of accepted points according to the Metropolis criterion and, consequently, in subsequent loops the step-lengths will become smaller, which refines the search. At the time of each temperature reduction the latest accepted point is stored thus creating a succession of vectors $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots$. Note that to reach such a terminal point $p \times N^s \times N^T$ function evaluations are needed. Finally, after sufficiently many iterations, the latest N^ε recorded terminal points are compared with the current optimal one. The algorithm is stopped once the values of the likelihood function evaluated at these points are within a distance of size ε apart, $\varepsilon > 0$ sufficiently small. This condition is set to increase probability that the found optimum is reasonably close to the global one. The quantities N^s, N^T, v, as well as the stopping rule ε, are all user defined; some guidelines how to choose them, as well as detailed description of the procedure can be found in Corana, Marchesi, Martini, and Ridella (1987).

An advantage of the simulated annealing algorithm is that it does not place any prerequisites on the shape of the objective function, nor on its differentiability. In fact, the function does not even need to be defined at all points. When the algorithm chooses a new candidate point $\boldsymbol{\theta}'$, a new random number r is drawn until the point is within the domain of the function. Therefore the simulated annealing algorithm can explore areas that would be impossible for other methods. The step size gives more information about the likelihood surface than a gradient at a given point does; the larger the step size in the direction of a certain parameter, the flatter the likelihood with respect to that parameter. Generally, simulated annealing is a procedure that can deal with functions that would be impossible to optimize using local optimizers. A drawback, however, is that, because finding an optimum is a slow process, the simulated annealing algorithm requires a vast amount of computation time to converge. Therefore, it may not be the first choice of algorithm one wants to try on an estimation problem unless, of course, the objective function is such that none of the local optimizers can be used. Initial checks of convergence in general, as well as sensitivity to starting values, can be carried out using faster algorithms, and if there are problems one can apply simulated annealing. However, due to ever increasing computing power, the simulated annealing algorithm can become a very appealing alternative to derivative-based methods in the near future.

3.2 Numerical optimization – problems

The gradient methods as well as the SQP methods require the knowledge of the first, and sometimes second, derivatives of the likelihood function. While finding analytic expressions for the derivatives can be a tedious task, once provided they substantially reduce the computation time. Gable, van Norden, and Vigfusson (1997) pointed out that the calculation of a gradient vector of size $p \times 1$ typically requires $p+1$ and the matrix of second derivatives p^2+1 likelihood evaluations. However, a slight modification of the model to be estimated will unavoidably alter the derivatives, which should then

be recalculated and reprogrammed. Therefore, when searching for a model that fits the data well, it may be argued that even at a cost of longer estimation time, numerical derivatives are a practical choice. However, in simulation experiments analytical ones prove very useful as they may save weeks or even months of computation time. In practice, when a carefully written estimation algorithm is used, the analytical and numerical derivatives yield rather similar parameter and standard error estimates. Use of numerical derivatives tend to slightly increase the standard errors. There are analytical derivatives available for some specific models, see for instance Lucchetti (2002) for first-order derivatives for the BEKK model, Hafner and Herwartz (2005) for first- and second order derivatives for the VEC and BEKK models, as well as CCC– and DCC–GARCH models, Silvennoinen and Teräsvirta (2005, 2006) for first-order derivatives for the STCC– and the DSTCC–GARCH models, and Nakatani and Teräsvirta (2006) for first- and second-order derivatives of the ECCC–GARCH model.[1]

In the previous section we discussed the problem of finding a local as opposed to the global maximum. Another issue related to the shape of the likelihood is that the surface can be flat in one or several directions of parameters.[2] That is, even a large change in the value of a parameter that is large in proportion to the scale of the other parameters may have a negligible effect on the value of the likelihood. The maximizing algorithm may then get stuck in a loop, searching through the same points over and over again, or push the parameter value in question towards positive or negative infinity. To alleviate this problem, one may rescale the problematic parameters such that the rescaled values are close to the values of the other parameters. In that way the numerical optimizer can be forced to take smaller steps and possibly find the optimal point. If, however, this does not bring results, one may consider fixing these parameters to specific values and estimating the remaining ones conditionally on these values. In that case a reasonably fine grid of fixed values should be considered and optimization carried out for each one of them in order to obtain a reasonably close approximation to the optimal solution.

Suitable rescaling of the parameters may have additional advantages. The differences in scale can lead to close to singular matrices, which the numerical optimizer then fails to invert. This is the case in particular when calculating standard errors of the parameters. The problem can often be avoided if the parameters have roughly the same numerical size. Instead of rescaling some of the parameters, one may try to rescale the likelihood function. This may become useful for instance when the likelihood surface appears flat in all directions. Suitable rescaling can enforce the contours of the surface of the likelihood and thereby facilitate the work of the numerical optimizer.

As discussed in the previous section, local optimizers can be especially sensitive to the starting values. A well-planned grid of initial values can benefit the overall estimation result by providing many individual results for comparison. Using a grid of starting values in the estimation routine can in principle be done automat-

[1] For model definitions and references, see Silvennoinen and Teräsvirta (2007).

[2] For related discussion in context of models involving smooth transitions, see van Dijk, Teräsvirta, and Franses (2002).

ically by programming such a grid into the algorithm. However, as the number of parameters increases, the number of reasonable starting points explodes and the estimation quickly becomes infeasible. It is therefore crucial to understand the role of different parameters in the model at hand and use that information as a guideline to first construct a crude grid. The estimation results will then hopefully reveal a subset of the parameter space in which the grid should be refined and to which the search is focused. It may be argued that such techniques are redundant and that global optimizers should always be used, particularly where high-powered computers are available. However it should be noted that global optimizers do not always find the global solution and may be prone to the same issues as the local ones. Global optimizers are designed to escape local maxima and focus on the most promising area for finding a global optimum. However, the resulting approximation of the optimal point may be numerically poor. Therefore global optimizers can be used for the same purpose as the grid: to produce starting values for local, derivative-based algorithms which can then be used for the final refinement of the parameter estimates.

As already mentioned, MGARCH models often involve parameter restrictions. Many algorithms are designed for unconstrained optimization and using such methods means that the restrictions have to be forced into the estimation routine from outside of the optimizer. This causes the estimation algorithm to work in quite an inefficient way. The objective function is usually not behaving nicely close to the boundaries of the parameter space, because those restrictions are set to ensure, say, the existence of certain moments. Once getting close to the boundary, the likelihood function may behave badly. For instance, the derivatives can give the impression that the optimal solution is even outside the parameter space. This is one reason why an algorithm can get stuck on the boundary. Some built-in algorithms may have a way of identifying the problem of one or more parameters reaching the boundaries of the constrained space. In that case they are likely to contain some 'standard' procedure for getting away from the boundary. Whether these features are present and how they actually are implemented in the estimation program may not, however, be clear from the documentation of the maximizing function, which emphasizes the importance of an open source code. The methods designed for constrained optimization should have the restrictions well implemented. However, it may become necessary to modify even those implementations in order to be able to break away from the boundary. Each MGARCH model can have its own set of parameters that are more likely to remain on the boundary and the reasons for this can be very different from one model to another. For this reason it is not advisable to treat all models alike. One should try to understand why estimates of certain parameters of a model tend to move towards the edge of the parameter space and what type of a precautionary strategy would be likely to eschew the problem in that particular model.

The conditional covariance matrix of an MGARCH model should be positive definite at each point in time. Some models have this property by construction, whereas in some models the requirement of positive definiteness causes restrictions in some part of the model. Ensuring positive definiteness can sometimes be expressed as parameter restrictions. As discussed before, restricting the parameter space increases the risk of getting stuck on the boundary, and therefore it often is more convenient to

implement a numerical check instead. Requiring the conditional covariance matrix to be positive definite at each step during the iterations leads to a positive definite covariance matrix of the estimated model. However, checking the positive definiteness numerically is computationally demanding as it involves procedures, for instance computing the Cholesky decomposition of a potentially large matrix, that are time-consuming. It is not necessary, however, to require positive definiteness throughout the estimation, although one has ensure non-singularity of the matrices that have to be inverted. Once the model is estimated, one has to check that positive definiteness is not violated before accepting the parameter estimates. Whether this approach works in practice depends on the model and can only be found out by trial and error.

4 Programming

There are a number of software programs available for econometric modelling. Many of these can be enhanced by writing new functions. Eviews, RATS, and SAS are examples of such software. Brooks (1997) compared properties of some of the packages in a Monte Carlo study where the standard univariate GARCH(1,1) model was estimated using the BHHH algorithm. The comparison focused on the availability of pre-programmed functionality, and user-friendliness was described as being inversely related to the number of lines of computer code required for running an estimation routine. A similar comparison in MGARCH context is reported in Brooks, Burke, and Persand (2003). The study conducted in their paper requires some programming skills but still the focus is on the availability of pre-programmed packages. However, one may find statistical or mathematical packages not flexible enough to let the user deal with issues discussed in the previous section, and their modification may be difficult, if not impossible. Although many of those built-in procedures may be helpful, the estimation of MGARCH models may require fine-tuning that forces one to program the estimation routine from scratch. The user-friendliness in those situations comes from the ease of modifying the existing code, using custom-written functions, or even making use of procedures written in other languages. Another important consideration is the number of useful base functions and the quality of the documentation provided. Most of the languages have discussion boards and user mailing lists that provide support on programming related problems and also help users to identify bugs and suggest improvements.

4.1 Aspects to consider when choosing software

There are some basic criteria that should be considered when choosing a software program. Many of the programs providing statistical computing and graphics also provide packages for estimation. Four popular and flexible matrix oriented languages we consider are Matlab (www.mathworks.com), GAUSS (www.aptech.com), Ox (www.doornik.com), and R (www.r-project.org). 'Matrix-oriented' means that special native support for matrix operations is available, which naturally is a benefit

when it comes to modelling multivariate time series. Matlab has a highly developed graphical user interface, which makes it easier to use than the other three. The others require the user to have basic programming skills. Such skills will be required by any user who wishes to extend the basic functionality provided with the software program. All of them come with a comprehensive mathematical and statistical function library of pre-programmed functions. Only Ox and R support the use of object-oriented programming. This provides speed and well-designed syntax, which leads to programs that are easy to maintain. There is not much difference between the programs when it comes to the speed of the native functions, e.g. matrix inversion. Matlab, Ox, and R are compatible with C/C++ and Fortran, in that they can use existing programs written in those languages. Ox and R are also compatible with GAUSS. Matlab and GAUSS only have commercial versions available. GAUSS, Ox, and R provide console versions with limited graphical features. For Ox this version is free for academic use, whereas the advantage of R is that, not depending on the user group, the software program is free. The professional version of Ox offers enhanced functionality and a richer user interface, but is not free.

Matlab, GAUSS, Ox, and R provide open source code for many of their functions, which makes modification and enhancement of functions easy. However, the selection of open source functions in Matlab, GAUSS, and OxProfessional is limited by copyright restrictions and modifications may not always be permitted. Open source code also encourages collaboration between users on web forums. There are numerous functions available on the internet in the form of either accessible or compiled C-code. Because of the compatibility with the C-language, one can supplement the original function library provided by the software program, with new, improved, and fast alternatives. Specific parts of the estimation procedures for MGARCH models may need to be adjusted or refined to work more efficiently in different models. Access to the source code of certain functions may help to understand, for instance, why an optimizing algorithm fails in some cases and how it needs to be improved. All these programs have a range of numerical maximizers to choose from, such as Newton-Raphson, BFGS, and SQPF algorithms that include the option of using either analytical (provided by the user) or numerical derivatives. Simulated annealing is also available.[3]

These software programs provide a selection of ready-to-use univariate GARCH estimation routines. Very limited support is provided for estimating MGARCH models.[4] Although Brooks, Burke, and Persand (2003) claim that implementing multivariate extensions to existing univariate packages 'would not be a difficult exercise', we would like to point out that even a simple model such as the BEKK model, or even the diagonal version of it, can turn out to be almost impossible to estimate reliably, not to mention the problems that arise when the code is extended to cover the VEC model. Even though the two models are mathematically quite closely related, any 'trivial' modification of a code for estimating BEKK is likely to result in nonsensical VEC estimates. Interestingly enough, as reported in Brooks, Burke, and

[3]See also Lester Ingber's homepage, http://www.ingber.com/.

[4]Sébastien Laurent, Jeroen Rombouts, Annastiina Silvennoinen, and Francesco Violante are currently working on producing an estimation package for MGARCH models for Ox.

Persand (2003), the parameter estimates from the diagonal VEC model using various ready-to-use packages differ remarkably from each other.

4.2 Object-oriented programming

Object Oriented Programming (OOP) is a programming paradigm designed to address some of the limitations encountered in procedural and event-based programming models. An object is a self-contained building block. Objects, by definition, contain properties, methods and events. Properties are used for storing data, e.g. a name or a value. Methods define the actions an object can do, i.e. functions. Events provide support for non-sequential programming. All these entities are written into a program known as a class. This class is now the template for creating objects of this type. Programmers will create many classes, or building blocks, and then use them together with the pre-programmed ones to achieve their goals. Ox and R provide pre-programmed classes for data management and simulation, which makes these complex tasks much simpler than they otherwise would be. For instance, when investigating properties of, say, a model, tests, or an estimation algorithm, running simulations by a proper use of objects can make the job relatively easy.

Two core concepts associated with OOP are encapsulation and inheritance. Encapsulation means that the purpose of the object is defined and its boundaries are fixed. Inheritance provides an elegant mechanism to enhance and modify the functionality of an existing object. Together these concepts, when applied intelligently, go a long way towards providing simple code re-usability. Another advantage of OOP is that objects are inherently self-documenting, making the code easier to read and sparing programmers from having to provide a lot of documentation when they want to share their code. The object-oriented features in Ox and R are not as sophisticated as the ones of some high-level languages. This avoids the complexity of a language such as C++, while still providing most of the benefits. Both languages can, however, be used without taking advantage of the OOP features, thereby providing functionality similar to that of Matlab and GAUSS.

4.3 Observations on programming

All software programs require a fair amount of work to learn how the language works and how to make most efficient use of it. Given the similarity of their basic features and the large initial effort, the choice between them should be based on the potential user group, possibilities of sharing parts of the code with colleagues, support, potential benefits from using objects and classes, etc.

The speed of an estimation routine is a function of the efficiency of the computer code and the computer it is running on. Estimation routines can be both processor and memory intensive operations. The matrix operations will use a lot of the CPU time. Working with large dimensional matrices will require a lot of computer memory. In order to write efficient code the programmer should take these things into account.

Generally one should avoid the unnecessary use of loops, for example, by exploiting matrix operations. R is particularly inefficient in executing loops. Where loops are inavoidable and the performance is critical, R users may be well served by writing these sections in a language like C. The code can then be compiled and called from within the R program. If looping is required to, say, perform a memory intensive operation for all the elements in a time series, the programmer must be diligent in the use of memory. In large-dimensional models the memory may need to be released between each iteration of the loop. Failing to do this can result in the computer's memory being used up, which in turn will cause the program to run increasingly slowly and ultimately to crash. Good practice is to release the memory used by variables and matrices as soon as they are no longer needed. Where processing power is the constraining factor, it may sometimes pay off to save certain large matrices in order to avoid the overhead of recalculating them in every iteration of a loop. Also, one should use native functions where possible as they generally are efficient both from the speed and memory-saving point of view. When running simulations code execution speed is critical. To this end, it is necessary to have good knowledge of the properties of the software and how its features can be exploited in a most efficient way to obtain a speed-optimized computer code.

As a general guideline, the largest effort should be directed towards improving the efficiency of the innermost loops in the program. Numerical optimization is itself an iterative process in which the whole parameter vector is updated in each iteration. When the number of parameters increases, dealing with all of them simultaneously can create numerical problems for the algorithm. This can be avoided by splitting the parameter vector into subvectors, each of which is updated in turn while keeping the other ones fixed to their previously updated values. The division of the parameters into subsets should be based on the role or scale of the parameters. For instance, in dynamic conditional correlation MGARCH models one group could contain the parameters involved in modelling the conditional volatilities whereas the other one would consist of the correlation parameters. The downside of such a division is that the number of iterations needed for convergence may grow. However, the procedure increases the efficiency of the estimation.

5 Concluding remarks

In this paper we draw attention to aspects in the estimation of MGARCH models that thus far have not been directly addressed. Parametric (quasi) maximum likelihood estimation of MGARCH models can be regarded as a straightforward task. There are several numerical algorithms designed for tackling such problems. However, due to the nonlinear nature of the model as well as the potentially large number of parameters, the optimization can turn out to be non-trivial. A key problem with the maximization of the likelihood function is that it often has several local maxima. A direct application of any numerical optimizer can lead to unreliable model estimates. Therefore it is crucial to understand the properties of the model to be estimated as well as the properties of the algorithm used in the estimation procedure.

Because the problems in the estimation of MGARCH models are not only theoretical but also depend on the actual implementation of the estimation routine, we also consider aspects that should be given attention when choosing a software program. Due to numerical difficulties in the estimation of MGARCH models, the modeller should be prepared to learn to program new procedures or modify existing ones. The computer code should also be both memory and speed efficient. An important aspect to consider before studying programming is the possibility to extend the functionality of a software program. Object oriented programming (OOP) is a relatively new tool in the estimation of econometric models, although the concept has been in use for decades in commercial software development. The benefits of OOP are already visible in the reduced effort required to carry out simulations and manage data, but there is potential for much more.

References

Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006): "Multivariate GARCH models: A survey," *Journal of Applied Econometrics*, 21, 79–109.

Berndt, E. K., B. H. Hall, L. R. E. Hall, and J. A. Hausman (1974): "Estimation and inference in nonlinear structural models," *Annals of Economic and Social Measurement*, 3, 653–666.

Boggs, P. T., and J. W. Tolle (1995): "Sequential quadratic programming," *Acta Numerica*, 4, 1–51.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994): "ARCH models," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2959–3038. Elsevier Science, Amsterdam.

Brooks, C. (1997): "GARCH modelling in finance: a review of the software options," *Economics Journal*, 107, 1271–1276.

Brooks, C., S. P. Burke, and G. Persand (2003): "Multivariate GARCH models: software choice and estimation issues," *Journal of Applied Econometrics*, 18, 725–734.

Brooks, S. P., and B. J. T. Morgan (1995): "Optimization using simulated annealing," *The Statistician*, 44, 241–257.

Corana, A., M. Marchesi, C. Martini, and S. Ridella (1987): "Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm," *ACM Transactions on Mathematical Software*, 13, 262–280.

Engle, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987–1006.

——— (ed.) (1995): *ARCH: Selected Readings*. Oxford University Press.

Gable, J., S. van Norden, and R. Vigfusson (1997): "Analytical derivatives for Markov switching models," *Computational Economics*, 10, 187–194.

Gill, P. E., W. Murray, and M. H. Wright (1981): *Practical Optimization*. Academic Press, New York.

Goffe, W. L., G. D. Ferrier, and J. Rogers (1994): "Global optimization of statistical functions with simulated annealing," *Journal of Econometrics*, 60, 65–99.

Hafner, C. M., and H. Herwartz (2005): "Analytical quasi maximum likelihood inference in multivariate volatility models," Unpublished manuscript.

Lawrence, C. T., and A. Tits (2001): "A computationally efficient feasible sequential quadratic programming algorithm," *SIAM Journal on Optimization*, 11, 1092–1118.

Lucchetti, R. (2002): "Analytical score for multivariate GARCH models," *Computational Economics*, 19, 133–143.

Nakatani, T., and T. Teräsvirta (2006): "Testing volatility interactions in a constant conditional correlation GARCH model," Unpublished manuscript.

Palm, F. C. (1996): "GARCH models of volatility," in *Handbook of Statistics*, ed. by G. S. Maddala, and C. R. Rao, vol. 14, pp. 209–240. Amsterdam: Elsevier Sciences.

Shephard, N. (1996): "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Finance and Other Fields*, ed. by D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielsen, pp. 1–67. Chapman and Hall, London.

Silvennoinen, A., and T. Teräsvirta (2005): "Multivariate autoregressive conditional heteroskedasticity with smooth transitions in conditional correlations," SSE/EFI Working Paper Series in Economics and Finance No. 577.

——— (2006): "Modelling multivariate autoregressive conditional heteroskedasticity with the double smooth transition conditional correlation GARCH model," Unpublished manuscript.

——— (2007): "Multivariate GARCH models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch. Springer, New York.

Teräsvirta, T. (2007): "Univariate GARCH models," in *Handbook of Financial Time Series*, ed. by T. G. Andersen, R. A. Davis, J.-P. Kreiss, and T. Mikosch. Springer, New York.

van Dijk, D., T. Teräsvirta, and P. H. Franses (2002): "Smooth transition autoregressive models – a survey of recent developments," *Econometric Reviews*, 21, 1–47.

EFI, The Economic Research Institute

Published in the language indicated by the title. A complete publication list can be found at www.hhs.se/efi/. Books and dissertations can be ordered from EFI via e-mail: EFI.Publications@hhs.se.

Reports since 2002

2006

Books

Thodenius, Björn. *Organisering av kunskap. En studie av Wallenberg Consortium North.*

Wijkström, Filip och Torbjörn Einarsson. *Från nationalstat till näringsliv? Det civila samhällets organisationsliv i förändring.*

Östman, Lars. *Lysande ögonblick och finansiella kriser. Dramaten under ett sekel.*

Dissertations

Argenton, Cedric. *Quality provision in duopoly.*

Beckerman, Carina. *The Clinical Eye. Constructiong and Computerizing an Anesthesia Patient Record.*

Borglund, Tommy. *Aktievärden i fokus - internationell påverkan på intressentrelationer vid förvärv och fusion.*

Breman, Anna. *The Economics of altruism, paternalism and self-control.*

Edquist, Harald. *Technological breakthroughs and productivity growth.*

Frostenson, Magnus. *Legitimitetskontrollen - en studie av etiska värderingars roll i gränsöverskridande förvärv och fusioner.*

Gaspar, Raquel M. *Credit Risk and Forward Price Models.*

Hopkins, Elisabeth. *Is a higher degree of local currency pricing associated with lower exchange rate pass-through? A study of import pricing in 51 Swedish industries.*

Kling, Ragnar. *Developing Product Development in Times of Brutal Change.*

Langenskiöld, Sophie. *Peer Influence on Smoking: Causation or Correlation?*

Lychnell, Lars-Olof. *IT-relaterat förändringsarbete I informationsintensiva småföretag - en klinisk fallstudie från resebranschen.*

Meitz, Mika. *Five Contributions to Econometric Theory and the Econometrics of Ultra-High-Frequency Data.*

Mendicino, Caterina. *Financial Market Imperfections, Business Cycle Fluctuations and Economic Growth.*

Ovanfors, Anna. *Essays on Nonlinear Time Series Analysis and Health Economics.*

Rogberg, Martin. *Den modeföljande organisationen. Om acceptansen av TQM och andra populära managementmodeller.*

Slinko, Irina. *Essays in Option Pricing and Interest Rate Models.*

Wilander, Fredrik. *Essays on Exchange Rates and Prices.*

2005

Books

Andersson, Per, Susanne Hertz and Susanne Sweet (eds). *Perspectives on market networks - boundaries and new connections.*

Charpentier, Claes. IT inom omsorgen. *Förväntade effekter av införande av IT-system för utförarna inom äldre- och handikappomsorgen.*

Dembrower, Maria. *Entreprenörskap i industriella nätverk.*

Lind, Johnny och Göran Nilsson (redaktörer).*Ekonomistyrningens metoder, sammanhang och utveckling - En vänbok till Lars A Samuelson.*

Samuelson, Lars A. *Organizational governance and control - a summary of research in the Swedish society.*

Svedberg Nilsson, Karin, Roger Henning och Karin Fernler (eds). *En illusion av frihet? Företag och organisationer i regelsamhället. EFIs Årsbok 2005. EFI/Studentlitteratur.*

Dissertations

Andersson, Martin. *Making a Difference - Project Result Improvement in Organizations.*

Arvidsson, Per. *Styrning med belöningssystem - Två fallstudier om effekter av belöningssystem som styrmedel.*

Bems, Rudolfs. *Essays in International Macroeconomics.*

Berg-Suurwee, Ulrika. *Nya trender, nya nmnder - effekter av en stadsdelsnmndsreform inom kultur och fritid.*

Björkman, Hans. *Learning from members - Tools for strategic positioning and service innovation in trade unions.*

Bodnaruk, Andriy. *Essays in Empirical Corporate Finance and Portfolio Choice.*

Clapham, Eric. *Essays in Real Estate Finance.*

Dareblom, Jeanette. *Prat, politik och praktik - Om individers möten med strukturer i en kommunal satsning på kvinnors företagande.*

Fromm, Jana. *Risk Denial and Neglect: Studies in Risk Perception.*

Hjelström, Anja. *Understanding International Accounting Standard Setting - A Case Study of IAS 12, Accounting for Deferred Tax.*

Hortlund, Per. *Studies on Swedish Banking 1870-2001.*

Lindahl, Therese. *Strategic and Environmental Uncertainty in Social Dilemmas.*

Linnarsson, Håkan. *Alliance for Innovation. A structural perspective on new business development in cooperative ventures.*

Madestam, Andreas. *Developing Credit Markets.*

Nilsson, Roland. *The Market Impact of Short-Sale Constraints.*

Nordfält, Jens. *Is consumer decision-making out of control? Non-conscious influences on the consumer decision-making process for fast moving consumer goods.*

Nordin, Fredrik. *Externalising Services - Walking a Tightrope between Industrial and Service Logics.*

Parmler, Johan. *Essays in Empirical Asset Pricing.*

Simbanegavi, Witness. *Price Discrimination, Advertising and Competition.*

Thodenius, Björn. *Användning av ledningsinformationssystem: en longitudinell studie av svenska storföretag.*

Tolis, Christofer. *Framing the Business - Business Modelling for Business Development.*
Östberg, Per. *Corporate Disclosure and Investor Recognition.*

2004
Books
Ahrne, Göran och Nils Brunsson (red). *Regelexplosionen.*
Lind, Johnny. Strategi och ekonomistyrning. *En studie av sambanden mellan koncernstrategi, affärsstrategi och ekonomistyrning.*
Lind, Johnny och Walter Schuster (red). *Redovisningens teori, praktik och pedagogik. En vänbok till Lars Östman.*
Sevón, Guje och Lennart Sjöberg (red). *Emotioner och värderingar i näringslivet. EFIs Årsbok 2004.*
Wijkström, Filip and Stefan Einarsson. *Foundations in Sweden - Their scope, roles and visions.*

Dissertations
Anderson, Anders. *Essays in Behavioral Finance.*
Balsvik, Gudrun. *Information Technology Users: Studies of Self-Efficacy and Creativity among Swedish Newspaper Journalists.*
Blix, Magnus. *Essays in Mathematical Finance - Modelling the Futures Price.*
González Gómez, Andrés. *Nonlinear dynamics and smooth transition models.*
Grönqvist, Erik. *Selection and Moral Hazard in Health Insurance: Taking Contract Theory to the Data.*
Ivarsson Westerberg, Anders. *Papperspolisen - varfr ökar administrationen i moderna organisationer.*
Jutterström, Mats. *Att påverka beslut - företag i EUs regelsättande.*
Jönsson, Kristian. *Macroeconomic Aspects of Capital Flows to Small Open Economies in Transition.*
Larsson, Pär. *Förändringens villkor. En studie av organisatoriskt lärande och förändring inom skolan.*
Lagerwall, Björn. *Empirical Studies of Portfolio Choice and Asset Prices.*
Malmsten, Hans. *Properties and Evaluation of Volatility Models.*
Marshall, Cassandra. *Dating for Innovation. Creating and Recognizing Opportunities through Collaborative Interorganizational Relationships in Fluid Environments.*
Mattsson, Susanna. *På gränsen mellan ordning och oordning - tingens betydelse vid marknadsombildningar. En studie av svenska postväsendets ombildning under 1990-talet.*
Nilsson, Charlotte. *Studies in Environmental Economics: Numerical Analysis of Greenhouse Gas Policies.*
Nilsson, Hans. *Medborgaren i styrsystemet - beskrivning av VAD och HUR i styrning av kommunal verksamhet.*
Nystedt, Jens. *Competition, Regulation and Integration in International Financial Markets.*
Pajuste, Anete. *Corporate Governance and Controlling Shareholders.*
Richtnér, Anders. *Balancing Knowledge Creation. Organizational Slack and Knowledge Creation in Product Development.*

Salabasis, Mickael. *Bayesian Time Series and Panel Models - Unit Roots, Dynamics and Random Effects.*

Sandberg, Rickard. *Testing the Unit Root Hypothesis in Nonlinear Time Series and Panel Models.*

Skallsjö, Sven. *Essays on Term Structure and Monetary Policy.*

Strikholm, Birgit. *Essays on Nonlinear Time Series Modelling and Hypothesis Testing.*

Söderström, John. *Från Produkt till Tjänst. Utveckling av affärs- och miljöstrategier i produktorienterade företag.*

Talia, Krim. *The Scandinavian Currency Union, 1873-1924 - Studies in Monetary Integration and Disintegration.*

2003

Books

Lundahl, Mats (editor). *Globalization and Its Enemies. EFIs Årsbok 2003.*

Sundgren, Bo, Pär Mårtensson, Magnus Mähring and Kristina Nilsson (editors). *Exploring Patterns in Information Management. Concepts and Perspectives for Understanding IT-Related Change.*

Dissertations

Andersson, Henrik. *Valuation and Hedging of Long-Term Asset-Linked Contracts.*

Bergman, Malin. *Essays on Human Capital and Wage Formation.*

Damsgaard, Niclas. *Deregulation and Regulation of Electricity Markets.*

Eklund, Bruno. *Four Contributions to Statistical Inference in Econometrics.*

Hakkala, Katariina. *Essays on Restructuring and Production Decisions in Multi-Plant Firms.*

Holgersson, Charlotte. *Rekrytering av företagsledare. En studie i homosocialitet.*

Ivaschenko, Iryna. *Essays on Corporate Risk, U.S. Business Cycles, International Spillovers of Stock Returns, and Dual Listing.*

Lange, Fredrik. *Brand Choice in Goal-derived Categories - What are the Determinants?*

Le Coq, Chloé. *Quantity Choices and Market Power in Electricity Market.*

Magnusson, Peter R. *Customer-Oriented Product Development - Experiments Involving Users in Service Innovation.*

Meisiek, Stefan. *Beyond the Emotional Work Event Social Sharing of Emotion in Organizations.*

Mårtensson, Anders. *Managing Mission-Critical IT in the Financial Industry.*

Nilsson, Göran. *Processorientering och styrning - Regler, mål eller värderingar?*

Sandberg, Robert. *Corporate Consulting for Customer Solutions Bridging Diverging Business Logics.*

Sturluson, Jon Thor. *Topics in the Industrial Organization of Electricity Markets.*

Tillberg, Ulrika. *Ledarskap och samarbete - En jämförande fallstudie i tre skolor.*

Waldenström, Daniel. *Essays in Historical Finance.*

Wallén, Ulrika. *Effektivitet i grundskolan i anslutning till en stadsdelsnämndsreform.*

Ögren, Anders. *Empirical Studies in Money, Credit and Banking - The Swedish Credit Market in Transition under the Silver and the Gold Standards, 1834-1913.*

2002

Books

Schuster, Walter. *Fretagets Valutarisk - En studie av horisontella och vertikala styrprocesser.*

Sjöstrand, Sven-Erik och Pernilla Petrelius. *Rekrytering av koncernstyrelsen - Nomineringsförfaranden och styrelsesammansättning med fokus på kvinnors ställning och möjligheter. EFI/SNS Förlag*

Löwstedt, Jan och Bengt Stymne (red). *Scener ur ett företag - Organiseringsteori för kunskapssamhället. EFIs Årsbok 2002.EFI/Studentlitteratur.*

Dissertations

Barinaga, Ester. *Levelling Vagueness - A Study of Cultural Diversity in an International Project Group.*

Berglund, Johan. *De otillräckliga - En studie av personalspecialisternas kamp för erkännande och status.*

Bolander, Pernilla. *Anställningsbilder och rekryteringsbeslut.*

Damjanovic, Tatiana. *Essays in Public Finance.*

Ekman, Mattias. *Studies in Health Economics - Modelling and Data Analysis of Costs and Survival.*

Heyman, Fredrik. *Empirical Studies on Wages, Firm Performance and Job Turnover.*

Kallifatides, Markus. *Modern företagsledning och omoderna företagsledare.*

Kaplan, Michael. *Acquisition of Electronic Commerce Capability - The Cases of Compaq and Dell in Sweden.*

Mähring, Magnus. *IT Project Governance.*

Nilsson, Mattias. *Essays in Empirical Corporate Finance and Governance.*

Schenkel, Andrew. *Communities of Practice or Communities of Discipline - Managing Deviations at the resund Bridge.*

Skogsvik, Stina. *Redovisningsmått, värderelevans och informationseffektivitet.*

Ternström, Ingela. *The Management of Common-Pool Resources - Theoretical Essays and Empirical Evidence.*

Tullberg, Jan. *Reciprocitet - Etiska normer och praktiskt samarbete.*

Westling, Gunnar. *Balancing Innovation and Control - The Role of Face-to-face Meetings in Complex Product Development Projects.*

Viklund, Mattias. *Risk Policy - Trust, Risk Perception, and Attitudes.*

Vlachos, Jonas. *Risk Matters - Studies in Finance, Trade and Politics.*