

ESSAYS ON MULTIVARIATE VOLATILITY AND DEPENDENCE MODELS
FOR FINANCIAL TIME SERIES

ESSAYS ON MULTIVARIATE VOLATILITY AND DEPENDENCE MODELS
FOR FINANCIAL TIME SERIES

DIAA NOURELDIN
NUFFIELD COLLEGE
UNIVERSITY OF OXFORD

THIS THESIS IS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF ECONOMICS
AT THE UNIVERSITY OF OXFORD

DEPARTMENT OF ECONOMICS
UNIVERSITY OF OXFORD
TRINITY TERM
2011

CONTENTS

List of Figures	iii
List of Tables	iv
Preface	v
Abstract	vii
Statement of Publication and Word Count	viii
Chapter 1: Introduction	1
Chapter 2: Forecasting Changes in the Term Structure of Interest Rates	14
Chapter 3: Multivariate High-Frequency-Based Volatility (HEAVY) Models	70
Chapter 4: Flexible Covariance-Targeting Volatility Models Using Rotated Returns	124

LIST OF FIGURES

Chapter 2

1	Actual and fitted yields at selected dates	22
2	Three-dimensional plot of the term structure	39
3	Estimated DNS factors	41
4	Selected features of the estimated DNS factors	44
5	Conditional dependence measures: time-varying t copula model	52
6	Forecast RMSE: D-VAR(2) versus L-VAR(1)	54
7	Directional forecast accuracy: D-VAR(2) versus L-VAR(1)	55

Chapter 3

1	SPY and BAC realised measures	97
2	Leverage effect in model residuals	99
3	SPY-BAC conditional correlation: one-step and multi-step forecasts	100
4	Predictive ability tests' t -statistics for the covariance targeting HEAVY and GARCH models	109

Chapter 4

1	Simulated path of the conditional variance, covariance and correlation	136
2	XOM and AA actual and rotated returns	155
3	Conditional correlations from the diagonal BEKK, OGARCH and DCC models	161
4	Conditional variances, correlation and beta from the diagonal OGARCH and DCC models	164

LIST OF TABLES

Chapter 2

1	Summary statistics for the yields at different maturities	40
2	Summary statistics for the DNS factors	42
3	Estimation results for the marginal models	47
4	Estimation results for the copula models	50
5	Models considered in conditional density forecasting	56
6	Predictive likelihood t -statistics for OOS conditional density forecasts	60

Chapter 3

1	SPY-BAC: scalar HEAVY and GARCH estimation and forecast results	98
2	SPY-BAC: scalar HEAVY and GARCH estimation and forecast results using different realized measures	102
3	Other asset pairs: scalar HEAVY and GARCH estimation and forecast results	103
4	SPY-BAC: diagonal HEAVY and GARCH estimation and forecast results	105
5	Other asset pairs: diagonal HEAVY and GARCH estimation and forecast results	106
6	Ten DJIA assets: covariance targeting scalar HEAVY and GARCH model	108

Chapter 4

1	XOM-AA: unconditional covariance, eigenvectors and eigenvalues	156
2	XOM-AA: BEKK, OGARCH and GOGARCH estimation results	157
3	XOM-AA: DCC estimation results	160
4	SPY-XOM: BEKK, OGARCH, GOGARCH and DCC estimation results	162
5	Ten DJIA assets: First two eigenvectors	164
6	Ten DJIA assets: BEKK, OGARCH and DCC estimation results	166

PREFACE

Writing this thesis has been a great journey of exploration. After four years of a deep dive into financial econometrics, I grew more confident that there is much more to explore. From this point onward, I look forward to doing just that. This journey would not have been successful without the support of many people, to whom I would like to express my gratitude.

My greatest and most sincere thanks go to my two supervisors, Neil Shephard and Kevin Sheppard, for their invaluable guidance and advice. The deep insights and inexhaustible supply of ideas provided by Neil Shephard were instrumental in shaping the scope and content of this thesis. I am greatly indebted to him for his enthusiasm, encouragement and unwavering support. I also sincerely thank Kevin Sheppard for his intellectual engagement, approachableness and great efforts to explain things clearly and simply. It has been a great privilege working under their supervision, and I can only hope to be able to work with them on future research.

I would also like to thank my MPhil thesis supervisor, Andrew Patton. Writing my MPhil thesis under his supervision has been a remarkable learning experience, as I was taking my first steps into the world of academic research. I am deeply grateful for his solid research advice and continuing support. My thanks also extend to my DPhil thesis examiners, Jennifer Castle and Asger Lunde, for their comments and insightful suggestions.

My DPhil experience has been greatly enhanced by my fellow students at the Department of Economics and Nuffield College. The stimulating and pleasant environment they provided has been a significant driver to seeing this thesis through to fruition. I would like to particularly thank Cavit Pakel, Kasper Lund-Jensen, Kadambari Prasad, Jouni Sohkanen, Martina Kirchberger, Andrew Whitby, Sean Muller, Anna Stupnytska, Aino Levonmaa, Alexander Taylor, Khatchig Karamanoukian, Sriram Lakshman and Joel Strange.

I would also like to sincerely thank my college supervisor, John Muellbauer, for the many interesting discussions on cross-cutting themes in finance and macroeconomics, as well as his guidance and support. My thanks also extend to Stephanie Wright, Nuffield College, and Julie Minns, Department of Economics, for their help at various points during this journey.

I owe a great debt to the Oxford-Man Institute (OMI), at which I was based during my DPhil research. The unequalled research environment at OMI has been a source of inspiration and motivation. I thank my OMI colleagues for their pleasant company, particularly Bahman Angoshtari, Nathaniel Korda, Arnaud Lionnet, Sylvestre Burgos, Youness Boutaib, Gechun Liang and Michael Streatfield. I would also like to express my sincere thanks to the OMI administration team, particularly Lucy Mullins, Diane Feast, Helen Crockford, Gerd Heber and Hussein El-Hindi.

I have benefited tremendously from the excellent research environment at Oxford. The Nuffield Econometrics Seminar and the Econometrics Group Lunch have been

an important platform for intellectual engagement with leading academics in my field of interest. I have learned a great deal from the insights of David Hendry, Bent Nielsen, Steve Bond, Jennifer Castle, Jurgen Doornik and Debopam Bhattacharya during these discussions. I am also thankful for the comments I have received from David Hendry and Steve Bond on Chapter 3 of this thesis.

Particular thanks go to the staff of the Social Sciences Library, Nuffield Library and the Department of Statistics Library for their assistance. In addition, I gratefully acknowledge financial support from the Oxford-Man Institute, the Department of Economics and Nuffield College.

Before embarking on my journey at Oxford, I enjoyed a rewarding learning experience during my studies in Egypt. I would like to particularly express my appreciation to William Mikhail, my former supervisor at the American University in Cairo, who first taught me econometrics. I would also like to immensely thank Adel Beshai for his continuing advice and encouragement.

It is undoubtedly difficult to achieve this without the support of family and friends. I thank my mother, Farida Lotfy, and my brothers, Ahmed, Hatem and Ayman, for always being there for me, in the most difficult of times. I wish my father, Mohamed Nouredin, was with us to witness the conclusion of this important phase in my career. I am eternally thankful for the wonderful father and great person he was. Fortunately, I met my father-in-law, Gamal Okail, who deserves special mention for all his support. I also thank my dear friends in Egypt and Oxford, especially Ahmed Hafez, Katherine Allard and Nick Lott, for being exceptional companions.

Last, and most importantly, I am deeply grateful to my wife, Nancy Okail, whose constant support and enthusiasm made the journey a truly pleasant experience. As always, she has been an enduring source of comfort and encouragement. Our children, Adam and Farida, were born during my studies. I thank them for their lovely morning smiles, and I am happy to dedicate this thesis to them.

Diaa Nouredin
September 2011

ABSTRACT

This thesis investigates the modelling and forecasting of multivariate volatility and dependence in financial time series. The first paper proposes a new model for forecasting changes in the term structure (TS) of interest rates. Using the level, slope and curvature factors of the dynamic Nelson-Siegel model, we build a time-varying copula model for the factor dynamics allowing for departure from the normality assumption typically adopted in TS models. To induce relative immunity to structural breaks, we model and forecast the factor changes and not the factor levels. Using US Treasury yields for the period 1986:3-2010:12, our in-sample analysis indicates model stability and we show statistically significant gains due to allowing for a time-varying dependence structure which permits joint extreme factor movements. Our out-of-sample analysis indicates the model's superior ability to forecast the conditional mean in terms of root mean square error reductions and directional forecast accuracy. The forecast gains are stronger during the recent financial crisis. We also conduct out-of-sample model evaluation based on conditional density forecasts.

The second paper introduces a new class of multivariate volatility models that utilizes high-frequency data. We discuss the models' dynamics and highlight their differences from multivariate GARCH models. We also discuss their covariance targeting specification and provide closed-form formulas for multi-step forecasts. Estimation and inference strategies are outlined. Empirical results suggest that the HEAVY model outperforms the multivariate GARCH model out-of-sample, with the gains being particularly significant at short forecast horizons. Forecast gains are obtained for both forecast variances and correlations.

The third paper introduces a new class of multivariate volatility models which is easy to estimate using covariance targeting. The key idea is to rotate the returns and then fit them using a BEKK model for the conditional covariance with the identity matrix as the covariance target. The extension to DCC type models is given, enriching this class. We focus primarily on diagonal BEKK and DCC models, and a related parameterisation which imposes common persistence on all elements of the conditional covariance matrix. Inference for these models is computationally attractive, and the asymptotics is standard. The techniques are illustrated using recent data on the S&P 500 ETF and some DJIA stocks, including comparisons to the related orthogonal GARCH models.

Keywords: multivariate volatility; HEAVY; GARCH; orthogonal GARCH; DCC; realized covariance; dynamic Nelson-Siegel model; copula; covariance targeting; common persistence; directional forecasting; predictive likelihood; Wishart distribution.

JEL Codes: C32; C52; C53; C58; E43.

STATEMENT OF PUBLICATION AND WORD COUNT

Parts of this thesis have been previously submitted for an academic degree as well as for publication. An earlier version of Chapter 2 has been submitted for the degree of Master of Philosophy in the subject of Economics at the University of Oxford. Also, a slightly modified version of Chapter 3 is forthcoming in the *Journal of Applied Econometrics*, with Neil Shephard and Kevin Sheppard.

Some of the chapters have also been presented at conferences and seminars. An earlier version of Chapter 2 has been presented at the 15th annual conference of the African Econometric Society, Cairo, 7-9 July, 2010. Earlier versions of Chapter 3 have been presented at the Department of Economics, University of Oxford, the Oxford-Man Institute, the 2nd Humboldt-Copenhagen conference on financial econometrics, Copenhagen, 13-14 May, 2011, and the 4th annual conference of the Society for Financial Econometrics (SoFiE), Chicago, 15-17 June, 2011.

In the production of this thesis, the following applications were used: Scientific Word for setting the text; Matlab (R2008b, R2009b, R2010a) and OxMetrics (5.0, 6.0) for the computations and production of charts and figures. The majority of computations were done by author-written code in Matlab. The *Copula toolbox* of Andrew Patton, and the *MFE toolbox* and *Realized Measures toolbox* of Kevin Sheppard were also used in some of the computations. OxMetrics was used for producing most of the charts and figures.

WORD COUNT

The body of this thesis comprises 169 pages (including references), with a typical page of text containing around 300 words, making the thesis approximately 50,700 words long.

ESSAYS ON MULTIVARIATE VOLATILITY AND DEPENDENCE MODELS
FOR FINANCIAL TIME SERIES

CHAPTER I

INTRODUCTION

The last twenty years have witnessed many exciting developments in financial econometrics. The breadth and depth of research in this area is a reflection of both the importance of understanding financial risk, and the challenging nature of the research questions. The growth in financial markets coupled with fast-paced financial innovation has fuelled the need to understand how to price newly-created products, necessitating a reliable characterisation of their payoff risks. Technological advancement in data storage made available vast data sets of intra-day financial transactions which revolutionised our understanding of how to measure and model return volatility. The proliferation of large portfolios spanning different asset classes prompted thinking about dependence modelling in high dimensions. In parallel, recurring episodes of financial turmoil brought financial risk management to the forefront.

Recent progress is evident in three interrelated research streams: estimation, modelling and forecasting. The first objective of this introductory chapter is to highlight some of the recent advances in these areas. Providing a comprehensive account of the recent literature is not an easy task given space limitations, so this is a selective overview focusing on the contributions that are most related to this thesis. The second objective is to discuss the contribution of the three chapters, constituting the body of the thesis, in relation to modelling and forecasting volatility and dependence using financial time series.

With regard to estimation, the advent of high-frequency (HF) data has led to revolutionary developments in the measurement of volatility. With improvements in data storage and computational capacities, it is now possible to obtain intra-day observations on quotes, transaction prices and traded quantities for most financial securities. This also brought to the forefront issues related to market microstructure effects, which is now a well-established research area in finance.

For univariate time series, the use of *ex post* measures of volatility, known as ‘realised’ measures, has become widespread. These nonparametric estimators, computed from HF data, are found to significantly improve on the traditional squared daily return in terms of the signal-to-noise ratio (Andersen and Bollerslev, 1998). The realised variance, systematically studied in Andersen et al. (2001) and Barndorff-Nielsen and Shephard (2002), is the first and one of the most popular realised measures. In the absence of market microstructure noise, it is a consistent estimator of the quadratic variation of the underlying price process. To account for market microstructure noise, refined estimators are proposed in, for example, Zhang et al. (2005), Zhang (2006), Barndorff-Nielsen et al. (2008) and Jacod et al. (2009). Also multivariate equivalents have been proposed in Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen et al. (2011). Recent surveys of the econometrics of realised measures can be found in Barndorff-Nielsen and Shephard (2007) and Andersen et al. (2010).¹

In volatility and dependence modelling, there have been parallel developments focusing mainly on: (i) enhancing model dynamics, (ii) modelling large dimensional systems, and (iii) improved modelling of cycles and structural breaks. Starting with the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and the generalised ARCH (GARCH) model of Bollerslev (1986), research in this area has proliferated providing various extensions in terms of functional forms, allowing for asymmetric effects, and the distinction between long-run and short-run volatility components. Bollerslev et al. (1992), Teräsvirta (2009) and Francq and Zakoian (2010) provide extensive surveys of related theory and applications.

The advent of realised measures opened two doors: firstly, a literature focused on

¹See also the special issue on realised volatility in the *Journal of Econometrics* 160, 2011.

modelling and forecasting the time series of realised measures, which primarily builds on previous progress in GARCH-type models. Secondly, by allowing the use of realised measures to model and forecast the conditional variance of daily returns. Replacing the squared return, which drives the ARCH model dynamics, by a realised measure is shown to provide significant in- and out-of-sample (OOS) gains. See, for example, Engle (2002b), Engle and Gallo (2006), Cipollini et al. (2007), Brownlees and Gallo (2010), Shephard and Sheppard (2010) and Hansen et al. (2011).

In parallel, multivariate volatility models have witnessed significant developments.² The two main challenges in multivariate volatility models is guaranteeing positive semidefiniteness of the estimated conditional covariance matrix, and the ‘curse of dimensionality’ in reference to the - often exponential - increase in the number of parameters with the dimensionality of the problem. Reviews of the multivariate GARCH literature are given by, for example, Bauwens et al. (2006), Engle (2009) and Silvennoinen and Teräsvirta (2009). Several proposals have been offered that successfully meet the first challenge, e.g. the BEKK model of Engle and Kroner (1995). However, the curse of dimensionality remains a pressing problem, which the dynamic conditional correlations (DCC) model of Engle (2002a) partly addresses. This model has become a workhorse for much research in the last few years aiming to introduce more flexible dynamics for the conditional correlations; see, for example, Cappiello et al. (2006), Billio et al. (2006), Billio and Caporin (2009) and Hafner and Franses (2009).

On a related front, the use of copulas for dependence modelling has been a promising development for financial risk management. Copulas allow for the decomposition of the

²Developments in both univariate and multivariate volatility models are also evident in the stochastic volatility literature recently reviewed in Asai et al. (2006) and Chib et al. (2009). In this overview, we focus only on the developments in the GARCH literature.

probability density function of random vectors into marginal densities and a copula that summarises dependence among the variables. It has become a popular tool for studying asymmetric and extreme forms of dependence, and has been applied to risk management models particularly in credit risk modelling. An introduction to copula theory can be found in Joe (1997) and Nelsen (2006), while Patton (2009) provides a recent survey of copula applications in finance.

Forecasting and forecast evaluation is also a fast-growing area of research with many applications. There is growing consensus that model selection in finance should be based on OOS criteria. Most models are constructed for forecasting future volatility and dependence, and evaluating the model's potency OOS seems a natural approach. There is no consensus, however, on whether model selection should be based on statistical or economic criteria; see the discussion in Granger and Machina (2006). On another front, one challenge in OOS evaluation in these models is that volatility is unobserved. The recent contributions of Hansen and Lunde (2006) and Patton (2011) show that certain loss functions can provide consistent ranking of forecasting models as these loss functions are robust to noise in the proxy for the latent volatility. Multivariate extensions of these concepts are discussed in Patton and Sheppard (2009) and Laurent et al. (2009).

Forecast evaluation tends to be based on some loss function, whether statistical or economic. Predictive likelihood is also one approach which considers OOS model fit comparisons. Central to forecast evaluation is the use of some test statistic for decision making. An early contribution is the test of Diebold and Mariano (1995), which was later formalised by West (1996) and Giacomini and White (2006). Using the predictive likelihood approach and building on the ideas of in-sample testing in Cox (1961) and Vuong (1989), there have

been similar propositions for OOS model evaluation, e.g. Amisano and Giacomini (2007).³

Despite significant progress in these three research streams, there are still quite a few unsolved problems. The focus of the thesis is on modelling and forecasting multivariate financial time series, in an attempt to address some of these problems. The models we develop build on recent advances in the literature, particularly the use of copula methods, utilising HF data for modelling time-varying conditional covariance matrices, and techniques that allow for introducing more flexible dynamics in multivariate volatility models.

Chapter 2, *Forecasting Changes in the Term Structure of Interest Rates*, proposes using copula methods to model and forecast changes in the shape of the term structure (TS) of interest rates. The TS is naturally a large-dimensional object since it contains information about numerous fixed-income instruments traded in bond markets every day. Our approach to the problem is based on dimension reduction. We utilise the popular dynamical Nelson-Siegel (DNS) model of Diebold and Li (2006), which builds on the model of Nelson and Siegel (1987). This model allows us to focus on the systematic component of yield variation explained by the DNS factors which are interpreted as the slope, level and curvature of the TS. We focus on modelling and forecasting the changes in the DNS factors. This marks a point of distinction compared to existing models which focus on the factor levels.

Our approach is motivated by evidence of non-stationarity in the factors, and the objective is to build a model that is relatively immune to structural breaks. We model the joint density of the factor changes using copulas, which enables us to relax the assumptions typically maintained in existing studies. For example, normality is usually assumed in TS models for tractability in estimation without necessarily being empirically plausible. Us-

³For model comparisons involving more than two models, multiple hypothesis testing approaches have been offered in White (2000) and Hansen et al. (2011).

ing copulas we create multivariate distributions that allow for individual and joint extreme behaviour in the factor changes, which provides statistically significant gains in goodness of fit. In addition, TS models typically assume that the factors' dependence structure is time-invariant. We show that this is rejected empirically by using time-varying copula specifications. The additional flexibility of the model proves very useful in characterizing the factor dynamics during the recent financial crisis, when the TS shows rather unprecedented behaviour.

We show the model's favourable OOS performance, particularly for longer horizon forecasts. Compared to two widely-used benchmark models – including the random walk model – we show reductions in root mean squared error and improved directional forecasting, particularly during the financial crisis. We also report conditional density forecasting results, which aim for assessing the usefulness of the model's new features in real-time forecasting. Although the results are statistically insignificant, the direction of the predictive likelihood gains confirm the in-sample findings, namely that models with time-varying dependence that permit extreme joint movements in the factors perform better.

Chapter 3, *Multivariate High-Frequency-Based Volatility (HEAVY) Models*, presents a new class of multivariate volatility models which utilise HF data. We develop a multivariate extension of the model of Shephard and Sheppard (2010). The model replaces the outer product of daily returns, used in multivariate GARCH models to drive the dynamics of the conditional covariance, with a multivariate realised measure. Focusing on a simple specification akin to a GARCH(1,1) model, we discuss its dynamic properties and provide closed-form forecasting formulas. We also discuss estimation and inference assuming the model's innovations follow a Wishart distribution, with the model's stochastic representation being a matrix variate generalisation of the univariate multiplicative error model

(MEM) of Engle (2002b) and the vector MEM of Cipollini et al. (2007). We also discuss covariance targeting using a rotation technique for the realised measure to make covariance targeting operational. Inference for this case is straightforward based on the results on two-step Generalised Method of Moments estimation discussed in, for example, Newey and McFadden (1994).

As a new proposition in the context of multivariate volatility models, we develop an OOS model evaluation strategy based on the margins-copula decomposition of the predictive joint likelihood. This enables us to decompose the OOS gains/losses due to variance forecasts, correlation forecasts or a combination of both. Thus we are able to evaluate the usefulness of HF data to forecast conditional variances and conditional correlations independently.

We apply the scalar and diagonal versions of the model to the returns of the S&P 500 index and 10 of the most liquid stocks in the Dow Jones Industrial Average (DJIA) index, and find that the use of the realised measure provides significant OOS gains, particularly at short forecast horizons. Interestingly, in most cases the gains in forecasting the conditional correlations persist for longer horizons up to 22 days into the future. We also use covariance targeting as a useful device for modelling large systems, and apply the model to all 10 DJIA stock returns. We reach the same conclusions, and in this case the copula-margins decomposition turns out to be a useful approach in evaluating multivariate volatility models.

Chapter 4, *Flexible Covariance-Targeting Volatility Models Using Rotated Returns*, focuses on the covariance targeting technique in multivariate volatility models. One of the problems with using covariance targeting is identifying parameter restrictions to ensure positive semidefiniteness of the target, which are easy to impose in practice. This is

straightforward with scalar models but not with more richly parameterised models. Our approach to solving the problem is a simple one, yet it turns out to be potent in practice. Using the spectral decomposition for the unconditional covariance matrix, we use the matrices of eigenvalues and eigenvectors to rotate the raw daily returns and re-centre them around the identity matrix. For diagonal models, using both the BEKK and DCC parameterisations, this technique allows us to do covariance targeting under easy-to-impose parameter restrictions. The set of parameter restrictions ensuring covariance stationarity of the conditional covariance matrix also guarantee positive semidefiniteness of the target.

In addition, we propose a new specification for multivariate volatility models which we call the ‘common persistence’ volatility model. Compared to diagonal models, this specification reduces the number of dynamic parameters by about a half, while remarkably resulting only in a slight deterioration in fit. This model is motivated by the empirical observation that persistence levels in return volatilities are less heterogeneous than their smoothness levels. Our model imposes a common persistence parameter on all elements of the conditional covariance matrix, but allows them to have different smoothness levels as they are filtered in-sample. The rotation approach coupled with the extensions for the BEKK and DCC models give specifications which are currently among the most flexible to model multivariate volatility in moderately large dimensions.

In assessing the model’s performance, we run a horse race using various specifications of the BEKK and DCC models adapted to our extension. We also apply the same specifications to the orthogonal GARCH (OGARCH) model of Alexander and Chibumba (1997) and Alexander (2001), and also the generalised OGARCH (GOGARCH) model of van der Weide (2002) since both are also based on the idea of rotating the raw returns. We show that our model leads to statistically significant gains; in particular, the diagonal DCC and

common persistence DCC extensions provide considerable improvement in the goodness of fit of the conditional correlations.

Since each of the following chapters focuses on a unique problem, the thesis is structured such that each chapter is a stand-alone reading. Each chapter provides its own introduction, notation and references. When needed, end-of-chapter appendices are included to introduce related concepts or provide proposition proofs, making the chapter self-contained. There is no concluding chapter to the thesis as each chapter has its own conclusions.

References

- Alexander, C. (2001). Orthogonal GARCH. In *Mastering Risk*, Volume 2, pp. 21–38. Prentice Hall.
- Alexander, C. O. and A. M. Chibumba (1997). Multivariate orthogonal factor GARCH. Mimeo: University of Sussex, UK.
- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25, 177–190.
- Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold (2010). Parametric and nonparametric measurement of volatility. In Y. Aït-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*, pp. 67–138. Amsterdam: North Holland.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001). The distribution of exchange rate volatility. *Journal of the American Statistical Association* 96, 42–55. Correction published in 2003, volume 98, page 501.
- Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: a review. *Econometric Reviews* 25, 145–175.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–169.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.

- Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.
- Barndorff-Nielsen, O. E. and N. Shephard (2007). Variation, jumps and high frequency data in financial econometrics. In R. Blundell, T. Persson, and W. K. Newey (Eds.), *Advances in Economics and Econometrics. Theory and Applications, Ninth World Congress*, Econometric Society Monographs, pp. 328–372. Cambridge University Press.
- Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.
- Billio, M. and M. Caporin (2009). A generalized dynamic conditional correlation model for portfolio risk evaluation. *Mathematics and Computers in Simulation* 79, 2566–2578.
- Billio, M., M. Caporin, and M. Gobbo (2006). Flexible dynamic conditional correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters* 2, 123–130.
- Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics* 51, 307–327.
- Bollerslev, T., R. Y. Chou, and K. F. Kroner (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.
- Brownlees, C. T. and G. M. Gallo (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* 8, 29–56.
- Cappiello, L., R. F. Engle, and K. Sheppard (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 4, 537–572.
- Chib, S., Y. Omori, and M. Asai (2009). Multivariate stochastic volatility. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 365–400. Springer-Verlag.
- Cipollini, F., R. F. Engle, and G. M. Gallo (2007). A model for multivariate non-negative valued processes in financial econometrics. Unpublished paper: Stern School of Business, New York University.
- Cox, D. R. (1961). Tests of separate families of hypotheses. *Proceedings of the Berkeley Symposium* 4, 105–123.
- Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica* 50, 987–1007.
- Engle, R. F. (2002a). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.

- Engle, R. F. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R. F. (2009). *Anticipating Correlations*. Princeton University Press.
- Engle, R. F. and J. P. Gallo (2006). A multiple indicator model for volatility using intra daily data. *Journal of Econometrics* 131, 3–27.
- Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Francq, C. and J. M. Zakoian (2010). *GARCH Models*. Chichester: Wiley.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Granger, C. W. J. and M. J. Machina (2006). Forecasting and decision theory. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 81–98. North-Holland: Amsterdam.
- Hafner, C. M. and P. H. Franses (2009). A generalized dynamic conditional correlation model: simulation and application to many assets. *Econometric Reviews* 28, 612–631.
- Hansen, P. R., Z. A. Huang, and H. H. Shek (2011). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*. Forthcoming.
- Hansen, P. R. and A. Lunde (2006). Consistent ranking of volatility models. *Journal of Econometrics* 131, 97–121.
- Hansen, P. R., A. Lunde, and J. M. Nason (2011). The model confidence set. *Econometrica*. Forthcoming.
- Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic Processes and Their Applications* 119, 2249–2276.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. London: Chapman and Hall.
- Laurent, S., J. V. K. Rombouts, and F. Violante (2009). On loss functions and ranking forecasting performances of multivariate volatility models. Working paper: Centre Interuniversitaire sur le Risque, les Politiques Economiques et l’Emploi (CIRPEE).
- Nelsen, R. B. (2006). *An Introduction to Copulas* (2 ed.). New York: Springer.
- Nelson, C. R. and A. F. Siegel (1987). Parsimonious modelling of yield curve. *Journal of Business* 60, 473–489.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics*, Volume 4, pp. 2111–2245. North-Holland.
- Patton, A. J. (2009). Copula-based models for financial time series. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 767–785. Springer-Verlag.

- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
- Patton, A. J. and K. K. Sheppard (2009). Evaluating volatility and correlation forecasts. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 801–838. Springer-Verlag.
- Shephard, N. and K. K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.
- Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer-Verlag.
- Teräsvirta, T. (2009). An introduction to univariate GARCH models. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 17–42. Springer-Verlag.
- van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17, 549–564.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
- White, H. (2000). A reality check for data snooping. *Econometrica* 68, 1097–1126.
- Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli* 12, 1019–1043.
- Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005). A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.

CHAPTER II

FORECASTING CHANGES IN THE TERM STRUCTURE OF INTEREST RATES

1 Introduction

Modelling and forecasting the term structure (TS) of interest rates has been one of the most active research areas in finance and financial econometrics in recent years. The TS of interest rates, often referred to as the yield curve, contains a vast amount of information about financial and macroeconomic variables as well as their expected future values. Understanding the dynamics of the TS is fundamental for a wide range of objectives and applications from monetary policy conduct to financial risk management. It is of central importance in the pricing of fixed income derivatives and hedging strategies.

The TS literature is too broad to be thoroughly covered in this paper, so we focus only on the main contributions to the literature and pay more attention to the models related to ours. Two of the early TS models are Vasicek (1977) and Cox et al. (1985). They are considered ‘equilibrium models’ as opposed to the ‘no-arbitrage models’ of Hull and White (1990) and Heath et al. (1992). Equilibrium models focus on modelling the dynamics of the short rate, and rates at longer maturities are determined by additive risk premia which are potentially time-varying, while no-arbitrage models are more focused on finding the best fitting model to the cross-section of yields ignoring the TS dynamics. More recently, the class of affine TS models has become the mainstream in finance for modelling and forecasting the TS; see, for example, Duffie and Kan (1996) and Dai and Singleton (2000).¹

In the TS literature, different model specifications aim to strike a balance between parsimony and goodness of fit. The choice in practice is driven by the objective of application. A parsimonious model is more suitable to analyse the time series dynamics of yields or for

¹ Also some recent models link the TS to the macroeconomy. These are known as ‘macro-finance’ models, which are recently surveyed in Rudebusch (2010).

forecasting purposes, while a model providing a close-to-perfect fit is needed in the pricing of bonds and fixed-income derivatives.

Some of the popular TS models have a factor-type structure assuming that yields at different maturities are driven by a few latent factors. In a three-factor model, the factors are usually interpreted as the level, slope and curvature of the yield curve. The Nelson-Siegel (NS) model - introduced by Nelson and Siegel (1987) - uses exponential components to estimate the latent factors and it has been a popular model among practitioners. Diebold and Li (2006) introduced the dynamic Nelson-Siegel (DNS) model, which has become the workhorse for much research in the last few years. Refinements and extensions of the DNS model include, for example, Diebold et al. (2006, 2008), Hautsch and Ou (2008), Christensen et al. (2009, 2011), and Koopman et al. (2010). Diebold et al. (2005) link the Diebold and Li (2006) model to affine TS models, a direction reinforced by the work of Christensen et al. (2009, 2011).

The exponential components functional form proposed by Nelson and Siegel (1987) leads to dimension reduction, and the model is readily interpretable as a factor model. It also tends to provide a good fit to the cross section of yields given the flexibility of its functional form. For example, it is able to fit upward sloping, inverted and humped yield curves. Svensson (1995) extends the NS model by including a second curvature factor to be able to fit a double-humped yield curve; this model is currently used by the Federal Reserve to estimate the factors of the US Treasuries TS. De Pooter (2007) cites and develops additional extensions of the NS model. All of these extensions may marginally improve the cross-section fit, but at the risk of overparameterisation when modelling the factor dynamics. In this paper, we focus only on the three-factor DNS model of Diebold and Li (2006).

Our objective is to develop a model that is capable of producing accurate forecasts of the TS. We aim for a parsimonious model that is relatively immune to structural breaks. The first requirement is a well-known ingredient of successful forecasting models. The second requirement is due to the complexity underlying the object of interest which makes its dynamics inherently vulnerable to structural breaks. Motivated by these requirements for a successful forecasting model, our model is built for the *changes*, and not the *levels*, of the TS factors. We view the DNS model as a useful dimension reduction device to extract from the large-dimensional yield curve three factors that represent systematic variation in yields at different maturities.² We focus on modelling and forecasting the joint density of the factor changes, thus we are essentially modelling and forecasting changes in the shape of the TS. This marks the first contribution of the paper and a point of contrast to many existing models which focus on modelling the factor levels.³

The second contribution is an attempt to provide a complete characterisation of the joint density of the factor changes. To this end, we use the margins-copula decomposition for the joint density, which enables us to achieve flexibility in matching the moments of the data. Existing models typically assume conditional normality of the factors, and in the models that allow for factor interaction, it is usually assumed that the factors' dependence structure is time-invariant and does not permit joint extreme movements in the factors.⁴ These assumptions are needed to ensure tractability when estimating the

²As we show later in the empirical analysis, the three factors capture close to 92 percent of the variation in yields. We consider the remaining variation to be idiosyncratic due to factors extraneous to our model.

³Non-stationarity is evident in yields at different maturities, which prompted some studies of the TS to use cointegration in modelling and forecasting; see Pagan et al. (1996) for a review. A more recent example is Bowsher and Meeks (2008). However, most TS models, particularly those related to the DNS model and affine TS models, focus on the dynamics of the factor levels.

⁴We use the term 'dependence structure' to refer to a metric that sufficiently summarizes the degree and direction of association between two or more variables. For instance, under the assumption of multivariate normality, the correlation matrix sufficiently captures the dependence structure of the variables under study. For extreme joint movements in random variables, often referred to as 'tail events', copulas are becoming

model parameters, but have not been tested in practice. Our framework allows us to statistically evaluate the validity of these assumptions. Indeed, Dai and Singleton (2000) have noted that flexible models of the factor correlations have a substantial impact on goodness of fit.

The third contribution of this paper is going beyond conventional forecast evaluation based solely on the accuracy of the mean forecast for yields at different maturities. We propose an additional criterion based on directional forecast accuracy, that is the model's ability to predict the direction of change in the TS. In addition, we study out-of-sample (OOS) conditional density forecasts to understand the importance of the model's new features in a setting that emulates real-time forecasting. In the TS literature, there is a dearth of studies on density forecasts despite their importance in pricing applications and hedging strategies. One notable contribution in this regard is Egorov et al. (2006) where they focus on affine TS models.

The additional flexibility of our model is not a free lunch. The price to pay is having to estimate the model parameters in two steps, which entails a loss of efficiency. Two-step estimation is typically used in copula models where the marginal densities are fitted in a first step, and then the probability integral transforms from the fitted densities are used in the second step to estimate the copula parameters. As we discuss later, our inference method does take into account the accumulation of parameter estimation error from the first step estimation. In addition, simulation studies in Patton (2006a) and Joe and Xu (1996) indicate that the loss in estimation efficiency in this case is rather negligible.

The empirical analysis utilises a relatively long sample period of US Treasury yields, extending from 1986:3 to 2010:12, which enables us to discern the difficulty of developing

increasingly popular tools to capture this kind of behaviour in financial time series.

a model that is immune to structural breaks. Our data sample also includes the recent financial crisis, and the associated decline in short term interest rates to unprecedented levels. The behaviour of the TS during this period provides additional insights about TS dynamics, especially with the short end of the yield curve currently being pinned down close to the zero level. The period of the financial crisis and its aftermath is part of our OOS period used for conditional mean and conditional density forecast evaluation. This is one of the first accounts of the behaviour of the TS during this turbulent period.

The paper is organised as follows: Section 2 provides the theoretical background on the DNS model. In Section 3 we present our model and discuss estimation and inference. Section 4 presents the results of the empirical analysis, while Section 5 offers some concluding remarks. Appendix A provides an overview of copula theory making the paper self-contained.

2 Theoretical Background

2.1 The Dynamic Nelson-Siegel Model

We utilise the DNS model proposed by Diebold and Li (2006),

$$Y_t^{NS}(\tau_j) = L_t + \left(\frac{1 - e^{-\lambda\tau_j}}{\lambda\tau_j} \right) S_t + \left(\frac{1 - e^{-\lambda\tau_j}}{\lambda\tau_j} - e^{-\lambda\tau_j} \right) C_t, \quad j = 1, \dots, n, \quad (1)$$

where $Y_t^{NS}(\tau_j)$ is the NS-*smoothed* yield on a bond with maturity τ_j at time t . L_t , S_t and C_t are latent factors representing, respectively, the level, slope and curvature of the TS, while $\lambda > 0$ is a free parameter. We assume τ_j is measured in months where one month equals 30.4375 days. This model was first introduced by Nelson and Siegel (1987) to fit the

cross-section of yields at a point in time, and later extended by Diebold and Li (2006) by adding the temporal dimension to the yields and the factors. Some extensions of the DNS model assume that λ is time-varying, e.g. Hautsch and Ou (2008) and Koopman et al. (2010). We discuss this extension in Section 2.3.

In (1), $Y_t^{NS}(\tau_j)$ represents the component of an n -dimensional yield curve that is smoothed according to the NS curve, hence the superscript NS. Actual yields do not, of course, lie on a smooth curve, and thus will differ from (1) by maturity-specific (or idiosyncratic) components. Therefore we work in practice with the following representation of the actual yields $Y_t(\tau_j)$, $j = 1, \dots, n$:

$$\begin{pmatrix} Y_t(\tau_1) \\ Y_t(\tau_2) \\ \vdots \\ Y_t(\tau_n) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1} \\ 1 & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} & \frac{1-e^{-\lambda\tau_n}}{\lambda\tau_n} - e^{-\lambda\tau_n} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} e_t(\tau_1) \\ e_t(\tau_2) \\ \vdots \\ e_t(\tau_n) \end{pmatrix}, \quad (2)$$

assuming $e_t(\tau_j) \stackrel{i.i.d.}{\sim} N(0, \sigma_t^2)$ for a given time t . This is similar to the representation in Diebold et al. (2006) but we interpret $e_t(\tau_j)$ differently as discussed shortly. We will use the following notation: Y_t to denote the n -dimensional vector of actual yields, $X_t = (L_t, S_t, C_t)'$ to denote the vector of factor levels and e_t to denote the n -dimensional vector of idiosyncratic yield components.

Figure 1 plots the actual TS and the fitted NS curve at selected dates from the data sample used in empirical analysis. The dates are chosen to illustrate the ability of the NS curve to fit a variety of TS shapes typically observed in bond markets, such as upward-sloping (top left), inverted (top right), humped (middle left) and double-humped (middle

right) term structures. The NS curve provides a very good fit for all but the double-humped TS. This has prompted Svensson (1995) to adjust the NS curve by adding a second curvature factor. This can be useful during some episodes in the market; however, the occurrence of a double-humped TS remains relatively infrequent.

The interpretation of the factors relates to the behaviour of their loadings. As shown in Figure 1 (bottom left), the factor loading on L_t is 1 independently of τ_j , thus an increase in L_t equally increases the yields on all maturities. It is also considered a long-term factor since $\lim_{\tau_j \rightarrow \infty} Y_t(\tau_j) = L_t$. The loading on S_t starts at 1 and decays monotonically to zero as τ_j increases. It is easily seen that $\lim_{\tau_j \rightarrow \infty} Y_t(\tau_j) - \lim_{\tau_j \rightarrow 0} Y_t(\tau_j) = -S_t$ indicating that: (i) negative values of S_t imply an upward-sloping yield curve or positive yield spread, and (ii) it is a short-term factor loading more heavily on short maturities. The loading on C_t is concave in τ_j reaching its maximum at some medium-term maturity τ^* . Thus C_t is a medium-term factor which approximates the shape, particularly the curvature, of the yield curve.

The parameter λ determines the rate of exponential decay in the factor loadings. A higher λ leads to faster decay and results in a better fit of the short end of the curve, whereas a smaller λ allows for a better fit at long maturities. λ also determines where the factor loading on C_t attains its maximum. Therefore, the choice of λ determines the ability of (2) to provide a good cross-section fit to Y_t . However, as shown in Figure 1 (bottom right), a high λ can potentially cause identification problems at longer maturities since the loadings on S_t and C_t become very similar resulting in the estimated S_t and C_t factors being almost identical up to a scaling factor. We discuss this point further in Section 2.3.

Empirically, the estimated DNS factors (L_t, S_t, C_t) tend to closely match some widely-used estimates of the level, slope and curvature of the TS. Using the yields at maturities

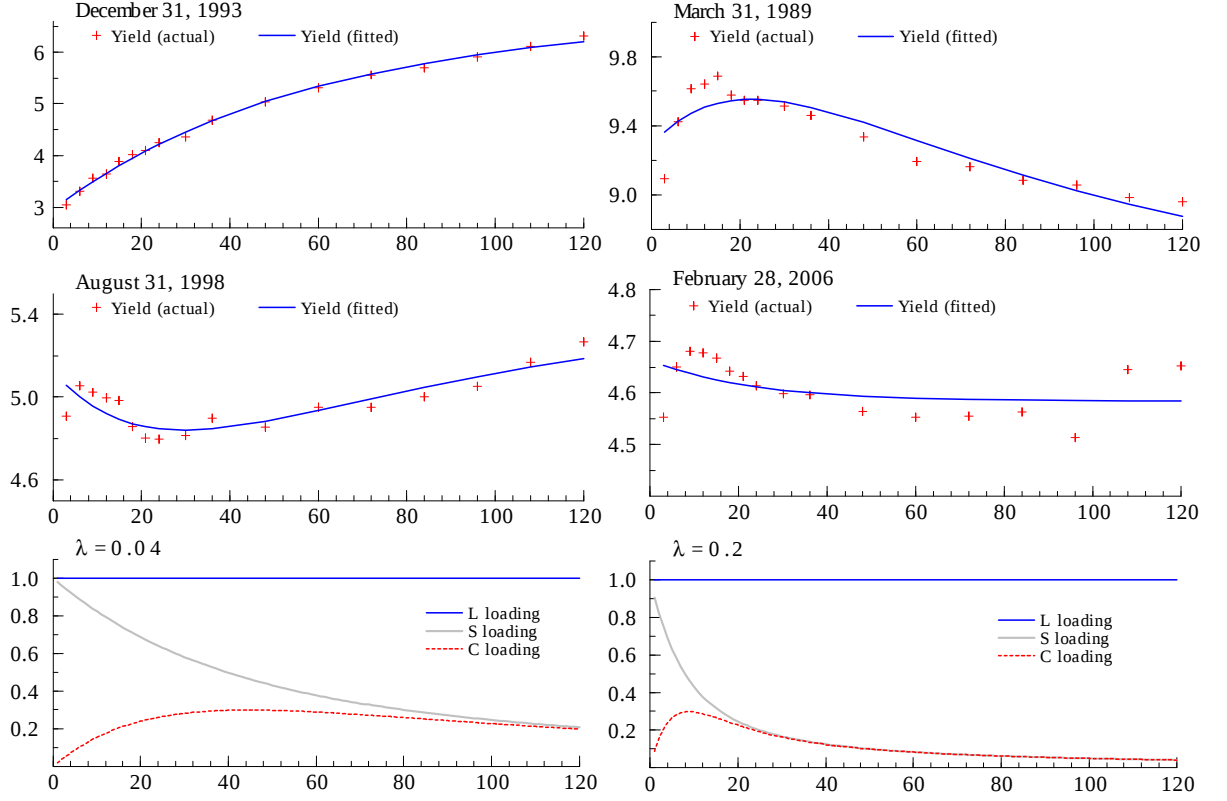


Figure 1: Top and middle panels: actual and fitted yield curves at selected dates showing (in order) an upward sloping, inverted, humped and double-humped yield curves. Bottom panel: the factor loadings at different maturities for $\lambda = 0.04$ (left) and $\lambda = 0.20$ (right).

of 3, 24 and 120 months, these are computed as $(Y_t(3) + Y_t(24) + Y_t(120))/3$ for the level, $Y_t(3) - Y_t(120)$ for (minus) the slope, and $2Y_t(24) - Y_t(3) - Y_t(120)$ for the curvature. The estimated DNS factors are also highly correlated with the first three principal components extracted from the TS, as shown in the empirical analysis.

Whether or not the DNS model is superior to other factor models of the TS is an open question. One general criticism of the DNS model is that it allows for arbitrage; however, Diebold and Li (2006) downplay the importance of this by showing that the DNS model outperforms various competing models out of sample. More recently, Christensen et al. (2009, 2011) developed a no-arbitrage version of the DNS model which is discussed in Section 2.5. Factors based on principal component (PC) analysis are not based on a

smooth yield curve function, which makes it difficult to compute yields at maturities other than the observed ones. We discuss alternative factor representations of the TS in the following subsection.

2.2 Factor Model Representation

We can write (2) as

$$Y_t = ZX_t + e_t, \tag{3}$$

which gives the familiar factor model representation with a loading matrix Z , and X_t typically has a lower dimension relative to Y_t . In our model, we adhere to this interpretation of the DNS model and consider e_t as the idiosyncratic component of yield movements, which reflects the impact of liquidity, maturity and risk preferences of bond investors; see Duffee (1996) for a related discussion.

This is a different interpretation compared to Diebold and Li (2006) and subsequent extensions of the DNS model. In these models, e_t is assumed to be measurement error which is convenient since it implies that (3) is the measurement equation in a state-space system, which is then supplemented by a transition equation for the dynamics of X_t . In contrast, we think of the TS as comprised of two components: (i) systematic variation in the observed yields captured by ZX_t which gives the smooth NS curve, and (ii) idiosyncratic variation captured by e_t . By focusing the analysis on ZX_t , we are essentially modelling and forecasting the systematic component of the TS. As we show in the empirical analysis, ZX_t explains more than 92 percent of the variation in Y_t in our data sample.

The DNS equation differs from traditional factor models since in the latter the factors are assumed to be either observable or latent, and in both cases the factor loadings are

usually unknown and need to be estimated. In the DNS model, the factor loadings are known for a given λ , while the factors are latent. As explained later, we choose λ to minimise the error in cross-section fit, which is equivalent to maximizing the systematic variation component (ZX_t) of the TS.

Of course, one can use PC analysis or factor analysis for the same objective. This can be done using the spectral decomposition of the unconditional covariance of Y_t as follows:

$$\text{Var}[Y_t] := \bar{Y} = P\Lambda P',$$

where P is a matrix of eigenvectors, and the eigenvalue matrix Λ is diagonal with non-negative elements $\lambda_1, \lambda_2, \dots, \lambda_n$. We assume that the eigenvalues in Λ are ordered such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Total variation is measured by $\text{tr}(\bar{Y})$, which is the sum of the variances of the yields at each maturity. In practice, Λ tends to have one or two dominant eigenvalues while the remaining eigenvalues tend to be very close to zero making $\bar{Y}$ an ill-conditioned matrix. This reflects the very high correlations among the yields at different maturities, and it is the reason why the first few PCs capture almost all of the variation in Y_t . The standardised PCs, denoted by Y_t^* , are computed as

$$Y_t^* = \Lambda^{-1/2} P' Y_t.$$

For dimension reduction, one usually considers the first m PCs, $m < n$, as representing

systematic yield variation.⁵ We decompose

$$Y_t^* = \begin{pmatrix} Y_{1,t}^* \\ Y_{2,t}^* \end{pmatrix}, \quad P = \begin{pmatrix} P_1 & P_2 \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix},$$

where $Y_{1,t}^*$ is $m \times 1$, $Y_{2,t}^*$ is $(n - m) \times 1$, P_1 is $n \times m$, P_2 is $n \times (n - m)$, Λ_1 is $m \times m$ and Λ_2 is $(n - m) \times (n - m)$. We can write the factor model in this case as

$$Y_t = Z^* X_t^* + e_t^*,$$

where $Z^* = P_1 \Lambda_1^{1/2}$, $X_t^* = Y_{1,t}^*$, and e_t^* represents idiosyncratic variation in yields due to the remaining $(n - m)$ PCs. This approach is adopted in, for example, Litterman and Scheinkman (1991) and Bliss (1997).

Although the DNS factors explain less of the total variation in Y_t than the first three PCs, the advantage of the DNS equation is that Z is time-invariant if we assume that λ is fixed, whereas Z^* is not since it is constructed in-sample. Another advantage of the DNS model is the underlying smooth yield curve which means that one can easily interpolate to compute yields at any maturity even if not observed in-sample, which is not possible when using PCs.

Bowsher and Meeks (2008) follow a closely related approach based on dimension reduction. They use cubic splines to estimate the systematic component in yields and model the dynamics of the resulting factors, which are the yields at the prespecified knots in their model. This allows them to model a 36-dimensional yield curve using only 5 or 6

⁵For instance, in our data sample the first three PCs explain about 99.98 percent of the variation in yields. They are also interpreted as the level, slope and curvature of the yield curve because of their factor loadings which bear surprising resemblance to the factor loadings of the DNS model. This is further discussed in Section 4.2.

factors that are assumed to be cointegrated. Although they show that their approach yields superior mean forecasts to the DNS model, they conclude that this is primarily due to better capture of the factor dynamics using a cointegrated VAR rather than improved cross-section fit due to using splines. We prefer the DNS equation due to the convenient interpretation of the DNS factors; the knot-yields in the model of Bowsher and Meeks (2008) do not have a clear interpretation. In addition, the DNS model is relatively parsimonious with only 3 factors.

2.3 Time-Varying versus Constant λ

Allowing for a time-varying λ introduces time variation in the factor loadings; however, there is no consensus in the literature as to whether λ should be fixed or left to vary. Nelson and Siegel (1987), Diebold and Li (2006) and Diebold et al. (2006), among others, assume a fixed λ . Hautsch and Ou (2008) use both fixed and time-varying λ specifications and conclude that allowing for time variation in λ does not change the main results. Koopman et al. (2010) assume a time varying λ , which is considered a fourth latent factor that allows for changes in the factor loadings, and report an improvement in goodness of fit.

A time-varying λ , as assumed in Koopman et al. (2010) and in one of the models in Hautsch and Ou (2008), implies that (3) should be written as

$$Y_t = Z_t X_t + e_t,$$

with the factor loading matrix, Z_t , now being stochastic. In our framework this extension poses two complications. First, our analysis is based on $x_t := \Delta X_t$. When λ is fixed (i.e. $Z_t = Z$), it holds that $y_t = Zx_t + \Delta e_t$ where $y_t := \Delta Y_t$. If λ is time-varying, then we have

$y_t = \Delta(Z_t X_t) + \Delta e_t$ which does not enable us to translate factor changes (x_t) into yield changes (y_t) unless we also model the law of motion of λ . This brings us to the second complication, which is the difficulty of modelling the dynamics of λ , especially that we would want λ to be restricted to a region in which the S_t and C_t factors can be identified at long maturities; see the discussion in Section 2.1 and also De Pooter (2007).

For instance, Koopman et al. (2010) estimate the DNS model assuming a time-varying λ using US Treasury yields over the period 1972:1-2000:12. They find the estimated λ to be quite volatile especially during the period prior to 1985; see their Figure 1. The values of λ , reaching as high as 0.5 around 1974, result in the loadings on S_t and C_t being almost identical at longer maturities. By fixing λ at a value that maximises the yield variation attributable to the DNS factors, we circumvent both complications.

Since we are also interested in density forecasts of the TS, a time-invariant Z implies that the density of the systematic yield component is proportional to the density of the factors. The former is given by $|\det(Z)|^{-1} f(X_t)$, where $f(X_t)$ denotes the joint density of the factors. This follows directly from the method of transformation of random variables, e.g. Casella and Berger (2001), where Z is the Jacobian of the transformation.⁶ Thus, models that achieve better density forecasts for the factors will also provide better density forecasts for the systematic component of yields.

Our approach of treating λ as fixed is also consistent with Christensen et al. (2009, 2011). In the no-arbitrage version of the DNS model developed in Christensen et al. (2009), λ is shown to be a constant controlling the rate of mean-reversion of S_t and C_t ,

⁶A technical point here is that Z is only invertible when it is a 3×3 matrix, i.e. when we include at most three yields in Y_t . Using the singular value decomposition, it is straightforward to show that if $\dim(Y_t) > 3$, then some of the yields in Y_t will have a degenerate distribution. This, however, does not affect the essence of our argument.

and also captures the effect of deviations of C_t from its mean on the mean of S_t .

2.4 Factor Dynamics

The objective is to use flexible models to characterise the dynamics of $X_t = (L_t, S_t, C_t)'$. Existing models focus on modelling the factor levels, and they can be classified as either ‘independent-factor models’ or ‘correlated-factor models’; see Christensen et al. (2011).⁷ The models of Diebold and Li (2006), Diebold et al. (2008) and Hautsch and Ou (2008) are ‘independent-factor models’. Diebold and Li (2006) assume that the factors follow the VAR(1) process

$$X_t = \Phi_{X,0} + \Phi_{X,1}X_{t-1} + \Omega^{1/2}\eta_t, \quad \eta_t \stackrel{i.i.d.}{\sim} N(0, I), \quad (4)$$

where $\Phi_{X,1}$ is diagonal and the conditional covariance matrix of X_t is given by Ω , which is assumed to be diagonal. Diebold et al. (2008) consider a two-factor DNS model, where the factor dynamics are given by (4) but X_t includes only L_t and S_t . Hautsch and Ou (2008) assume that Ω is time varying with the following stochastic volatility specification

$$vech(\log \Omega_t) = \Phi_{\Omega,0} + \Phi_{\Omega,1}vech(\log \Omega_{t-1}) + \zeta_t, \quad \zeta_t \stackrel{i.i.d.}{\sim} N(0, \Upsilon). \quad (5)$$

They assume that $\Phi_{X,1}$ and $\Phi_{\Omega,1}$ are diagonal, and (5) now allows for stochastic volatility in each factor.

Among the DNS ‘correlated-factor models’, Diebold et al. (2006) introduce correlations among the factors by assuming $\Phi_{X,1}$ is unrestricted and $\Omega^{1/2}$ lower triangular. Koopman et al. (2010) extend the model of Diebold et al. (2006) by: (i) allowing for a time-varying

⁷This classification also applies to the class of affine TS models; see Dai and Singleton (2000).

λ as a fourth latent factor, and (ii) introducing time-varying volatility in the yields via the idiosyncratic component, $e_t(\tau_j)$, and also through the factors themselves as in Hautsch and Ou (2008).

Diebold and Li (2006) estimate the model in 2 steps, first estimating the factors by cross-section regressions in (2) holding λ fixed, then modelling the factor dynamics in a second step. The other models discussed above are set up as a state-space system. Combining (2) and (4), creates a state-space system in which (2) is the measurement equation and (4) is the transition equation. The parameters are estimated using the Kalman filter, with the exception of Koopman et al. (2010) which is estimated by a modified Kalman filter since the transition equation in their model is nonlinear in the state variables as λ is assumed to be a fourth state variable. The model of Hautsch and Ou (2008) is estimated by Monte Carlo Markov Chain methods.

2.5 The Arbitrage-Free Dynamic Nelson-Siegel Model

As indicated earlier, the DNS model admits arbitrage since the model does not provide a perfect fit to the cross section of yields. Empirically, the fitted values will differ from the observed yields, $Y_t(\tau_j)$, in (2) by the idiosyncratic component $e_t(\tau_j)$. Christensen et al. (2009, 2011) develop an arbitrage-free DNS (AF-DNS) model given by

$$Y_t(\tau_j) = L_t + \left(\frac{1 - e^{-\lambda\tau_j}}{\lambda\tau_j} \right) S_t + \left(\frac{1 - e^{-\lambda\tau_j}}{\lambda\tau_j} - e^{-\lambda\tau_j} \right) C_t - \frac{\Gamma(\tau_j)}{\tau_j}, \quad (6)$$

where $\Gamma(\tau_j)$ is a maturity-specific adjustment term that allows the model to perfectly fit the cross-section of observed yields. Despite the attractiveness of the no-arbitrage feature of the AF-DNS model, one limitation is that $\Gamma(\tau_j)$ also depends on Ω , or its stochastic

volatility specification if it is assumed to be time-varying. This means that only system estimation is possible in this framework where (6) and (4), and possibly (5), form a state-space system.

The Kalman filter can be used for estimation in this case; however, the results in Christensen et al. (2010) show that it is quite challenging to accurately estimate the stochastic volatility parameters given the large standard errors they report; see their Tables 6 and 18. In addition, the magnitude of some of the parameter estimates strongly suggests potential convergence problems in estimation.

3 A Dynamic Nelson-Siegel Model for Changes in the Term Structure

In this section, we discuss our modelling approach. The reason we model and forecast the factor changes rather than the factor levels is simply to boost immunity to structural breaks. It is well-known that most forecast failures happen due to changes in the unconditional moments of time series (Clements and Hendry, 2006). By modelling the factor changes, our model is potentially less vulnerable to structural breaks since the moments of factor changes seem more stable compared to the moments of the factor levels. For instance, both L_t and S_t are found to have bimodal unconditional distributions. In the case of L_t , this is primarily due to the general moderation of interest rate levels during the last two decades. It is perhaps surprising that existing models of the TS have thus far paid little attention to this issue.

In addition, we allow for departures from the assumption of multivariate normality typically adopted in the literature, and also allow the dependence structure of the factor

changes to be time varying as well as permit joint extreme dependence, where the latter is unattainable under multivariate normality. Through statistical testing we are able to discern the empirical plausibility of these assumptions. Thorough analysis of model misspecification is also undertaken. This is also one of the features which distinguish this paper from most existing studies of the TS.

From a practical viewpoint, it is acceptable to have a misspecified model that does well in forecasting. The random walk model, which is a widely used and difficult-to-beat benchmark for many time series, is a case in point. As shown in the empirical analysis, our model fares rather well in both point and directional forecasting, in addition to passing conventional misspecification tests.

3.1 Estimating the DNS Factors

We estimate the DNS factors using (2). We keep λ fixed (see the discussion in Section 2.3) but we choose it optimally to maximise the systematic variation in yields due to the DNS factors, which is equivalent to minimizing the cross-section fit error in (2) at each point in time. However, we restrict λ to a range of values where L_t and S_t can be identified at long maturities.

We perform a grid search over $\lambda \in [0.03, 0.15]$ to find the value of λ that minimises error in the cross-section fit, given by $e_t' e_t$, over the entire sample. The lower and upper bounds of the grid correspond to the curvature factor attaining its maximum at 12 and 60 months, respectively. This region for λ ensures that the slope and curvature factors can be identified at longer maturities; see De Pooter (2007) for a related discussion. Next, we focus on modelling the joint density of $x_t := \Delta X_t$.

As mentioned earlier, most of the existing models focus on the factor levels. There

are a few exceptions which are worth mentioning at this point. Bliss (1997) undertakes factor analysis of the changes in the US Treasury yields for the period 1970:1-1995:12. Remarkably, he finds that the estimated factors correspond to changes in the level, slope and curvature of the TS. These three factors are able to explain most of the variation in the changes of the TS despite significant structural changes in the TS dynamics over this period. Audrino and Trojani (2007) use nonparametric methods to estimate the conditional mean and conditional variance of changes in the yields, and then use multivariate GARCH processes to model the time-varying correlations of the standardised factor changes.

Fabozzi et al. (2005) use regressions to assess whether the DNS model factor changes are forecastable using various economic and financial indicators. Throughout, they assume that the factor changes are independent. Bowsher and Meeks (2008) point to the potential non-stationarity in the TS and use a cointegrated VAR model for the factor dynamics, where the factors in their model are estimated using spline methods.

3.2 Modelling the Joint Density of the Factor Changes

We use the copula-margins decomposition to model the joint density of the factors; see Appendix A for an overview of copula theory. Thus we have

$$f(x) = \prod_{i=1}^3 f_i(x_i) \cdot c(F_1(x_1), F_2(x_2), F_3(x_3)), \quad (7)$$

where $x_t = (x_{1,t}, x_{2,t}, x_{3,t})'$, $x_{1,t} = \Delta L_t$, $x_{2,t} = \Delta S_t$ and $x_{3,t} = \Delta C_t$. $f(\cdot)$ denotes the joint density of the random vector x_t , F_i and f_i denote, respectively, the *cdf* and *pdf* of x_i . Throughout we use $i = 1, 2, 3$, to index the factor levels ($X_{i,t}$) or factor changes ($x_{i,t}$). The copula density is denoted by $c(\cdot)$. The joint, marginal and copula densities

are all conditional on the information set at time $t - 1$, denoted by $\mathcal{F}_{t-1}$, which includes past observations of x_t . The joint density is parameterised by the parameter vector $\theta = (\psi'_1, \psi'_2, \psi'_3, \kappa')'$, where ψ_i denotes the parameter vector for the i -th marginal density, and κ denotes the copula parameter vector. The dimension of θ depends on the dimensions of the constituting vectors, which in turn depend on the assumed marginal and copula models. We assume that the functional form of the copula is time-invariant, but we allow the copula parameter (or dependence measure) to be potentially time-varying.

For the marginal models, we assume that $x_{i,t}$ follows a t -GARCH(1,1) process given by

$$x_{i,t} = \mu_{i,t} + \sigma_{i,t}\varepsilon_{i,t}, \quad (8)$$

$$\mu_t := (\mu_{1,t}, \mu_{2,t}, \mu_{3,t})' = \Phi_0 + \sum_{j=1}^J \Phi_j x_{t-j}, \quad (9)$$

$$\sigma_{i,t}^2 = (1 - \alpha_i - \beta_i)\bar{\sigma}_i + \alpha_i(\sigma_{i,t-1}\varepsilon_{i,t-1})^2 + \beta_i\sigma_{i,t-1}^2, \quad (10)$$

$$\varepsilon_{i,t} \stackrel{i.i.d.}{\sim} F_i(0, 1, \nu_i). \quad (11)$$

The marginal model in (8)-(11) defines a t -GARCH(1,1) process for $x_{i,t}$ where $\mu_{i,t}$ is the conditional mean, $\sigma_{i,t}^2$ is the conditional variance and $\varepsilon_{i,t}$ is an *i.i.d.* standardised innovation. $\varepsilon_{i,t}$ is assumed to be orthogonal to the idiosyncratic components, $e_t(\tau_j)$, in (2). We assume that F_i is given by the standardised Student's t distribution, where ν_i denotes the degrees of freedom (d.o.f.) parameter.

The following parameter restrictions are assumed to hold: $\bar{\sigma}_i > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$, and $\alpha_i + \beta_i < 1$. When $\alpha_i = 0$, β_i is unidentified and is set equal to zero, which implies that $x_{i,t}$ is conditionally homoskedastic in this case. For the second moment of $\varepsilon_{i,t}$ to exist, $\nu_i > 2$ must hold. The t -GARCH process given by (8)-(11) is chosen for its flexibility as

it nests the cases of a constant variance if $\alpha_i = \beta_i = 0$, and also the normal distribution when $\nu_i \rightarrow \infty$. Our specification of the marginal models is closely related to Jondeau and Rockinger (2006), Patton (2006b), and Rodriguez (2007); however, the applications in these studies are not related to TS models.

In the specification of the conditional mean vector, μ_t , we allow it to depend on past values of x_t and the lag structure in (9) is chosen to ensure that all serial correlation in the residuals of each equation, $\sigma_{i,t}\varepsilon_{i,t}$, is adequately captured. This is important to ensure correct specification of each marginal model; otherwise misspecification of any of the marginal densities will extend to the copula. This occurs since F_i is used to obtain the estimated probability integral transforms, $\hat{u}_{i,t} = F_i(x_{i,t}; \hat{\psi}_i)$, which enter the copula. Patton (2006b) extends the theorem of Sklar (1959) - see Appendix A - to conditional distributions and conditional copulas. He shows that partitioned information sets for the margins are valid under certain conditions; however in our model, we do not attempt any partitioning and include the same lags for all variables.

The copula specification, discussed in the following subsection, permits flexibility in modelling the joint density of x_t . It enables us to create multivariate densities which are richer than the multivariate normal distribution typically assumed in the literature; see Appendix A for details. This is the case when, for instance, mixing Student's t margins with the normal copula, which results in a *meta* normal distribution using the terminology of Demarta and McNeil (2005). Using the t copula gives a *meta t* distribution which allows for joint extreme dependence (or tail events) between the factor changes. To our knowledge, this is one of the first applications of copula theory to TS models with the objective of increasing the flexibility of modelling the factor dynamics.

3.3 A Time-Varying Copula Model

For the choice of an appropriate copula, we focus on the normal and t copulas. The advantage these copulas offer in the trivariate and higher dimensions is that the dependence structure is summarised in a matrix Σ , where its off-diagonal elements are pairwise dependence measures; the normal and t copulas are introduced in (A.2) and (A.4) in Appendix A. Many of the commonly-used copulas in finance, e.g. the Clayton and Gumbel copulas, have only one dependence parameter which is difficult to interpret in dimensions greater than two.

For our chosen copulas, the copula density in (7) is given by (A.3) for the normal copula, and (A.5) for the t copula. To parameterise time variation in the dependence parameter, we focus on the covariance matrix implied by Σ_t , which we denote by Q_t . This decomposition is popular in the multivariate GARCH literature, and is based on the dynamic conditional correlations (DCC) model of Engle (2002). So we have

$$Q_t = (1 - \delta)\bar{Q} + \delta Q_{t-1} + a(V_{t-1}V'_{t-1} - Q_{t-1}), \quad (12)$$

$$\Sigma_t = (Q_t \circ I)^{-\frac{1}{2}} Q_t (Q_t \circ I)^{-\frac{1}{2}},$$

where $\circ$ denotes the Hadamard elementwise product, $0 < a < 1$ and $0 < \delta < 1$ are scalar parameters satisfying $\delta > a$. This is a reparameterisation of the scalar DCC model of Engle (2002) expressing the conditional covariance dynamic equation in an ARMA(1,1)-type representation, where δ measures the covariance/correlation persistence and a is the loading on the innovations, $V_{t-1}V'_{t-1} - Q_{t-1}$. Note that by construction $E[V_t V'_t | \mathcal{F}_{t-1}] = Q_t$, so $V_{t-1}V'_{t-1} - Q_{t-1}$ is a vector martingale difference sequence. This parameterisation also

retains the correlation targeting property by setting $\overline{Q} = E[V_t V_t']$, and the restriction $0 < \delta < 1$ ensures both covariance stationarity in (12) and positive semidefiniteness of the target $(1-\delta)\overline{Q}$.⁸ For the normal copula $V_{t-1} = (\Phi^{-1}(u_{1,t-1}), \Phi^{-1}(u_{2,t-1}), \Phi^{-1}(u_{3,t-1}))'$, and for the t copula $V_{t-1} = (t_{v_1}^{-1}(u_{1,t-1}), t_{v_2}^{-1}(u_{2,t-1}), t_{v_3}^{-1}(u_{3,t-1}))'$, where we use the notation in Appendix A.

This is one of the first papers to explicitly model time-variation in the dependence structure of the factors. Christensen et al. (2010) introduce restricted forms of multivariate stochastic volatility similar to the square-root diffusion of Cox et al. (1985). Their model assumes that volatility and co-volatility in the factors depend on the levels of the factors. Our specification is more general in that it allows Σ_t to be filtered in-sample while matching the unconditional correlations among the standardised factor changes. Dai and Singleton (2000) argue that capturing the factor correlations is crucial to effectively characterise the TS dynamics. In their affine TS models' framework, they had to impose restrictive assumptions on the correlation structure such as ruling out positive correlations among the factors. Our approach generalises this by removing any restrictions on the factors' dependence structure, in addition to allowing the dependence structure to be potentially time-varying.

3.4 Estimation and Inference

The state-space system given by (2) and (7) can in principle be estimated with a Kalman filter since the measurement equation is linear in the latent factors (for a fixed λ), but with our proposed model for the factor dynamics, this poses a considerable challenge in

⁸Covariance stationarity in (12) follows directly from the results of Engle and Kroner (1995) on BEKK models.

estimation due to the nonlinearities in the joint density. In the spirit of Diebold and Li (2006), we overcome this computational challenge by first estimating the factors for a fixed λ , and then modelling the joint density of x_t in (7).

From (7), the t -th observation joint log-likelihood can be written as

$$l_{x,t}(\theta; x) = \sum_{i=1}^3 l_{x_i,t}(\psi_i; x_i) + l_{c,t}(\kappa; u), \quad u = (u_1, u_2, u_3), \quad (13)$$

where $u_i = F_i(x_i; \psi_i)$, $i = 1, 2, 3$. The t -th observation log-likelihood of the joint, i -th margin and copula are respectively denoted by $l_{x,t}(\theta; x)$, $l_{x_i,t}(\psi_i; x_i)$ and $l_{c,t}(\kappa; u)$. The parameters of the joint density are collected in the vector $\theta = (\psi'_1, \psi'_2, \psi'_3, \kappa')'$. The dimension of the parameter vector for each margin's log-likelihood, $l_{x_i,t}(\psi_i; x_i)$, depends on the number of lags in the conditional mean specification in (9). There are 4 parameters in the t -GARCH model for the conditional variance: 3 parameters in (10) and the d.o.f. parameter for F_i in (11). The dimension of κ depends on the chosen copula: for the time-invariant normal and t copulas, κ contains 3 and 4 parameters, respectively; see Appendix A for details. For the time-varying copula specification in (12), κ contains 8 parameters in the normal copula model (6 parameters in $\bar{Q}$ in addition to a and δ), while the t copula model has 9 parameters, the additional parameter being the d.o.f. of the t copula.

The joint log-likelihood in (13) can be estimated by maximum likelihood (ML) in two steps. In the first step, ψ_i , $i = 1, 2, 3$, is estimated and the probability integral transforms obtained as $\hat{u}_i = F_i(x_i; \hat{\psi}_i)$. In the second step, κ is estimated by using $l_{c,t}(\kappa; \hat{u})$, $\hat{u} = (\hat{u}_1, \hat{u}_2, \hat{u}_3)$. This results in consistent estimation of the model parameters, but entails a loss in efficiency compared to the one-step ML estimator. Despite the loss in efficiency, this approach is more tractable and considerably simplifies estimation. Inference in this case

is a direct application of two-step GMM discussed in, for example, Newey and McFadden (1994).

Let $\widehat{\psi}_i$ and $\widehat{\kappa}$ be the ML estimators for the i -th margin and copula densities, respectively, and let $\widehat{\theta} = (\widehat{\psi}_1', \widehat{\psi}_2', \widehat{\psi}_3', \kappa')'$. Patton (2006a) shows that

$$\sqrt{T}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1} \mathcal{J} (\mathcal{I}^{-1})'),$$

where

$$\mathcal{J} = \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial l_t}{\partial \theta'} \right],$$

$$\mathcal{I} = \text{E} \begin{bmatrix} \frac{\partial^2 l_{x_1,t}}{\partial \psi_1 \partial \psi_1'} & 0 & 0 & 0 \\ 0 & \frac{\partial^2 l_{x_2,t}}{\partial \psi_2 \partial \psi_2'} & 0 & 0 \\ 0 & 0 & \frac{\partial^2 l_{x_3,t}}{\partial \psi_3 \partial \psi_3'} & 0 \\ \frac{\partial^2 l_{c,t}}{\partial \psi_1 \partial \kappa'} & \frac{\partial^2 l_{c,t}}{\partial \psi_2 \partial \kappa'} & \frac{\partial^2 l_{c,t}}{\partial \psi_3 \partial \kappa'} & \frac{\partial^2 l_{c,t}}{\partial \kappa \partial \kappa'} \end{bmatrix}.$$

The lack of dependence between the parameters of each margin and the parameters of the other margins is well motivated by the copula modelling approach which stipulates that each margin only contains univariate information. In estimating κ , we account for the accumulation of estimation error from the first step of estimation.

4 Empirical Analysis

4.1 Data Set

The data set of nominal yields used in estimation is for US Treasury bills and bonds obtained from the Center for Research in Security Prices (CRSP). End-of-month yields and

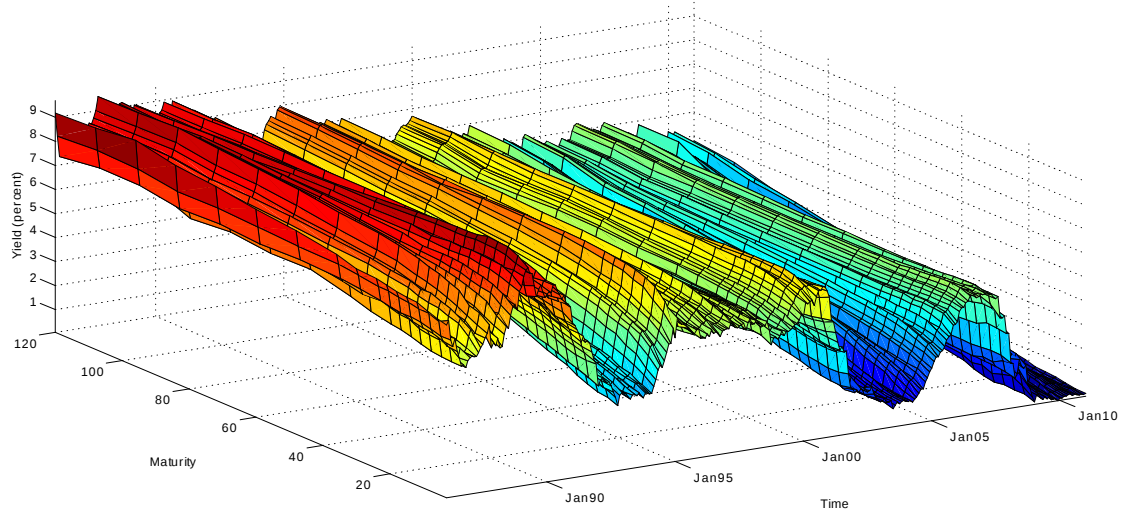


Figure 2: 3D plot of the US Treasuries term structure. Sample period: 1986:3-2010:12.

time to maturity are obtained for the period 1986:3 to 2010:12.⁹ At any given date, there are numerous Treasury bills and bonds trading in the market with different maturities. Following Diebold and Li (2006), we fix the maturities at 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, where one month is equivalent to 30.4375 days. Then, we linearly interpolate the yields at nearby maturities to compute the yields at the desired maturities.

Figure 2 presents a 3D plot of the yield curve. It shows an overall moderation in the level of the TS during the second part of the sample, as well as substantial variation in the slope with the short end of the curve closely tracking the interest rate policy cycles during the recessions and expansions starting the 1990s.¹⁰ Starting 2009 the short end of the curve is effectively pinned down at a level close to zero, a feature that is unprecedented

⁹We use the yields reported directly in CRSP which are based on price quotes for US Treasuries. We do not use any filters, e.g. Fama and Bliss (1987), to compute zero-coupon yields for synthetic bonds. Thus potential coupon effects will be included in the idiosyncratic yield components.

¹⁰During this period, the Federal Reserve pursued an activist monetary policy to smooth output fluctuations, with the Federal Reserve Rate (on overnight borrowing) becoming the main policy instrument.

Maturity	Mean	St. Dev.	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
3	4.186	2.260	0.055	9.095	0.996	0.830	0.450
6	4.319	2.282	0.171	9.423	0.998	0.835	0.462
9	4.433	2.293	0.224	9.615	0.997	0.840	0.483
12	4.526	2.300	0.253	9.642	0.997	0.850	0.515
15	4.600	2.301	0.272	9.689	0.996	0.860	0.552
18	4.678	2.283	0.256	9.578	0.996	0.868	0.586
21	4.735	2.247	0.314	9.548	0.995	0.873	0.612
24	4.808	2.225	0.333	9.548	0.994	0.879	0.642
30	4.944	2.179	0.397	9.515	0.994	0.890	0.689
36	5.083	2.129	0.534	9.460	0.993	0.900	0.736
48	5.288	2.033	0.932	9.334	0.992	0.906	0.793
60	5.483	1.950	1.296	9.243	0.991	0.912	0.834
72	5.646	1.874	1.613	9.402	0.990	0.914	0.858
84	5.752	1.797	2.074	9.525	0.990	0.914	0.863
96	5.891	1.807	2.165	9.684	0.989	0.911	0.878
108	6.094	1.634	2.716	9.610	0.988	0.911	0.885
120	6.180	1.594	2.879	9.582	0.988	0.923	0.907

Table 1: Summary statistics for the yields at different maturities. Full sample: 1986:3-2010:12. $\hat{\rho}(k)$ is the autocorrelation coefficient at lag k .

in recent history. The summary statistics reported in Table 1 indicate that, on average, the yield curve is upward sloping, and volatility tends to be higher at short maturities. There is also strong persistence in the yields as indicated by the autocorrelation coefficient $\hat{\rho}(k)$, for lags $k = 1, 12, 24$ months. Yields at the long end are more persistent relative to yields at the short end.

4.2 The DNS Model Factors

The optimal value of λ which maximises the yield variation due to the DNS factors is 0.040, which we use to construct the factor loadings. This value of λ implies that the curvature factor is maximised at $\tau_j = 48$ months. It is worth noting that Diebold and Li (2006) fix λ at 0.0609 which maximises the curvature factor loading at $\tau_j = 30$ months. We then estimate the factors from the cross-section by least squares optimisation. The mean of the cross-section residuals, $\hat{e}_t(\tau_j)$, in (2) vary by maturity from a minimum of 0.001 percent at $\tau_j = 30$ to a maximum of 0.063 percent at $\tau_j = 84$. Overall the estimated yield curve fits

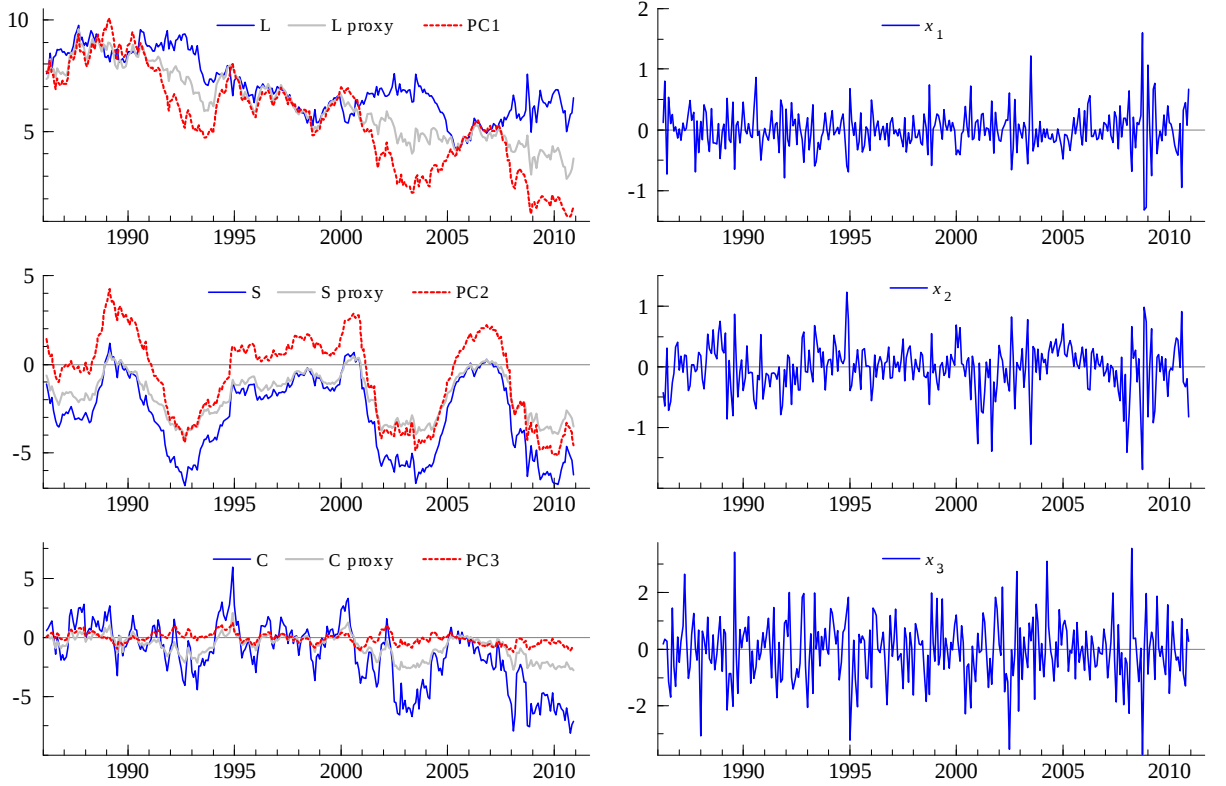


Figure 3: Left panel: estimated DNS factors (L_t, S_t, C_t), their empirical proxies and corresponding principal components (PCs). In the top left panel, the first PC is scaled down to match the means of L_t and its proxy. Right panel: first differences of the DNS model factors ($x_{1,t}, x_{2,t}, x_{3,t}$).

the cross section of yields quite well especially at intermediate maturities. Moving towards the short and long ends of the curve, the fit tends to deteriorate. A similar finding is reported in Diebold and Li (2006) and Koopman et al. (2010).

The estimated DNS factors are presented in Figure 3 along with their empirical proxies (see Section 2.1) and the first three PCs of the yields (see Section 2.2). The DNS level factor, L_t , generally moves in tandem with its empirical proxy and the first PC. They diverge in periods when S_t is strongly negative - indicative of a steep yield curve - which also coincide with the periods when C_t is strongly negative.

The slope factor, S_t , tracks well the cyclical movements in its empirical proxy and the

		Mean	St. Dev.	Skewness	Kurtosis	Min	Max	$\hat{\rho}(1)$	$\hat{\rho}(12)$	$\hat{\rho}(24)$
L	FS	6.984	1.351	0.196	2.058	4.205	9.760	0.968	0.823	0.687
	SS1	7.961	1.069	-0.459	2.159	5.373	9.760	0.977	0.916	0.861
	SS2	5.994	0.754	-0.319	2.384	4.205	7.585	0.871	0.412	0.005
S	FS	-2.823	2.176	-0.222	1.740	-6.861	1.179	0.986	0.592	-0.094
	SS1	-2.459	1.785	-0.502	2.558	-6.861	1.179	0.983	0.569	-0.126
	SS2	-3.192	2.463	0.099	1.402	-6.789	0.684	0.984	0.570	-0.212
C	FS	-1.415	2.631	-0.512	2.727	-8.129	5.941	0.919	0.516	0.365
	SS1	-0.021	1.667	0.052	3.425	-4.402	5.941	0.796	-0.063	-0.054
	SS2	-2.828	2.679	-0.018	2.082	-8.129	3.287	0.920	0.448	0.083
x_1	FS	-0.004	0.340	0.259	6.062	-1.315	1.606	-0.172	0.050	0.043
	SS1	-0.013	0.294	0.034	3.495	-0.790	0.858	-0.186	0.060	0.147
	SS2	0.006	0.381	0.325	6.448	-1.315	1.606	-0.163	0.036	-0.036
x_2	FS	-0.017	0.405	-0.621	4.761	-1.692	1.225	0.161	-0.005	0.025
	SS1	0.004	0.349	0.123	3.633	-0.861	1.225	0.193	0.004	0.042
	SS2	-0.039	0.455	-0.882	4.584	-1.692	0.980	0.138	-0.013	0.013
x_3	FS	-0.026	1.095	-0.048	3.945	-3.728	3.559	-0.035	0.151	0.035
	SS1	-0.016	1.074	-0.140	3.469	-3.218	3.413	0.004	0.131	0.077
	SS2	-0.036	1.118	0.037	4.349	-3.728	3.559	-0.074	0.152	-0.001

		L	S		x_1	x_2
FS	S	-0.231		x_2	-0.781	
SS1		-0.441			-0.724	
SS2		-0.674			-0.815	
FS	C	0.250	0.593	x_3	-0.427	0.336
SS1		-0.091	0.388		-0.401	0.408
SS2		-0.432	0.708		-0.451	0.286

Table 2: Top panel: summary statistics for the DNS model factors (L_t, S_t, C_t) and their first differences ($x_{1,t}, x_{2,t}, x_{3,t}$). Full sample is 1986:3-2010:12, subsample 1 is 1986:3-1998:8 and subsample 2 is 1998:9-2010:12. $\hat{\rho}(k)$ is the autocorrelation coefficient at lag k . Bottom left panel: unconditional correlations among (L_t, S_t, C_t). Bottom right panel: unconditional correlations among ($x_{1,t}, x_{2,t}, x_{3,t}$).

second PC, with the differences among the three variables being one of scale. As indicated earlier, these cycles also match very closely the movement of the Federal Reserve Rate (FFR) during the same period which greatly influences the short end of the curve, and overall S_t seems to inherit some persistence from these policy cycles. The curvature factor, C_t , is the most volatile tracking on average the movements in its empirical proxy and the third PC but showing more volatility. The extreme negative realisations of C_t (e.g. during 2002-2003 and 2008-2010) occur when the TS is rather steep.¹¹

Table 2 presents the summary statistics for the factor levels and the factor changes. We report the results for the full sample (FS) which is 1986:3-2010:12, the first subsam-

¹¹Christensen et al. (2011) interpret C_t as a time-varying mean for S_t .

ple (SS1) covering the period 1986:3-1998:8 and the second subsample (SS2) covering the period 1998:9-2010:12. In the FS, the mean level for the yield curve is 6.984 percent with an average spread of 2.823 percent and curvature equal to -1.415. The level factor is the least volatile and most persistent, while the curvature factor has the highest volatility and a relatively lower level of persistence. Their unconditional distributions (not reported for brevity) are bimodal, partly skewed and platykurtic. First differencing re-centers the series at about zero and expectedly generates less persistent time series.¹² The unconditional distributions of the factor changes are unimodal, marginally skewed and have excess kurtosis. The Kolmogorov-Smirnov test rejects normality in $(x_{1,t}, x_{2,t}, x_{3,t})$ at 5 percent. Subsample statistics indicate that the unconditional moments of the factor changes are generally more stable relative to the factor levels.

The FS unconditional correlations reported in the bottom panel of Table 2 reveal an interesting regularity that is rarely encountered with time series data. Typically the correlations between first-differenced time series tend to decline; however, for the DNS factors first differencing substantially increases the (negative) $x_{1,t}/x_{2,t}$ correlation, and also the $x_{1,t}/x_{3,t}$ correlation which switches sign. The $x_{2,t}/x_{3,t}$ correlation is lower relative to the S_t/C_t correlation. Equally important is the fact that the correlations among $(x_{1,t}, x_{2,t}, x_{3,t})$ are more stable across the two subsamples relative to the correlations among (L_t, S_t, C_t) .

We conclude this section by discussing Figure 4 which presents some interesting features of the estimated factors that are worth highlighting. The top left chart shows the mean of the actual yields at each maturity and the NS curve plotted using the mean values of L_t , S_t and C_t . This is indicative of the average cross-section fit which is very good particularly at

¹²Using the Augmented Dickey-Fuller unit root test with a constant and 12 lags, we do not reject H_0 of a unit root in (L_t, S_t, C_t) , but reject H_0 for $(x_{1,t}, x_{2,t}, x_{3,t})$.

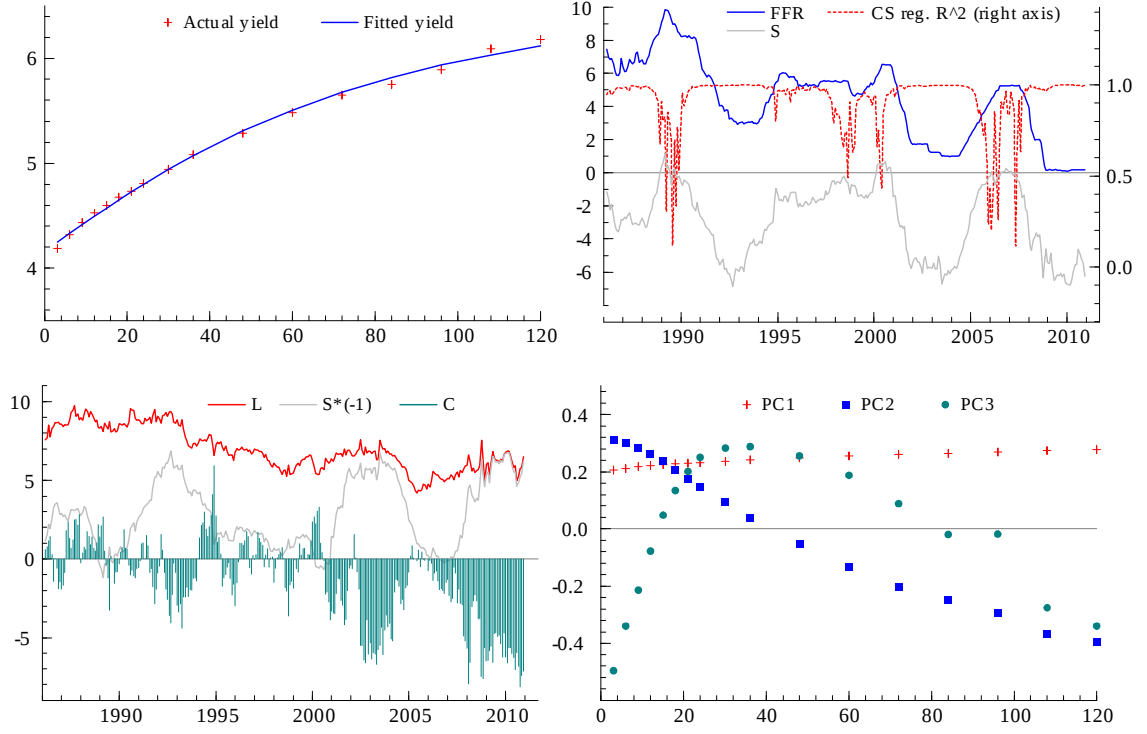


Figure 4: Top left: the mean of the actual yields for each maturity and the Nelson-Siegel curve implied by the mean values of L_t , S_t and C_t . Top right: time series of R^2 from the cross-section regressions, the Federal Funds Rate (FFR) and S_t . Bottom left: time series of L_t , (minus) S_t and C_t . Bottom right: the loadings of the first three principal components at each maturity.

medium-term maturities. The top right chart shows the time series of S_t , the FFR and the R^2 from the cross-section regressions. Interestingly, the NS curve provides a near-perfect fit to the observed yields at most of the sample points except during the periods where the TS is relatively flat. Equally interesting is that the periods where the fit worsens also coincide with periods of post-adjustment of the TS to contractionary policy by the Fed as indicated by movements in the FFR. The close association between S_t and the FFR over the whole sample period is clearly evident.

In the bottom left chart, we plot L_t , (minus) S_t and C_t to show another interesting feature that is peculiar to the recent behaviour of the short end of the yield curve. With short maturities effectively pinned down close to zero since 2009, any movement in yields

at longer maturities changes both the level and slope of the yield curve in an almost identical manner. In other words, the two factors become almost indistinguishable in this situation if we ignore the sign of S_t . This is immediately obvious from (1) since $\lim_{\tau_j \rightarrow 0} Y_t^{NS}(\tau_j) = L_t + S_t$, and if $Y_t^{NS}(\tau_j = 0) = 0$, then $L_t \approx -S_t$.¹³ During this period, C_t has also become strongly negative and is well below its unconditional mean. Later, we report the time-varying copula model results, where we find the $x_{1,t}/x_{2,t}$ conditional dependence measure approaching -1 towards the end of the sample period, a finding that is consistent with Figure 4.

Finally, the bottom right chart plots the factor loadings of the first three PCs. They show remarkable resemblance to the factor loadings of the NS curve despite the lack of normalisation. The loadings on PC3 (interpreted as curvature) attains a maximum at $\tau_j = 36$, while we find the optimal λ for the DNS factors to imply that the loading on C_t is maximised at $\tau_j = 48$.

4.3 In-Sample Analysis

Table 3 shows the estimation results for the marginal models. Two lags of x_t seem sufficient to filter out the conditional mean dynamics resulting in serially uncorrelated residuals. For $x_{1,t}$, its own first lag and both lags of $x_{2,t}$ are statistically significant at 1 percent. For $x_{2,t}$, its own first lag as well both lags of $x_{1,t}$ are statistically significant at 1 percent, while the second lag of $x_{3,t}$ is statistically significant at 10 percent. For $x_{3,t}$, the first lag of $x_{1,t}$ and both lags of $x_{2,t}$ are statistically significant at 1 percent, while its own second lag and the second lag of $x_{1,t}$ are statistically significant at 10 percent. In the two subsamples, the

¹³This holds only approximately since the estimated L_t and S_t will also be fitting the yields at intermediate and long maturities.

parameter estimates and their statistical significance do not show large changes, with the exception of the lags of $x_{1,t}$ whose influence on the conditional means of $x_{2,t}$ and $x_{3,t}$ shows rather large differences between the two subsamples.

With regard to the t -GARCH(1,1) model estimates, in all three cases, the estimate of the ARCH term (α_i) is statistically insignificant. When $\alpha_i = 0$, β_i becomes unidentified and should be set equal to zero. One exception is $x_{1,t}$ in SS1, for which α_i is statistically significant at 1 percent, while β_i is not statistically significant. Despite this exception, there is enough evidence to conclude that conditional homoskedasticity is a reasonable assumption for the factor changes. As for the estimates of the d.o.f. parameter for each margin, the large standard errors suggest that it is rather difficult to estimate this parameter given the limited number of observations used in estimation. However, the results suggest some ‘fattening’ of the tails going from SS1 to SS2. In SS1, the estimate of ν_i in the cases of $x_{1,t}$ and $x_{3,t}$ makes it virtually indistinguishable from the normal density. In SS2, the estimates of ν_i drop substantially in all three margins.

We also tried fitting the GARCH(1,1) model to $(x_{1,t}, x_{2,t}, x_{3,t})$ assuming a normal density for F_i . The likelihood ratio (LR) test can be used in this case since the Student’s t distribution nests the normal as $\nu_i \rightarrow \infty$. The test results indicate that the Student’s t density provides a statistically significant improvement in the likelihood compared to the normal in all cases in the FS and the two subsamples. This is perhaps not too surprising since the excess kurtosis in the unconditional distributions of $(x_{1,t}, x_{2,t}, x_{3,t})$ is evident, especially in SS2, as shown in Table 2.

Full sample: 1986:4-2010:12				Subsample 1: 1986:4-1998:8			Subsample 2: 1998:9-2010:12		
	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$	$x_{1,t}$	$x_{2,t}$	$x_{3,t}$
Constant	-0.013 (0.018)	0.022 (0.002)	-0.029 (0.062)	-0.024 (0.022)	0.017 (0.027)	-0.006 (0.084)	-0.005 (0.028)	-0.003 (0.035)	-0.041 (0.094)
$x_{1,t-1}$	-0.529 [†] (0.090)	0.651 [†] (0.110)	1.093 [†] (0.311)	-0.491 [†] (0.114)	0.593 [†] (0.136)	1.449 [†] (0.423)	-0.533 [†] (0.145)	0.676 [†] (0.180)	0.581 (0.483)
$x_{1,t-2}$	-0.095 (0.093)	0.301 [†] (0.113)	-0.569 [*] (0.321)	0.012 (0.120)	0.110 (0.144)	-0.279 (0.448)	-0.164 (0.145)	0.533 [†] (0.180)	-0.760 (0.482)
$x_{2,t-1}$	-0.416 [†] (0.074)	0.537 [†] (0.090)	0.909 [†] (0.256)	-0.383 [†] (0.099)	0.523 [†] (0.118)	1.070 [†] (0.368)	-0.408 [†] (0.115)	0.518 [†] (0.142)	0.628 [*] (0.381)
$x_{2,t-2}$	0.191 [†] (0.076)	0.071 (0.092)	-0.918 [†] (0.260)	0.167 [*] (0.101)	-0.043 (0.121)	-1.047 [†] (0.377)	0.196 [*] (0.114)	0.216 (0.142)	-0.850 [†] (0.380)
$x_{3,t-1}$	0.028 (0.018)	0.013 (0.022)	0.011 (0.062)	0.014 (0.024)	0.013 (0.028)	0.029 (0.088)	0.031 (0.035)	0.022 (0.035)	-0.050 (0.095)
$x_{3,t-2}$	-0.002 (0.018)	0.042 [*] (0.022)	-0.115 [*] (0.063)	0.009 (0.023)	0.014 (0.028)	-0.118 (0.087)	-0.001 (0.029)	0.072 [*] (0.036)	-0.103 (0.097)
<i>Conditional mean</i>									
Uncond. variance (σ_i)	0.062 [†] (0.012)	0.065 [†] (0.016)	0.800 [†] (0.145)	0.072 [†] (0.013)	0.059 [†] (0.026)	0.971 [†] (0.125)	0.061 [†] (0.016)	0.078 [†] (0.025)	0.708 [†] (0.272)
ARCH term (α_i)	0.130 (0.121)	0.082 (0.055)	0.030 (0.030)	0.296 [†] (0.133)	0.051 (0.044)	0.011 (0.067)	0.077 (0.079)	0.101 (0.090)	0.037 (0.046)
GARCH term (β_i)	0.376 (0.693)	0.721 [†] (0.257)	0.839 [†] (0.059)	0.000 (--)	0.849 [†] (0.226)	0.481 [†] (0.161)	0.255 (0.347)	0.532 (0.631)	0.852 [†] (0.049)
Degrees of freedom (ν_i)	7.803 [†] (3.351)	5.031 [†] (1.229)	9.338 [†] (3.993)	∞ (--)	7.710 [*] (4.094)	∞ (--)	5.141 [†] (2.239)	4.494 [†] (1.579)	6.456 [†] (2.745)
Log-likelihood	-54.506	-106.197	-427.721	-10.583	-35.331	-206.266	-36.261	-66.393	-214.794
<i>Conditional variance</i>									
<i>Misspecification Tests (p-values)</i>									
First moment ($k=1$)	0.256	0.130	0.251	0.467	0.263	0.716	0.594	0.980	0.947
Second moment ($k=2$)	0.263	0.538	0.863	0.444	0.835	0.825	0.888	0.526	0.703
Third moment ($k=3$)	0.279	0.436	0.608	0.762	0.517	0.775	0.843	0.926	0.992
Fourth moment ($k=4$)	0.239	0.537	0.708	0.615	0.864	0.673	0.698	0.441	0.799
Kolmogorov-Smirnov (t)	0.916	0.688	0.908	0.998	0.948	0.364	1.000	0.595	1.000
Anderson-Darling (t)	0.978	0.807	0.989	0.995	0.978	0.801	1.000	0.644	0.999
Kolmogorov-Smirnov (normal)	0.586	0.270	0.803	0.998	0.962	0.364	0.709	0.342	0.887
Anderson-Darling (normal)	0.559	0.160	0.756	0.995	0.891	0.801	0.541	0.242	0.740

Table 3: Parameter estimates of the marginal models for the full sample, and the two subsamples. Top panel: estimates of the conditional mean parameters. Middle panel: estimates of the t -GARCH(1,1) model for the conditional variance assuming a Student's t distribution for the innovations, and the log-likelihood for each margin. Standard errors are reported in brackets. †, ‡ and * mark statistically significant coefficients at the 1, 5 and 10 percent significance levels, respectively. Bottom panel: p-values of the misspecification tests for the dynamics of the probability integral transforms, and density misspecification tests for each model.

In the bottom panel of Table 3, we report the results of testing for serial and cross-margin independence in the probability integral transforms (PITs). The objective of this test is to ensure that the marginal models adequately capture the individual dynamics of each variable and do not include any dependence information, since the latter should only be captured by the copula. For this purpose, we use the Lagrange Multiplier test outlined in Patton (2006b). The test is conducted using $z_{x_{i,t}}^k = (\hat{u}_{i,t} - \bar{u}_{i,t})^k$ for $i = 1, 2, 3$, where $\hat{u}_{i,t} = F_i(x_{i,t}; \hat{\psi}_i)$ denote the PIT for the i -th margin given the estimated parameters, and $\bar{u}_i$ denotes its sample mean. For $k = 1, 2, 3, 4$, we regress each $z_{x_{i,t}}^k$ on 12 lags of itself and of the other two variables and compute R^2 from each regression. The test statistic is $(T - 36)R^2$ which is distributed under the null hypothesis as $\chi_{(36)}^2$. The null hypothesis of serial and cross-margin independence is rejected for any *p-value* less than 0.05. The results indicate that the null hypothesis of independent PITs (both serially and across margins) cannot be rejected.

We also apply the density misspecification Kolmogorov-Smirnov and Anderson and Darling (1952) tests to the PITs. The Anderson and Darling (1952) test has more power to detect misspecification in the tails of the distribution. We apply both tests to the PITs obtained when assuming either the normal or the Student's t density for the margins, which reveals interesting findings. The null hypothesis of a correct density specification cannot be rejected for all three margins in the FS as well both subsamples; however, in the FS, both tests give uniformly lower p-values when assuming normality providing evidence in favour of the Student's t density. This is particularly the case for $x_{2,t}$. This result is driven mainly by the results of SS2 suggesting that the factor changes may have been subject to more extreme movements during this period. This is consistent with the extreme realisations of $(x_{1,t}, x_{2,t}, x_{3,t})$ during the recent financial crisis; see the right panel of Figure 3.

Table 4 gives the constant and time-varying copula model estimation results for the FS and the two subsamples. For the constant copula models, the estimated dependence parameters, σ_{ij} , for the normal and t copulas are fairly similar. In SS2, there is an increase in $x_{1,t}/x_{2,t}$ and $x_{1,t}/x_{3,t}$ (negative) dependence given by σ_{12} and σ_{13} , respectively. This is coupled with a decline in $x_{2,t}/x_{3,t}$ dependence. This mirrors the previously-discussed behaviour of the level and slope factors toward the end of the FS. In both subsamples, there is an increase in some of the standard errors partly reflecting the small number of observations used in estimation. According to the likelihood ratio test, the t copula provides a statistically significant improvement in fit suggesting the possibility of joint extreme realisations in the factor changes. This is true in the FS and SS2, but not in SS1.

For the time-varying copulas, a is generally found to be statistically insignificant suggesting that δ is potentially unidentified in this case. However, in the FS the improvement in the likelihood over the constant copula model is statistically significant at 1 percent according to the LR test; this holds for both the normal and t copulas. This is true only in the FS suggesting that significant changes in the dependence structure have occurred between the two subsamples, rather than within each subsample. As in the case of the marginal models, the d.o.f. parameter is estimated with a large standard error, especially in SS2. However, in the FS and SS2 the time-varying t copula provides a statistically significant fit improvement relative to the time-varying normal copula, which suggests that joint extreme dependence is more prevalent in SS2.

	Full sample: 1986:4-2010:12				Subsample 1: 1986:4-1998:8				Subsample 2: 1998:9-2010:12			
	Constant copula		TV copula		Constant copula		TV copula		Constant copula		TV copula	
	normal	t	normal	t	normal	t	normal	t	normal	t	normal	t
σ_{12}	-0.785 [†] (0.033)	-0.788 [†] (0.041)	-	-	-0.693 [†] (0.061)	-0.688 [†] (0.097)	-	-	-0.861 [†] (0.040)	-0.862 [†] (0.062)	-	-
σ_{13}	-0.372 [†] (0.083)	-0.349 [†] (0.091)	-	-	-0.329 [†] (0.096)	-0.318 [†] (0.138)	-	-	-0.415 [†] (0.518)	-0.396 [†] (0.135)	-	-
σ_{23}	0.300 [†] (0.096)	0.277 [†] (0.091)	-	-	0.372 [†] (0.121)	0.363 [†] (0.182)	-	-	0.274 [†] (0.291)	0.248 [*] (0.145)	-	-
a	-	-	0.047 (0.044)	0.067 (0.056)	-	-	0.034 (0.245)	0.046 (0.265)	-	-	0.051 (0.158)	0.059 (0.308)
δ	-	-	0.940 [†] (0.057)	0.934 [†] (0.038)	-	-	0.896 [†] (0.248)	0.850 (0.600)	-	-	0.847 [*] (0.460)	0.865 (0.160)
ν	-	9.648 (10.997)	-	10.987 (6.758)	-	26.552 (19.762)	-	26.552 (22.033)	-	9.9438 (11.866)	-	10.364 (83.007)
Log-lik.	163.059	168.064	173.337	178.589	58.930	60.158	61.190	62.428	114.738	117.793	117.442	120.367

Table 4: Parameter estimates and log-likelihood of the constant and time-varying copula models. For the time-varying copula, we only report the estimates of the dynamic parameters. σ_{ij} denotes the off-diagonal elements of Σ , the copula dependence parameter. a and δ denote the parameters of the DCC copula model. ν denotes the d.o.f. parameter for the t copula. Standard errors are reported in brackets. †, ‡ and * mark statistically significant coefficients at the 1, 5 and 10 percent significance levels, respectively.

Figure 5 shows the conditional dependence measures estimated from the time-varying t copula model. The conditional dependence measures from the corresponding normal copula model are rather similar. The $x_{1,t}/x_{2,t}$ conditional dependence is strongly negative typically less than -0.750 and reaching about -0.900 towards the end of the sample. The interpretation of the strong negative association is that positive realisations of $x_{1,t}$, which indicate a level increase, tend to be associated with negative realisations of $x_{2,t}$ indicating a steepening of the TS. Given the highly negative association between $x_{1,t}$ and $x_{2,t}$, the $x_{1,t}/x_{3,t}$ and $x_{2,t}/x_{3,t}$ conditional correlations are almost a mirror image of one another reflected about a horizontal line slightly below the zero line.

4.4 Out-of-Sample Analysis

4.4.1 Conditional Mean Forecasts

To assess the predictive accuracy of our model for the conditional mean, we compare it to the Random Walk (RW) and VAR models both specified for the factor levels. The former is a widely-used benchmark which affine TS models often fail to outperform; see, for example, Duffee (2002) and Egorov et al. (2006). The VAR model with one-lag for the factor levels, which we denote by L-VAR(1), is the conditional mean specification commonly used in Diebold and Li (2006) as well its recent extensions discussed in Section 2.4.

Let s denote the forecast horizon. The RW model specifies that $X_{i,t+s|t} = X_{i,t}$, while the L-VAR(1) model specifies $X_{i,t+s|t} = \Phi_{X,0} + \Phi_{X,1}X_{i,t+s-1|t}$. We use $X_{i,t+s|t}$ to denote the s -period forecast given time t information set. Our model, denoted by D-VAR(2), is specified for the factor changes, and thus the s -period conditional mean forecast is given by $x_{i,t+s|t} = \Phi_0 + \Phi_1x_{i,t+s-1|t} + \Phi_2x_{i,t+s-2|t}$, which is then translated into a forecast for

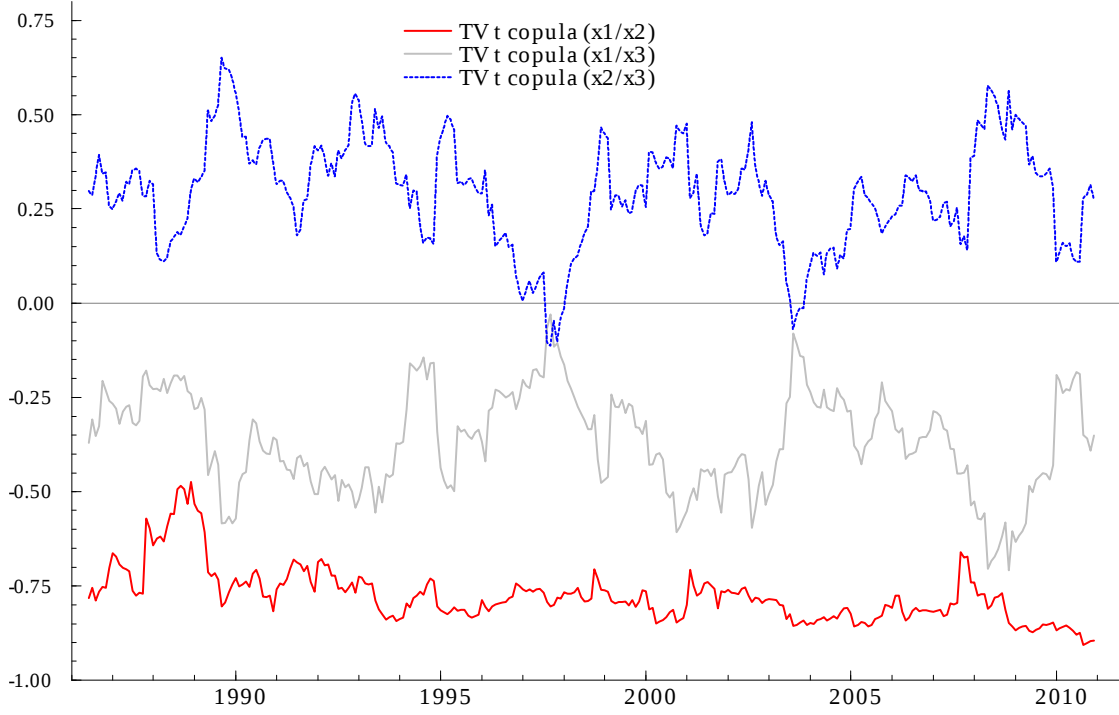


Figure 5: Conditional dependence measures from the time-varying t copula model.

the factor levels. We then use (1) to translate the factor level forecasts into NS-smoothed yield forecasts at the 17 maturities analysed previously. We compute root mean square error (RMSE) for the forecasted yields from each model at each of the 17 maturities for $s = 1, 6, 12$ months.¹⁴

We split the full sample equally, and use the first half to estimate the models' parameters by a rolling window. Our OOS is the period 1998:9-2010:12. To have an equal number of forecasts for each maturity, the last data point used in estimation is 2009:12. As a robustness check, we compute the RMSE from each model for two OOS subperiods: the first OOS subperiod is 1998:9-2004:5, and the second OOS subperiod is 2004:6-2010:12. The

¹⁴We also tried L-VAR(2) and D-VAR(1) specifications and found the results to be generally similar to the L-VAR(1) and D-VAR(2) specifications, respectively. These are not reported in the interest of brevity, but are available upon request.

reason for splitting the OOS is to independently assess the models' performance during the recent financial crises.

Our model improves modestly on the RW model only for some of the maturities particularly at the short and long ends of the yield curve. The performance improves for the 12-months forecast horizon especially in the second OOS subperiod, where it achieves an average reduction in RMSE of about 8 percent compared to the RW model. Although previous studies have found the L-VAR(1) specification to outperform the RW model, we found this not to hold in our OOS mainly due to inferior performance at medium-term maturities. This is particularly the case during the second OOS subperiod, which includes the financial crisis.

In Figure 6, we report the ratio of RMSE of our model to the RMSE of the L-VAR(1) model at each maturity for each forecast horizon. The results generally indicate superior performance relative to the L-VAR(1) models, with the exception of the 1-month forecasts in the first OOS subperiod. As the forecast horizon increases, the D-VAR(2) model achieves some reduction in RMSE. At the 12-months forecast horizon, in the second OOS subperiod including the recent financial crisis, the D-VAR(2) model reduces RMSE by more than 20 percent for medium-term maturities.

In addition to the RMSE criterion, we also assess the model's power in directional forecasting. Directional forecasting is concerned with the ability to predict the direction of change in the variable of interest, which is currently popular in short-term trading strategies. Let

$$d_{j,t} = \text{sign}(Y_t(\tau_j) - Y_{t-1}(\tau_j)), \quad \hat{d}_{j,t} = \text{sign}(\hat{Y}_t(\tau_j) - Y_{t-1}(\tau_j)),$$

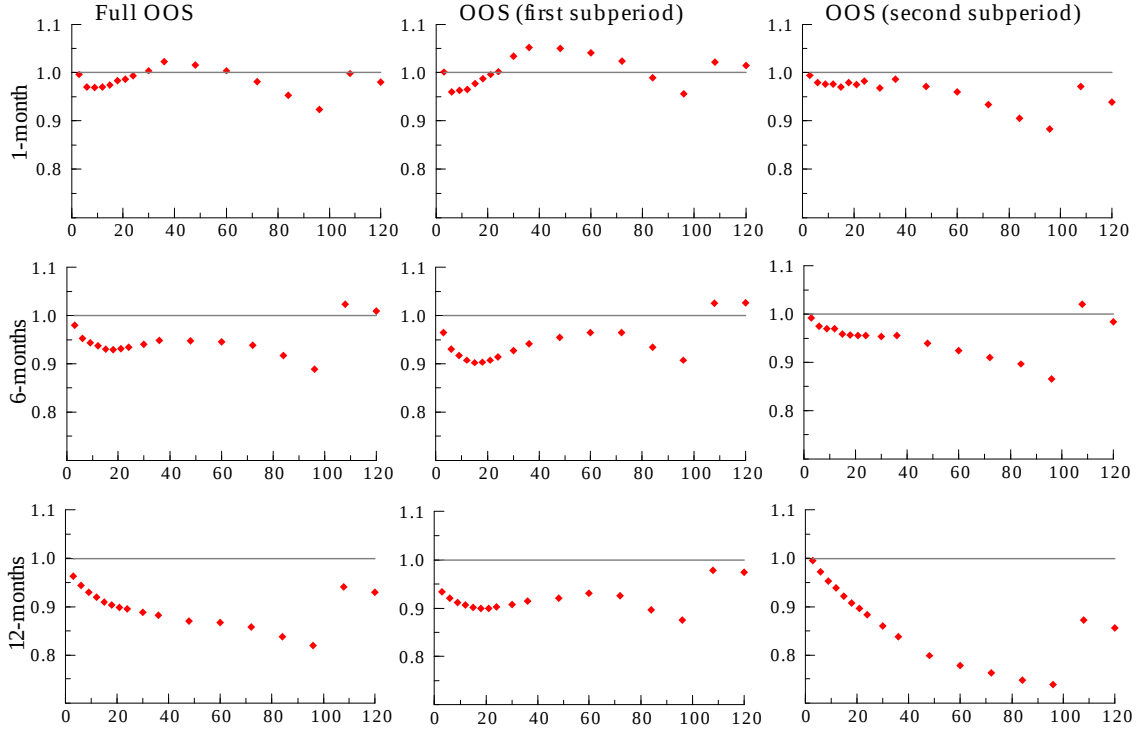


Figure 6: Ratio of the RMSE of D-VAR(2) model to the RMSE of the L-VAR(1) model at forecast horizons $s = 1, 6, 12$. The OOS period is 1998:8-2009:12. The first OOS subperiod is 1998:8-2004:5, while the second OOS subperiod is 2004:6-2010:12.

respectively denote the sign of the actual and forecasted yield changes for the j -th maturity

at time t . We define

$$c_{j,t} = \begin{cases} 1, & d_{j,t} = \hat{d}_{j,t} \\ 0, & \text{otherwise,} \end{cases},$$

and the count variable $c_t = \sum_{j=1}^n c_{j,t}$, which denotes the number of yields for which the sign is correctly forecasted at time t , thus c_t lies between 0 and 17. Let R^* denote the last data point in the rolling window used in estimation. We use c_t , $t = R^* + 1, R^* + 2, \dots$, as an additional criterion for evaluating the conditional mean forecasts from the competing models.

Figure 7 plots the ratio of c_t for both L-VAR(1) and D-VAR(2) relative to c_t of the RW

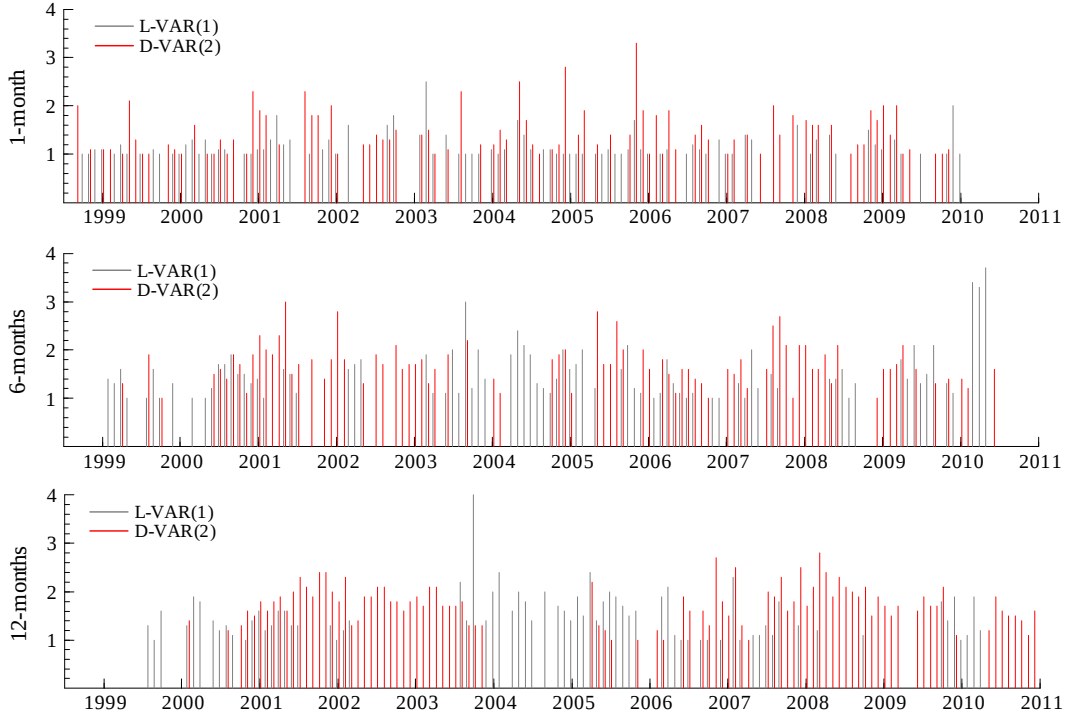


Figure 7: The ratio of c_t for both L-VAR(1) and D-VAR(2) relative to c_t of the RW model for the out-of-sample period 1998:9-2010:12. Any ratio less than one is set equal to zero.

model. To improve presentation, any ratio less than one has been set equal to zero.¹⁵ The results indicate that for the 1-month forecast horizon, both models appear, on average, to provide reliable directional forecasts in comparison to the RW model. Some months, e.g. in 2001 and 2002, are exceptions since both models do not improve on the number of correct sign predictions by the RW model. However, it is noticeable that the number of spikes above 1 for our model are greater than those of the L-VAR(1) model indicating a better ability in directional forecasting during these months.

At the 6-months forecast horizon, more heterogeneity among both models emerges as each performs better in certain periods. This can be seen during 2001-2002 in which

¹⁵Any ratio less than 1 indicates that at this point in time, the corresponding model did not match the number of correct sign predictions by the RW model. In these instances, setting the ratio equal to zero makes the charts less heavily populated with bars, hence easier to read.

	Constant copula		Time-varying copula	
	Normal	t	Normal	t
	(1)	(2)	(3)	(4)
No GARCH - normal density (A)	A1	A2	A3	A4
No GARCH - Student's t density (B)	B1	B2	B3	B4

Table 5: Models considered in conditional density forecasting.

our model shows better performance, and also during 2004 in which the L-VAR(1) model performs better. This distinction is even more evident in the 12-months forecasts, but in this case the D-VAR(2) model shows better performance relative to both the RW and L-VAR(1) models during the periods 2001-2003 and 2007-2010. During these two periods, the average number of maturities with correct sign predictions using D-VAR(2) was 14.8 and 13.8, respectively. For the RW model the figures are 8.4 and 9.1 maturities, while for the L-VAR(1) model the figures are 5.7 and 6.6 maturities, respectively.

4.4.2 Conditional Density Forecasts

For conditional density forecasts, we consider a number of models listed in Table 5. These models reflect the different features of our model, which enables us to determine the features that contribute, in a statistically significant manner, to improving the density forecast of the factor changes. Recall that in our model the density of the systematic yield component is proportional to the density of the factors since the factor loadings are fixed. Given the lack of evidence of a GARCH effect in $(x_{1,t}, x_{2,t}, x_{3,t})$, we focus on the marginal models without GARCH effects. Model A assumes that F_i in (11) follows a normal distribution, while model B assumes it follows a Student's t distribution. Combining these two models with different copulas gives the 8 different models in Table 5.

Since our objective is to emulate a real time forecasting exercise, we use a rolling-window of 238 observations to estimate the model parameters, thus the OOS period here

is 2006:1-2010:12. We use a larger estimation window compared to Section 4.4.1 to increase the accuracy of the parameters' point estimates. We consider forecast horizons $s = 1, 6, 12$, and so the last data point used in estimation is 2009:12.

In implementation, this exercise turns out to be challenging given the difficulty of estimating the d.o.f. parameters for each margin as well as the copula as discussed in Section 4.3. Attempting to estimate all the model parameters with a rolling-window led to unstable estimates of the d.o.f. parameters. Therefore, we fix the d.o.f. for each margin in model B and for the t copula at the estimated values in SS2. For the margins, these are reported in Table 3 and for the copula in Table 4. As a robustness check for this assumption, and given the wide confidence intervals around the point estimates, we also repeat the same exercise by fixing the d.o.f. for each margin at the lower and upper limits of the confidence intervals. However, if the lower bound for any of the marginal models is less than 2, we fix the d.o.f. at 2 to ensure that the second moment exists.¹⁶ In addition, we try different values for the t copula d.o.f. parameter.

The copula forecasts will differ according to whether the copula dependence parameter, Σ , is assumed to be constant or time-varying. For the constant copula model, the copula density will be computed as in (A.2) and (A.4), and the only change is that the PITs, u , will differ according to the specification of the marginal models. For the time-varying copulas, another dimension changing the density forecast is the dynamics of Σ_t . This is

¹⁶We also tried estimating a regime-switching model in which model B allows for a transition from a Student's t density with d.o.f. $\nu_{i,1}$ to a Student's t density with d.o.f. $\nu_{i,2}$ imposing that $\nu_{i,1} > \nu_{i,2}$. The idea is to estimate the latent conditional transition probabilities and allow the model to 'automatically' identify the regime switch over the in-sample period. Again, we found it difficult to accurately estimate the conditional transition probabilities given the difficulty of estimating the d.o.f. parameters.

computed as the forward recursion of (12) which gives

$$Q_{t+s|t} = (1 - \delta) \overline{Q} \sum_{j=1}^{s-1} (\delta)^{j-1} + (\delta)^{s-1} Q_{t+1|t},$$

$$\Sigma_{t+s|t} = (Q_{t+s|t} \circ I)^{-\frac{1}{2}} Q_{t+s|t} (Q_{t+s|t} \circ I)^{-\frac{1}{2}},$$

where $Q_{t+1|t}$ is $\mathcal{F}_t$ -measurable.

We use a predictive likelihood framework for model evaluation by comparing the OOS density forecasts for the models in Table 5. Define

$$L_{t,s} = \sum_{i=1}^3 L_{x_i,t,s} + L_{c,t,s},$$

where $L_{x_i,t,s} = l_{x_i}(\widehat{\psi}_{i,t}; x_{i,t+s})$ denotes log-likelihood of the i -th margin evaluated for time $t + s$ data $(x_{i,t+s})$ using parameters estimated at time t ($\widehat{\psi}_{i,t}$), and $L_{c,t,s} = l_c(\widehat{\kappa}_t; u_{t+s})$ is defined similarly for the copula. Thus $L_{t,s}$ denotes the predictive joint log-likelihood at time t for forecast horizon s . One difficulty with this approach is that although $x_{i,t+s}$ is observable OOS, u_{t+s} is not since it depends on the estimated marginal densities. Diks et al. (2008) propose to use the *forecast* standardised residual

$$\varepsilon_{i,t+s|t} = \frac{x_{i,t+s} - \mu_{i,t+s|t}}{\sigma_{i,t+s|t}},$$

and then compute $u_{i,t+s|t}$ using the empirical distribution function.¹⁷ Note that $\mu_{i,t+s|t}$ and $\sigma_{i,t+s|t}$ are forecasts based on parameters estimated at time t .

We propose an alternative approach in which we use the full sample parameter estimates

¹⁷In their method, they assume that marginal densities are estimated nonparametrically, while the copula is estimated parametrically.

to compute the conditional mean forecast, $\mu_{i,t+s|t}$, while the conditional variance forecast, $\sigma_{i,t+s|t}$, is assumed constant in models A and B. We conjecture that this approach is more consistent with the predictive likelihood framework which normally utilises the actual data in computation. The PITs computed using the full sample parameter estimates are the closest match to this setting since they condition on the entire history of the series. Although both the normal and Student's t densities pass the misspecification tests for the margins, we compute the PITs to be used for OOS comparison using the empirical distribution function. This ensures that our results will not be biased in favour of either model A or model B.

We undertake bivariate model comparisons based on the predictive likelihood differences denoted by

$$D_{t,s} = L_{t,s}^{m_1} - L_{t,s}^{m_2}, \quad t = R, R+1, \dots, T-s,$$

where $L_{t,s}^{m_1}$ and $L_{t,s}^{m_2}$ respectively denote the predictive likelihood from models m_1 and m_2 , and R is the size of the rolling estimation window. The average predictive likelihood gain is denoted by

$$\overline{D}_s = \frac{1}{T-R-s+1} \sum_{t=R}^{T-s} D_{t,s},$$

which is used to test the null hypothesis $H_0 : E[D_{t,s}] = 0$, for all s . The test statistic has the asymptotic distribution

$$\sqrt{T}(\overline{D}_s - \overline{D}_s^*) \xrightarrow{d} N(0, \Lambda_s),$$

where $\overline{D}_s^*$ denotes the average predictive likelihood gain at the true parameter value, and Λ_s is typically estimated using a HAC estimator. Significantly positive values of the test

		A1	A2	A3	A4	B1	B2	B3
A2	$s = 1$	-0.885						
	$s = 6$	-0.969						
	$s = 12$	0.862						
A3	$s = 1$	-1.021	0.762					
	$s = 6$	-0.884	1.151					
	$s = 12$	1.088	1.125					
A4	$s = 1$	-1.036	-1.208	-1.037				
	$s = 6$	-1.028	-1.058	-1.073				
	$s = 12$	-1.065	-1.030	-1.082				
B1	$s = 1$	2.477	0.920	1.090	1.053			
	$s = 6$	1.579	0.999	0.931	1.038			
	$s = 12$	1.221	-0.428	-1.075	1.087			
B2	$s = 1$	-1.003	-1.069	-0.999	-0.940	-1.015		
	$s = 6$	-1.063	-1.113	-1.122	0.100	-1.073		
	$s = 12$	-1.781	-1.515	-1.265	0.216	-1.677		
B3	$s = 1$	-1.011	0.723	-0.942	1.044	-1.069	1.001	
	$s = 6$	-0.904	1.034	0.823	1.055	-0.964	1.099	
	$s = 12$	1.154	1.283	-1.022	1.110	1.136	1.424	
B4	$s = 1$	-1.042	-1.132	-1.045	-1.053	-1.054	0.193	-1.049
	$s = 6$	-1.039	-1.073	-1.086	-1.475	-1.048	-0.388	-1.067
	$s = 12$	-1.082	-1.052	-1.087	-1.117	-1.097	-0.616	-1.113

Table 6: Predictive likelihood t -statistics for the out-of-sample conditional density forecasts for forecast horizons $s = 1, 6, 12$ months.

statistic indicate that model m_1 outperforms model m_2 out of sample. This test was introduced in Diebold and Mariano (1995), and later formalised by West (1996) and Giacomini and White (2006).

Table 6 reports the t -statistics from the predictive ability tests for each pair of models for forecast horizons $s = 1, 6, 12$ months. Significantly positive t -statistics favour the model listed horizontally, while significantly negative t -statistics favour the model listed vertically. The results indicate that, with only one exception, none of the models provides statistically significant OOS improvements in comparison to its competitors. This could potentially be due to either using a relatively small OOS period or fixing the d.o.f. parameters for the margins (in model B) and for the t copula. However, the results are qualitatively similar when using different estimates of the d.o.f. parameters.

The signs of the t -statistics are in agreement with our previous findings. For instance, models based on the t copula (A2, A4, B2 and B4) provide generally better density forecasts

compared to both the constant and time-varying normal copulas. The time-varying t copula model (A4 and B4) achieves, on average, an improvement in the predictive likelihood compared to the constant t copula model (A2 and B2). It is also notable that the B1 model seems inferior to all other models, suggesting that combining Student's t margins with the normal copula leads to the worst OOS performance. These statements, however, lack the lustre of statistical significance.

5 Conclusion

The objective of this paper is to introduce a forecasting model for the TS of interest rates that is parsimonious and relatively immune to structural breaks. We use the DNS model as a useful device to summarise the TS into three factors interpreted as the level, slope and curvature of the yield curve. The dimension reduction achieved through the DNS model allows us to develop a relatively parsimonious model of the TS by focusing on modelling and forecasting the DNS factors. Immunity to structural breaks is relatively achieved by modelling the factor changes rather than the factor levels, as we find the unconditional moments of the former to be relatively more stable.

The model we propose utilises recent advances in copula theory to allow for flexible modelling of the factor dynamics. Copula methods allow for a complete characterisation of the joint density of the factor changes and enable us to incorporate new features, namely allowing departures from multivariate normality, and allowing the factor changes to have a time-varying dependence structure that also permits joint extreme dependence. The in-sample analysis suggests that assuming a Student's t distribution for the margins, coupled with a time-varying t copula provides statistically significant gains in goodness of fit. Sub-

sample results indicate that this is potentially due to joint extreme realisations in the factor changes during the recent financial crises. There is also evidence of increasing dependence between changes in the level and slope factors during the same period, reflecting the fact that with the short end of the yield curve fixed close to zero, any changes in the yields at medium- and long-term maturities will increase both the level and steepness of the TS. The direct implication of this is that stress testing scenarios which view changes in the level and the slope of the TS as independent may provide wrong conclusions for financial risk management.

Although our model's conditional mean specification has quite a few parameters, it is able to provide RMSE and directional forecast improvements, relative to two widely-used and more parsimonious benchmarks. The forecast improvements are particularly evident at longer forecast horizons. For instance, during the recent financial crisis, our model is able to provide remarkably accurate 12-months-ahead predictions about the direction of change in the yield curve. These results highlight the attractiveness of well specified and stable models for forecasting changes in the TS.

References

- Anderson, T. W. and D. A. Darling (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics* 23, 193–212.
- Audrino, F. and F. Trojani (2007). Accurate short-term yield curve forecasting using functional gradient descent. *Journal of Financial Econometrics* 5, 591–623.
- Bliss, R. (1997). Movements in the term structure of interest rates. *Federal Reserve Bank of Atlanta Economic Review* 82, 16–33.
- Bowsher, C. G. and R. Meeks (2008). The dynamics of economic functions: modelling and forecasting the yield curve. *Journal of the American Statistical Association* 103, 1419–1437.
- Casella, G. and R. L. Berger (2001). *Statistical Inference* (2 ed.). Duxbury Press.
- Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2009). An arbitrage-free generalized Nelson-Siegel term structure model. *The Econometrics Journal* 12, 33–64.

- Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2011). The affine arbitrage-free class of Nelson-Siegel term structure models. *Journal of Econometrics*. Forthcoming.
- Christensen, J. H. E., J. A. Lopez, and G. D. Rudebusch (2010). Can spanned term structure factors drive stochastic volatility? Working paper: Federal Reserve Bank of San Francisco.
- Clements, M. P. and D. F. Hendry (2006). Forecasting with breaks. In G. Elliott, C. W. J. Granger, and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, pp. 605–657. North-Holland: Amsterdam.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–407.
- Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.
- De Pooter, M. (2007). Examining the Nelson-Siegel class of term structure models. Working paper: Tinbergen Institute.
- Demarta, S. and A. McNeil (2005). The t copula and related copulas. *International Statistical Review* 73, 111–129.
- den Goorbergh, R. W. J. V., C. Genest, and B. J. M. Werker (2005). Multivariate option pricing using dynamic copula models. *Insurance: Mathematics and Economics* 37, 101–114.
- Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130, 337–364.
- Diebold, F. X., C. Li, and V. Z. Yue (2008). Global yield curve dynamics and interactions: a dynamic Nelson-Siegel approach. *Journal of Econometrics* 146, 351–363.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Diebold, F. X., M. Piazzesi, and G. D. Rudebusch (2005). Modelling bond yields in finance and macroeconomics. *American Economic Review* 95, 415–420.
- Diebold, F. X., G. D. Rudebusch, and S. B. Aruoba (2006). The macroeconomy and the yield curve: a dynamic latent factor approach. *Journal of Econometrics* 131, 309–338.
- Diks, C., V. Panchenko, and D. van Dijk (2008). Out-of-sample comparison of copula specifications in multivariate density forecasts. Working paper: Australian School of Business, University of New South Wales.
- Duffee, G. R. (1996). Idiosyncratic variation of treasury bill yields. *Journal of Finance* 51, 527–551.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.
- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.
- Egorov, A. V., Y. Hong, and H. Li (2006). Validating forecasts of the joint probability density of bond yields: can affine models beat random walk? *Journal of Econometrics* 135, 255–284.

- Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Fabozzi, F. J., L. Martellini, and P. Priaulet (2005). Predictability in the shape of the term structure of interest rates. *Journal of Fixed Income*, 40–53. June 2005.
- Fama, E. F. and R. R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review* 77, 680–692.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Harvey, A. (2008). Dynamic distributions and changing copulas. Working paper: University of Cambridge.
- Hautsch, N. and Y. Ou (2008). Yield curve factors, yield volatility and the predictability of bond excess returns. Discussion Paper: Humboldt-Universitat zu Berlin.
- Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claim valuation. *Econometrica* 60, 77–105.
- Hull, J. and A. White (1990). Pricing interest rate derivative securities. *Review of Financial Studies* 3, 573–592.
- Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*. London: Chapman and Hall.
- Joe, H. and J. Xu (1996). The estimation method of inference functions for margins for multivariate models. Unpublished manuscript: Department of Statistics, University of British Columbia.
- Jondeau, E. and M. Rockinger (2006). The copula-GARCH model of conditional dependencies: an international stock market application. *Journal of International Money and Finance* 25, 827–853.
- Koopman, S. J., M. I. P. Mallee, and M. van der Wel (2010). Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *Journal of Business and Economic Statistics* 28, 329–343.
- Litterman, R. and J. Scheinkman (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 54–61. June 1991.
- Nelsen, R. B. (2006). *An Introduction to Copulas* (2 ed.). New York: Springer.
- Nelson, C. R. and A. F. Siegel (1987). Parsimonious modelling of yield curve. *Journal of Business* 60, 473–489.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2111–2245. North-Holland.

- Pagan, A. R., A. D. Hall, and V. Martin (1996). Modelling the term structure. In C. R. Rao and G. S. Maddala (Eds.), *Handbook of Statistics*, pp. 91–118. North-Holland: Amsterdam.
- Panchenko, V. (2005). Estimating and evaluating the predictive abilities of semiparametric multivariate models with application to risk management. Mimeo: University of Amsterdam.
- Patton, A. J. (2006a). Estimation of multivariate models for time series of possibly different lengths. *Journal of Applied Econometrics* 21, 147–173.
- Patton, A. J. (2006b). Modelling asymmetric exchange rate dependence. *International Economic Review* 47, 527–556.
- Patton, A. J. (2009). Copula-based models for financial time series. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 767–785. Springer-Verlag.
- Rodriguez, J. C. (2007). Measuring financial contagion: a copula approach. *Journal of Empirical Finance* 14, 401–423.
- Rudebusch, G. D. (2010). Macro-finance models of the interest rate and the economy. Working paper: Federal Reserve Bank of San Francisco.
- Sklar, A. (1959). Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut Statistique de l'Universite de Paris* 8, 229–231.
- Svensson, L. E. O. (1995). Estimating forward interest rates with the extended Nelson-Siegel method. *Sveriges Riksbank Quarterly Review* 3, 13–26.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

A Overview of Copula Theory

A.1 Modelling Dependence Using Copulas

Copulas offer a general framework for creating flexible distributions to allow for departures from normality, especially the modelling of nonlinear and extreme dependence. Copula methods enable a decomposition of the joint probability distribution of a group of variables into univariate marginal distributions, and a copula that includes only information about their dependence structure.

In this overview, we focus primarily on copula theory, while the copula applications related to our model are mentioned in Section 3.3 and towards the end of this appendix. An introduction to copula theory is given in Joe (1997) and Nelsen (2006). Patton (2009) provides a recent survey of copula applications in finance.

Let F and f respectively denote the cumulative and probability distribution functions of a random variable. For a random variable X_i , let the lowercase x_i denote a particular realisation of X_i . For a k -dimensional random vector $X = (X_1, X_2, \dots, X_k)'$, let F_i denote the i -th marginal distribution. The main result in copula theory is the Theorem of Sklar (1959), which states that

$$F(x) = \Pr(X_1 \leq x_1, X_2 \leq x_2, \dots, X_k \leq x_k) = C(F_1(x_1), F_2(x_2), \dots, F_k(x_k)). \quad (\text{A.1})$$

By defining $u_i = F_i(x_i)$, (A.1) can be expressed as $F(x) = C(u_1, u_2, \dots, u_k)$, where $u_i \sim U(0, 1)$ is the probability integral transform of X_i . This result shows that for any joint distribution F , there exists a copula C which is a joint distribution of the uniform margins. The copula is a function that satisfies the following properties:

1. $C : [0, 1]^k \rightarrow [0, 1]$;
2. C is grounded and k -increasing;
3. C has margins C_k which satisfy $C_k(u) = (1, \dots, 1, k, 1, \dots, 1) = u$ for all u in $[0, 1]$.

The first property states that C maps from the k -dimensional hypercube to the unit interval. The second property requires that the C -volume of all k -boxes whose vertices lie in $[0, 1]^k$ is non-negative ensuring non-negative joint probability. And the third property follows directly from (A.1).

One can also define the copula in terms of the joint and marginal *pdfs*. Assuming F is k -times differentiable, computing the k -th cross-partial derivative gives

$$f(x) = \prod_{i=1}^k f_i(x_i) \cdot c(F_1(x_1), F_2(x_2), \dots, F_k(x_k)),$$

where f and f_i denote the joint and marginal *pdfs*, respectively, and c is the copula density. This representation is important for modelling since one can model a joint *pdf* as the product of the marginal *pdfs* and the copula density. Each marginal density contains only information about the individual features and dynamics of its corresponding variable, while the copula contains only information about the dependence structure among all k variables. The relationship between copulas and some commonly used measures of dependence is discussed in Joe (1997) and Nelsen (2006). Patton (2006b) extends the Theorem of Sklar (1959) to the case of conditional distributions.

A.2 The Normal and t Copulas

For the copulas derived from known multivariate distributions, the most commonly used are the normal and t copulas. For a vector $u = (u_1, \dots, u_k)$, the normal copula is

$$C(u_1, \dots, u_k; \Sigma) = \Phi_{\Sigma, k}(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k)), \quad (\text{A.2})$$

where $\Phi_{\Sigma, k}$ is a k -dimensional standardised normal distribution with correlation matrix Σ , and Φ^{-1} is the inverse *cdf* of a standard normal variable. The normal copula density is

given by

$$c_t(u_1, \dots, u_k; \Sigma) = \frac{1}{\sqrt{|\Sigma|}} \exp \left\{ -\frac{(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k))'(\Sigma^{-1} - I_k)(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_k))}{2} \right\}, \quad (\text{A.3})$$

where I_k is the identity matrix of order k .

The t copula is given by

$$C(u_1, \dots, u_k; \Sigma, \nu) = t_{\Sigma, k}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_k)), \quad (\text{A.4})$$

where $t_{\Sigma, k}$ is a k -dimensional standardised Student's t distribution with d.o.f. ν and correlation matrix Σ , and t_ν^{-1} is the inverse *cdf* of a standardised Student's t distribution with d.o.f. ν . Its density is given by

$$c(u_1, \dots, u_d; \Sigma, \nu) = \frac{\Gamma((\nu + k)/2) [\Gamma(\nu/2)]^{k-1}}{\sqrt{|\Sigma|} [\Gamma((\nu + 1)/2)]^k} \left[1 + \frac{\tilde{Z}' \Sigma^{-1} \tilde{Z}}{\nu} \right]^{-(\nu+k)/2} \prod_{i=1}^k \left[1 + \frac{\tilde{z}_i^2}{\nu} \right]^{(\nu+1)/2} \quad (\text{A.5})$$

where $\tilde{Z} = (\tilde{z}_1, \dots, \tilde{z}_k)'$, $\tilde{z}_i = t_\nu^{-1}(u_i)$, and $\Gamma(\cdot)$ denotes the gamma function. The t copula nests the normal copula as $\nu \rightarrow \infty$. For $\nu < \infty$, the t copula generates tail dependence¹⁸ which does not obtain under the normal copula. The dependence structure in these two copulas is fully contained in the matrix Σ . The number of parameters in the normal copula is $\frac{1}{2}k(k-1)$, and in the t copula is $\frac{1}{2}k(k-1) + 1$ for the additional d.o.f. parameter.

When the normal (t) copula is coupled with univariate margins from a family other than the normal (Student's t) distribution, the resulting joint distribution is often called a meta-normal (meta- t) distribution (Demarta and McNeil, 2005). In this case, the off-diagonal

¹⁸See Joe (1997) for a formal definition of tail dependence.

elements of Σ may capture nonlinear forms of dependence.

Time variation in copulas can be introduced in different ways. One approach is to assume the functional form for the copula is constant, but its parameter evolves over time, which implies dynamic evolution in the dependence structure. This approach is adopted in, for example, Panchenko (2005), den Goorbergh et al. (2005) and Patton (2006b). Jondeau and Rockinger (2006) propose a regime-switching model for the copula parameter, while Rodriguez (2007) allows for a changing copula function. Harvey (2008) focuses on non-parametrically estimated copulas and tracks time variation by an exponentially-weighted moving average filter.

CHAPTER III

MULTIVARIATE HIGH-FREQUENCY-BASED VOLATILITY

(HEAVY) MODELS

1 Introduction

This paper introduces a new class of multivariate volatility models capable of producing precise multi-step forecasts of the conditional covariance matrix of daily returns. Multivariate volatility models have been the focus of a voluminous literature summarised recently by Bauwens et al. (2006) and Asai et al. (2006), where the focus in the latter is on multivariate stochastic volatility.

The covariance matrix of daily asset returns is a key input in portfolio allocation, option pricing and financial risk management. An interesting question is whether the increasing availability of high-frequency financial data enables the development of more accurate forecasting models for the conditional covariance of daily returns. We address this question by studying a new class of models which utilise high-frequency data for the objective of multi-step volatility forecasting. We call this class multivariate High-frEquency-bAsed VolatilitY (HEAVY) models.

Volatility forecasts from HEAVY models have some properties that distinguish them from those of multivariate GARCH models. HEAVY models have a relatively short response time which means they are likely to perform well in periods where the level of volatility or correlation is subject to abrupt changes. HEAVY models also have short-run momentum effects so that volatility forecasts may exhibit a continuation of upward (or downward) trends before mean reverting. The latter distinction pertains to comparing the HEAVY model to a baseline specification such as the GARCH(1,1) model. More richly parameterised GARCH models could, of course, also exhibit momentum effects.

The univariate HEAVY model was introduced in Shephard and Sheppard (2010) where it is shown - for a wide spectrum of asset classes - that the HEAVY model outperforms

the GARCH model in- and out-of-sample. The forecast gains tend to be more pronounced at short forecast horizons, typically the first few days. In the empirical section of this paper, we show similar results in a multivariate setting. The multivariate analysis poses additional interesting questions such as whether the forecast gains are due to the variance forecasts of individual assets, their correlations or a combination of both. We develop a novel out-of-sample model evaluation strategy to address this question.

To highlight the distinction between HEAVY and GARCH models, and how HEAVY models differ from recently proposed models which also utilise high-frequency data, we start with a brief overview of the univariate HEAVY model of Shephard and Sheppard (2010). Let $\mathcal{F}_t^{LF}$ and $\mathcal{F}_t^{HF}$ respectively denote the information set generated by low-frequency (i.e. daily) and high-frequency (i.e. intra-daily) data up to time t , where $t = 1, 2, \dots$, indexes days. Also let r_t denote the (de-means) daily return and v_t denote the realised measure (e.g. realised variance) at time t . The univariate HEAVY model in its linear specification is the 2-equation system

$$\mathbb{E}[r_t^2 | \mathcal{F}_{t-1}^{HF}] := h_t = c_h + b_h h_{t-1} + a_h v_{t-1},$$

$$\mathbb{E}[v_t | \mathcal{F}_{t-1}^{HF}] := m_t = c_m + b_m m_{t-1} + a_m v_{t-1},$$

while the GARCH model is

$$\mathbb{E}[r_t^2 | \mathcal{F}_{t-1}^{LF}] := h_t^* = c_g + b_g h_{t-1}^* + a_g r_{t-1}^2.$$

The primary distinction between HEAVY and GARCH models is the conditioning information set used in modelling the conditional variance of daily returns. The first

equation of the HEAVY model uses the lagged realised measure, v_{t-1} , to drive the dynamics of h_t , whereas the GARCH model uses the lagged squared return. The second equation of the HEAVY model is needed for multi-step forecasts of h_t .

The HEAVY model utilises recently developed estimators of ex post volatility of daily returns that have proven to be more precise compared to squared returns. Realised variance is the first realised measure to be systematically studied and used in modelling and forecasting the volatility of daily returns. Andersen and Bollerslev (1998) show that the realised variance has a much lower noise-to-signal ratio than the daily squared return when used as proxy for the unobserved variance, while Barndorff-Nielsen and Shephard (2002) formalise the econometrics of the realised variance. In the context of multi-step forecasting, Shephard and Sheppard (2010) show that the use of the realised kernel of Barndorff-Nielsen et al. (2008) leads to notable in- and out-of-sample improvements in predicting h_t , especially at short forecast horizons.

Univariate HEAVY models are related to recently proposed models by Engle (2002), Engle and Gallo (2006), Cipollini et al. (2007), Brownlees and Gallo (2010) and Hansen et al. (2011). Engle (2002) models volatility using a multiplicative error model (MEM).¹ He applies this model to squared returns and realised volatility as separate models, but they were not considered as a system for multi-step forecasting of the conditional variance of daily returns. These models are usually referred to as GARCH-X models when both v_{t-1} and r_{t-1}^2 appear in the h_t equation. Engle and Gallo (2006) model a 3-variable system comprising the squared return, the high-minus-low price range and the realised variance in an MEM setup. Cipollini et al. (2007) allow for contemporaneous correlations in a 4-

¹An MEM can be used for any non-negative valued process which can be modelled as *i.i.d.* innovations from a density with non-negative support scaled by a conditionally deterministic factor.

variable vector MEM including the absolute daily return and three realised measures, and tackle the problem of a suitable multivariate density choice using copulas.

The papers by Brownlees and Gallo (2010) and Hansen et al. (2011) are the closest in structure to the univariate HEAVY model. The model in Brownlees and Gallo (2010) has a HEAVY-like structure with the difference being that it uses a smoothed version of the realised measure to drive h_t by specifying the latter as an affine function of m_t . Hansen et al. (2011) treat the dynamics of the realised measure differently. While the HEAVY model postulates GARCH-type dynamics for the realised measure by modelling its conditional expectation, Hansen et al. (2011) relate the realised measure itself to h_t and a term that captures leverage effects.

Multivariate volatility models are becoming increasingly important not only because of their direct application in portfolio allocation and asset pricing, but also due to the insights they provide into risk management practices. Using low-frequency data, Brownlees and Engle (2010) portray the importance of modelling conditional correlations for systemic risk management, where they show that a rise in a firm's stock volatility and correlation with the market magnifies its contribution to their proposed measure of systemic risk. Highly leveraged financial companies in the recent financial crisis are a case in point. The work of Hansen et al. (2010), which is independent and concurrent, utilises realised measures in modelling a stock's conditional beta in a GARCH-like framework. Our primary empirical example focuses on the returns of Bank of America and the S&P 500 exchange traded fund (ETF) during the recent financial crisis, which relates to the applications in these papers.

There is some recent research that focuses only on modelling and forecasting the realised covariance matrix; see, for example, Voev (2008), Chiriac and Voev (2011) and Bauer and Vorkink (2011). The focus in these studies is on developing parsimonious models to forecast

the realised covariance matrix. In contrast, this paper develops a framework for forecasting the covariance of daily returns which also requires forecasts of the realised measure. We find the realised measure to be a more precise factor to drive the volatility dynamics for daily returns compared to the outer product of daily returns which is used in GARCH models.

Jin and Maheu (2010) pursue an objective similar to ours by utilizing realised measures to improve the density forecasts of multivariate daily returns; however, their model is different from ours as it is cast in the multivariate stochastic volatility framework. In addition, they propose a different nexus between the dynamics of daily returns and the realised measure. The implication of this is that our model is much easier to estimate and allows for straightforward out-of-sample model evaluation since we provide closed-form forecasting formulas.

The structure of the paper is as follows: Section 2 introduces multivariate HEAVY models with some detailed analysis of their properties using a linear specification. Section 3 discusses estimation and inference. In Section 4, we present the out-of-sample model evaluation framework. Section 5 contains the results of our empirical analysis, while Section 6 concludes the paper. Appendix A includes definitions and results from matrix algebra and matrix calculus which are used in some of the derivations and proofs. Appendix B derives the second moments' structure implied by the model. Appendix C gives a brief overview of the Wishart distribution which we employ in specifying the density of the innovations. All proofs are collected in Appendix D.

2 Multivariate HEAVY Models

2.1 Definitions and Notation

Let the multivariate log-price process be given by the $(k \times 1)$ vector Y_τ^* , where $\tau \in \mathbb{R}_+$ represents continuous time. Suppose we observe $m + 1$ intra-daily prices, assumed to be uniformly spaced, so that the j -th intra-daily vector of returns on day t is given by

$$R_{j,t} = Y_{(t-1)+\frac{j}{m}}^* - Y_{(t-1)+\frac{(j-1)}{m}}^*, \quad j = 1, \dots, m, \quad t = 1, 2, \dots$$

Assuming, for instance, 24-hour trading means $m = 1440$ for one-minute returns, and $R_{j,t}$ is the vector of returns for the j -th minute on day t . The vector of daily returns is $R_t = \sum_{j=1}^m R_{j,t}$. The outer product of daily returns is the $(k \times k)$ matrix denoted by $P_t = R_t R_t'$. The realised measure on day t is a $(k \times k)$ matrix denoted by V_t . One example of V_t which we use in this paper is the realised covariance (RC_t) matrix defined as

$$RC_t = \sum_{j=1}^m R_{j,t} R_{j,t}'.$$

Barndorff-Nielsen and Shephard (2004) show that, in the absence of market microstructure noise, RC_t is a mixed normal consistent estimator of the quadratic covariation of Y_τ^* as $m \rightarrow \infty$. In the presence of market microstructure noise, RC_t is a biased estimator. Therefore, in practice one needs to sample sparsely and use subsampling. An alternative is to use a noise-robust estimator such as the realised kernel of Barndorff-Nielsen et al. (2008, 2011).

Letting $\mathcal{F}_t^{LF}$ and $\mathcal{F}_t^{HF}$ be as defined previously, the HEAVY model is the 2-equation

system

$$\mathbb{E}[P_t | \mathcal{F}_{t-1}^{HF}] = \mathbb{E}[R_t R_t' | \mathcal{F}_{t-1}^{HF}] := H_t, \quad (1)$$

$$\mathbb{E}[V_t | \mathcal{F}_{t-1}^{HF}] := M_t, \quad (2)$$

where, for simplicity, we assume $\mathbb{E}[R_t | \mathcal{F}_{t-1}^{HF}] = 0$ so that H_t is the conditional covariance matrix of daily returns, or alternatively, the conditional expectation of the outer product of daily returns. We will occasionally use $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot | \mathcal{F}_t^{HF}]$ to denote the expectation conditional on $\mathcal{F}_t^{HF}$. Thus the conditional first moments (H_t, M_t) are assumed $\mathcal{F}_{t-1}^{HF}$ -measurable.

We shall call (1)-(2) the HEAVY-P and HEAVY-V equations, respectively. HEAVY models can be equivalently represented as

$$P_t = H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}}, \quad (3)$$

$$V_t = M_t^{\frac{1}{2}} \eta_t M_t^{\frac{1}{2}}, \quad (4)$$

where ε_t and η_t are $(k \times k)$ symmetric innovation matrices satisfying $\mathbb{E}_{t-1}[\varepsilon_t] = \mathbb{E}_{t-1}[\eta_t] = I_k$, where I_k is an identity matrix. We have defined the symmetric square root of a generic positive semidefinite matrix A , denoted by $A^{\frac{1}{2}}$, using the spectral decomposition such that $A^{\frac{1}{2}} = U \Lambda^{\frac{1}{2}} U'$ where U is a matrix containing the eigenvectors of A , and $\Lambda^{\frac{1}{2}}$ is a diagonal matrix containing the square root of the eigenvalues of A . The representation (3)-(4) is a matrix-variate generalisation of the univariate MEM introduced in Engle (2002) and the vector MEM presented in Cipollini et al. (2007).

Since our focus is on multivariate volatility models, we use the terms HEAVY and GARCH to refer to their multivariate formulation unless otherwise stated. The difference

between the HEAVY-P equation and the GARCH model is the conditioning information set. GARCH models condition on $\mathcal{F}_{t-1}^{LF}$ and thus H_t is influenced by the squares and cross products of past daily returns (i.e. lags of P_t). In the HEAVY-P equation, we condition on $\mathcal{F}_{t-1}^{HF}$ which enables us to use lags of V_t to project the path of H_t .

Equations (1)-(2), or equivalently (3)-(4), define a class of models which links the dynamics of H_t to the realised measure. This becomes clear once we specify the dynamic equations for H_t and M_t . Choosing a specification for the dynamics of H_t and M_t yields a particular model within the HEAVY class. For ease of presentation, we will focus in the rest of this paper on one particular specification within the HEAVY class which is akin to a multivariate GARCH(1,1) model, and we shall refer to it simply as the HEAVY model.

2.2 Model Parameterisation

A primary challenge in multivariate volatility modelling is to ensure that the conditional covariance matrix is positive semidefinite. In the GARCH literature, one of the ways this has been approached is the BEKK parameterisation introduced by Engle and Kroner (1995). We can adopt that approach to our model, which we call BEKK-type parameterisation although the models are distinct. The BEKK-type parameterisation is

$$H_t = \bar{C}_H \bar{C}_H' + \bar{B}_H H_{t-1} \bar{B}_H' + \bar{A}_H V_{t-1} \bar{A}_H', \quad (5)$$

$$M_t = \bar{C}_M \bar{C}_M' + \bar{B}_M M_{t-1} \bar{B}_M' + \bar{A}_M V_{t-1} \bar{A}_M'. \quad (6)$$

The $(k \times k)$ matrices $\bar{A}_H$, $\bar{B}_H$, $\bar{A}_M$ and $\bar{B}_M$ each have k^2 free parameters, while $\bar{C}_H$ and $\bar{C}_M$ are $(k \times k)$ lower triangular matrices each with $k^* = k(k+1)/2$ free parameters. The parameterisation in (5)-(6) guarantees that H_t and M_t are positive semidefinite for all

t assuming H_0 and M_0 are positive semidefinite. If, in addition, $\bar{C}_H$ and $\bar{C}_M$ are full rank matrices, then H_t and M_t are positive definite for all t . We refer to $\bar{A}_H$, $\bar{B}_H$, $\bar{A}_M$ and $\bar{B}_M$ as the dynamic parameters, which are of main interest to us. Sometimes we consider $\bar{C}_H$ and $\bar{C}_M$ to be "nuisance parameters".

Although our interest is to obtain multi-step forecasts of H_t , forecasts from (6) are needed due to the presence of V_{t-1} in (5). Forecasting the realised measure itself has been the focus of a number of recent studies, e.g. Andersen et al. (2003, 2007, 2011). We note that postulating GARCH-type dynamics for the realised measure is consistent with its empirical properties such as time-varying volatility of realised volatility and evidence of excess kurtosis; see Corsi et al. (2008). Therefore, (6) may produce accurate forecasts of M_t .

Of course, other parameterisations for (5)-(6) could be adopted. For instance, a higher order lag structure akin to GARCH(p, q) processes, or a component model which decomposes the conditional covariance matrix into long-run (permanent) and short-run (transitory) components as in Engle and Lee (1999). Also, a long memory model could be specified for (6) as proposed in Chiriac and Voev (2011).

The unrestricted BEKK-type parameterisation in (5)-(6) has $O(k^2)$ parameters. To avoid the curse of dimensionality one could impose that $\bar{A}_H$, $\bar{B}_H$, $\bar{A}_M$ and $\bar{B}_M$ are scalars or diagonal matrices, which yields the scalar or diagonal HEAVY model, respectively. In either case, the resulting equations for the diagonal elements of H_t and M_t would constitute univariate HEAVY models. The equations for the off-diagonal elements would also have a HEAVY structure in which the conditional covariances are driven by their own lags and the corresponding realised covariances. If the elements of $\bar{A}_H$, $\bar{B}_H$, $\bar{A}_M$ and $\bar{B}_M$ are unrestricted (i.e. a full HEAVY parameterisation), the multivariate HEAVY model

no longer comprises univariate HEAVY models, since in this case the evolution of every element in H_t and M_t will be influenced by own as well as cross-asset effects.

Example 1 For the H_t equation in the scalar HEAVY model, $\bar{A}_H = \bar{a}_H I_k$ and $\bar{B}_H = \bar{b}_H I_k$ where $\bar{a}_H$ and $\bar{b}_H$ are scalars, which gives the following parameterisation

$$H_t = \bar{C}_H \bar{C}_H' + \bar{b}_H^2 H_{t-1} + \bar{a}_H^2 V_{t-1}.$$

In the case of the bivariate diagonal HEAVY model, the H_t equation is given by

$$\begin{pmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{pmatrix} = \begin{pmatrix} \bar{c}_{11,H} & 0 \\ \bar{c}_{21,H} & \bar{c}_{22,H} \end{pmatrix} \begin{pmatrix} \bar{c}_{11,H} & 0 \\ \bar{c}_{21,H} & \bar{c}_{22,H} \end{pmatrix}' + \begin{pmatrix} \bar{b}_{11,H} & 0 \\ 0 & \bar{b}_{22,H} \end{pmatrix} \begin{pmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{pmatrix} \begin{pmatrix} \bar{b}_{11,H} & 0 \\ 0 & \bar{b}_{22,H} \end{pmatrix} + \begin{pmatrix} \bar{a}_{11,H} & 0 \\ 0 & \bar{a}_{22,H} \end{pmatrix} \begin{pmatrix} v_{11,t-1} & v_{12,t-1} \\ v_{21,t-1} & v_{22,t-1} \end{pmatrix} \begin{pmatrix} \bar{a}_{11,H} & 0 \\ 0 & \bar{a}_{22,H} \end{pmatrix},$$

where for any matrix A , a_{ij} denotes its (i, j) -th element.

To better understand the dynamics, we express (5)-(6) in vector form. Define $p_t := \text{vech}(P_t)$, $v_t := \text{vech}(V_t)$, $h_t := \text{vech}(H_t)$ and $m_t := \text{vech}(M_t)$, where the *vech* operator stacks the lower triangular part including the main diagonal of a $(k \times k)$ symmetric matrix into a $(k^* \times 1)$ vector, $k^* = k(k+1)/2$. These $(k^* \times 1)$ vectors retain the unique elements of the matrices of interest to us. An equivalent representation of (3)-(4) is

$$P_t = H_t + H_t^{\frac{1}{2}}(\varepsilon_t - I_k)H_t^{\frac{1}{2}}, \quad V_t = M_t + M_t^{\frac{1}{2}}(\eta_t - I_k)M_t^{\frac{1}{2}},$$

which, using the *vech* notation, can be expressed as

$$p_t = h_t + \tilde{\varepsilon}_t, \quad v_t = m_t + \tilde{\eta}_t,$$

where $\tilde{\varepsilon}_t = \text{vech}(H_t^{\frac{1}{2}}(\varepsilon_t - I_k)H_t^{\frac{1}{2}}) = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}})D_k \text{vech}(\varepsilon_t - I_k)$ and $\tilde{\eta}_t = \text{vech}(M_t^{\frac{1}{2}}(\eta_t - I_k)M_t^{\frac{1}{2}}) = L_k(M_t^{\frac{1}{2}} \otimes M_t^{\frac{1}{2}})D_k \text{vech}(\eta_t - I_k)$.² The matrices L_k and D_k are, respectively, the elimination and duplication matrices defined in Appendix A. This representation is particularly convenient since $\tilde{\varepsilon}_t$ and $\tilde{\eta}_t$ are a vector martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$.

Similarly, (5)-(6) can be written as

$$h_t = C_H + B_H h_{t-1} + A_H v_{t-1}, \tag{7}$$

$$m_t = C_M + B_M m_{t-1} + A_M v_{t-1}, \tag{8}$$

where $C_H = L_k(\bar{C}_H \otimes \bar{C}_H)D_k \text{vech}(I_k)$, $B_H = L_k(\bar{B}_H \otimes \bar{B}_H)D_k$ and $A_H = L_k(\bar{A}_H \otimes \bar{A}_H)D_k$. C_M , B_M , and A_M are defined similarly using the parameters of (6). C_H and C_M are $(k^* \times 1)$ vectors, while A_H , B_H , A_M and B_M are $(k^* \times k^*)$ matrices. The elimination and duplication matrices, L_k and D_k , are non-stochastic matrices of zeros and ones, so the parameters in (7)-(8) are uniquely identified from (5)-(6) and vice versa.

By substituting $h_t = p_t - \tilde{\varepsilon}_t$ and $m_t = v_t - \tilde{\eta}_t$ into (7)-(8), it is straightforward to show

²The second equality in each expression follows from the property that for any $(k \times k)$ matrices A and B , with B being symmetric, $\text{vech}(ABA') = L_k(A \otimes A)D_k \text{vech}(B)$; see (A.1) in Appendix A.

that the HEAVY model has the following VARMA(1,1) representation

$$\begin{pmatrix} p_t \\ v_t \end{pmatrix} = \begin{pmatrix} C_H \\ C_M \end{pmatrix} + \begin{pmatrix} B_H & A_H \\ 0 & B_M + A_M \end{pmatrix} \begin{pmatrix} p_{t-1} \\ v_{t-1} \end{pmatrix} + \begin{pmatrix} \tilde{\varepsilon}_t \\ \tilde{\eta}_t \end{pmatrix} - \begin{pmatrix} B_H & 0 \\ 0 & B_M \end{pmatrix} \begin{pmatrix} \tilde{\varepsilon}_{t-1} \\ \tilde{\eta}_{t-1} \end{pmatrix}$$

since $(\tilde{\varepsilon}'_t, \tilde{\eta}'_t)'$ is a vector martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$, assuming $\text{Var}[(\tilde{\varepsilon}'_t, \tilde{\eta}'_t)']$ exists. The coefficient matrix attached to $(p'_{t-1}, v'_{t-1})'$ determines the persistence of the HEAVY system. For covariance stationarity, the eigenvalues of this matrix must be less than one in modulus. Since it is block triangular, its eigenvalues are members of the multiset of the eigenvalues of B_H and $(B_M + A_M)$.³ In the following assumption we explicitly state this covariance stationarity condition, where for any $(k \times k)$ matrix A with eigenvalues $\lambda_1, \dots, \lambda_k$, $\rho(A) := \max_i |\lambda_i|$ denotes the spectral radius of A .

Assumption 1 *In the HEAVY model given by (7)-(8), $\rho(B_H) < 1$ and $\rho(B_M + A_M) < 1$.*

The covariance stationarity condition in Assumption 1 is analogous to the one given in Engle and Kroner (1995). This can be seen by noting that for any square matrix A , $D_k^+(A \otimes A)D_k$ and $(A \otimes A)$ have the same eigenvalues, where $D_k^+ = (D_k' D_k)^{-1} D_k'$ is the Moore-Penrose inverse of D_k ; see Magnus (1988, Theorem 4.10). Also, it holds that for any square matrix A , $D_k^+(A \otimes A)D_k = L_k(A \otimes A)D_k$; see Lütkepohl (1996, Section 9.5.5). Thus $B_H = L_k(\bar{B}_H \otimes \bar{B}_H)D_k$ and $(\bar{B}_H \otimes \bar{B}_H)$ have the same eigenvalues. A similar argument applies to $(B_M + A_M)$.

We can express the unconditional first moments of p_t and v_t in terms of the model parameters. By taking unconditional expectation of (7)-(8), it is straightforward to show

³A multiset is a set that allows for some or all of its elements to be repeated. This general definition is needed to allow for the case when B_H and $(B_M + A_M)$ have some common eigenvalues.

that

$$\omega_H := \mathbb{E}[p_t] = (I_{k^*} - B_H)^{-1} [C_H + A_H(I_{k^*} - B_M - A_M)^{-1} C_M], \quad (9)$$

$$\omega_M := \mathbb{E}[v_t] = (I_{k^*} - B_M - A_M)^{-1} C_M. \quad (10)$$

In Appendix B, we derive the unconditional second moments of p_t and v_t , which correspond to the fourth moments of the returns (i.e. kurtosis) and second moments of the realised measure (i.e. volatility of volatility).

2.3 Covariance Targeting

The covariance targeting parameterisation was introduced by Engle and Mezrich (1996) for the univariate GARCH model. This allows the unconditional moments of the model to be estimated by the empirical moments, and the dynamic parameters would then be estimated using a quasi-likelihood. The HEAVY model differs from ARCH-type models by using a shock other than the outer-product of returns to model the conditional covariance. This has an implication for the covariance targeting specification when the dynamics of the model are restricted from the full specification in (5), as is the case when $\bar{A}_H$ is assumed to be diagonal or scalar. We elaborate on this point after the following proposition, which gives two covariance targeting parameterisations of the HEAVY model.

Proposition 1 *Let $\Omega_H := \mathbb{E}[P_t] = \mathbb{E}[H_t]$ and $\Omega_M := \mathbb{E}[V_t] = \mathbb{E}[M_t]$. The covariance targeting parameterisation of the HEAVY model in (7)-(8) is*

$$h_t = (I_{k^*} - B_H - A_H \kappa) \omega_H + B_H h_{t-1} + A_H v_{t-1}, \quad (11)$$

$$m_t = (I_{k^*} - B_M - A_M) \omega_M + B_M m_{t-1} + A_M v_{t-1}, \quad (12)$$

where $\kappa = L_k(\bar{\kappa} \otimes \bar{\kappa})D_k$, $\bar{\kappa} = \Omega_M^{\frac{1}{2}}\Omega_H^{-\frac{1}{2}}$, $\omega_H := \text{vech}(\Omega_H)$, $\omega_M := \text{vech}(\Omega_M)$, and L_k and D_k denote respectively the elimination and duplication matrices of order k . An alternative covariance targeting parameterisation for (7) is

$$h_t = (I_{k^*} - B_H - A_H^*)\omega_H + B_H h_{t-1} + A_H^* \tilde{v}_{t-1}, \quad (13)$$

where $\tilde{v}_t = \kappa^{-1}v_t$ is a rotated realised measure such that $E[\tilde{v}_t] = \omega_H$.

While the covariance targeting specification in (11)-(12) is a reparameterisation of the original model in (7)-(8), the specification (13)-(12) corresponds to a different model which uses a rotated rather than the original realised measure. This is why the coefficient matrix on $\tilde{v}_{t-1}$ is now denoted by A_H^* . The two models are equivalent, implying $A_H^* = A_H \kappa$ holds, if and only if both A_H^* and A_H are fully parameterised matrices. When A_H and A_H^* are restricted to be scalar (diagonal), this equivalence does not hold unless $\kappa \propto I_k$ (κ is diagonal).

Using (13)-(12) has the advantage that it is easier to impose the condition $\rho(B_H + A_H^*) < 1$ during estimation; see Assumption 2 below. Imposing the condition $\rho(B_H + A_H \kappa) < 1$ is more involved, particularly in the diagonal and full HEAVY models since κ is a $(k^* \times k^*)$ matrix with non-zero elements. For covariance stationarity in (11)-(12), or alternatively (13)-(12), we replace Assumption 1 with the following assumption.

Assumption 2 *In the covariance targeting parameterisation of the HEAVY model given by (11)-(12), $\rho(B_H + A_H \kappa) < 1$ and $\rho(B_M + A_M) < 1$. In the covariance targeting parameterisation of the HEAVY model given by (13)-(12), $\rho(B_H + A_H^*) < 1$ and $\rho(B_M + A_M) < 1$.*

Estimating the model in its covariance targeting specification can be carried out in two steps, and we discuss the appropriate inference method in this case in Section 3.3.

2.4 Multi-Step Forecasting

We are primarily interested in forecasting the conditional covariance of daily returns, H_t . One-step forecasts are directly computable using (7), which expresses H_t in its *vech* form. To compute s -step forecasts for $s = 2, 3, \dots$, we need the forecasts from (8) as well to compute the s -step conditional expectation of the realised measure appearing in the right-hand side of (7). The s -step forecast of h_t is given in the following proposition.

Proposition 2 *Let the model be given by (7)-(8), then the s -step forecast of h_t is*

$$\begin{aligned} E_t[h_{t+s}] &= \sum_{i=1}^{s-1} B_H^{i-1} C_H + B_H^{s-1} h_{t+1} \\ &\quad + \sum_{i=1}^{s-1} B_H^{i-1} A_H \left\{ \sum_{j=1}^{s-i-1} (B_M + A_M)^{j-1} C_M + (B_M + A_M)^{s-i-1} m_{t+1} \right\} \end{aligned} \quad (14)$$

where h_{t+1} and m_{t+1} are $\mathcal{F}_t^{HF}$ -measurable. Alternatively, let the model be given by (11)-(12), then the s -step forecast of h_t is

$$E_t[h_{t+s}] = \omega_H + B_H^{s-1} (h_{t+1} - \omega_H) + \sum_{i=1}^{s-1} B_H^{i-1} A_H (B_M + A_M)^{s-i-1} (m_{t+1} - \omega_M). \quad (15)$$

The difference between (14) and (15) is that the latter is obtained under a covariance targeting specification in which the constant terms C_H and C_M are replaced with expressions involving ω_H and ω_M ; see Section 2.2. In (14), Assumption 1 implies $E_t[h_{t+s}] \rightarrow \omega_H$ as $s \rightarrow \infty$ since the coefficients on h_{t+1} and m_{t+1} will tend to zero, while the limit of the constant terms including C_H and C_M will be the right-hand side of (9). In (15), we also

have that $E_t[h_{t+s}] \rightarrow \omega_H$ as $s \rightarrow \infty$; however, in this case Assumption 2 is the operative assumption since the derivation of this equation is based on the covariance targeting specification.

In deriving (15), we focused on the covariance targeting specification given by (11)-(12) since it is more constructive to study the properties of the HEAVY model forecasts. For example, (15) can be used to compute the HEAVY model's half-life (of a deviation of the 1-step forecast of h_t from ω_H) and compare it to that of the GARCH model. The presence of the term $(m_{t+1} - \omega_M)$ also indicates that mean reversion of the forecast matrix is not necessarily monotonic. To forecast using the covariance targeting specification in (13)-(12), A_H^* will appear in (15) instead of A_H . Thus the term $(m_{t+1} - \omega_M)$ must be pre-multiplied by κ^{-1} to ensure positive definiteness of $E_t[H_{t+s}]$.

3 Estimation and Inference

3.1 The Distribution of ε_t and η_t

For the HEAVY model in (3)-(4),

$$P_t = H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}}, \quad V_t = M_t^{\frac{1}{2}} \eta_t M_t^{\frac{1}{2}},$$

the natural choice for the density of the innovation matrices, ε_t and η_t , is the Wishart distribution. It is an appropriate choice in models where the support of the random variable of interest is restricted to the space of positive semidefinite matrices.⁴ Appendix C provides an overview of the Wishart distribution including the definitions and notation

⁴Some recent multivariate stochastic volatility models also employ the Wishart distribution to model time-varying correlations; see Chib et al. (2009) and the references cited therein, and also Jin and Maheu (2010).

used in this section.

In GARCH models, the vector of daily returns is usually modelled as $R_t = H_t^{\frac{1}{2}} \xi_t$ with $\xi_t \stackrel{i.i.d.}{\sim} N(0, I_k)$, which motivates quasi-maximum likelihood estimation (QMLE). For the HEAVY-P equation, we have $P_t = R_t R_t' = H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}}$, where $\varepsilon_t = \xi_t \xi_t'$. The assumption that $\xi_t \stackrel{i.i.d.}{\sim} N(0, I_k)$ implies that ε_t follows a Wishart distribution.

One of the key results on the Wishart distribution is that if any matrix $S \sim W_k(n, \Sigma)$, then $ASA' \sim W_k(n, A\Sigma A')$ for any $(k \times k)$ nonsingular matrix A . Assuming a Wishart density for ε_t and η_t implies that P_t and V_t are assumed to be conditionally Wishart distributed. However, one distinction between the densities of ε_t and η_t relates to the differences in the ranks of P_t and V_t . The matrix $P_t = R_t R_t'$ has rank 1 by construction if there is at least one non-zero return in R_t . Whether using the realised covariance estimator or the realised kernel of Barndorff-Nielsen et al. (2011), the matrix V_t is guaranteed to be full rank under standard regularity conditions, provided that $k < m$, where m is the number of intra-daily returns. This difference in rank entails that ε_t should have a singular Wishart density and η_t a standardised Wishart density. The discussion in Appendix C makes it clear that this distinction is necessary for the two conditional moment assumptions, $E_{t-1}[\varepsilon_t] = I_k$ and $E_{t-1}[\eta_t] = I_k$, to be satisfied.

Therefore, we assume $\varepsilon_t \stackrel{i.i.d.}{\sim} SINGW_k(1, I_k)$ and $\eta_t \stackrel{i.i.d.}{\sim} SW_k(n, I_k)$. The densities of ε_t and η_t are given by, respectively, (C.2) and (C.1) in Appendix C. Thus $P_t | \mathcal{F}_{t-1}^{HF} \sim SINGW_k(1, H_t)$ and $V_t | \mathcal{F}_{t-1}^{HF} \sim SW_k(n, M_t)$. The distinction between the densities of ε_t and η_t is of no consequence to QMLE as we show in a moment. However, it is needed to have a correctly specified model satisfying $E_{t-1}[\varepsilon_t] = E_{t-1}[\eta_t] = I_k$.⁵

⁵One can test for the Wishart distribution assumption by making use of the property that if $S \sim W_k(n, \Sigma)$, then $\frac{a'Sa}{a'\Sigma a} \sim \chi_{(n)}^2$ for any $(k \times 1)$ vector $a \neq 0$; see Gupta and Nagar (2000). Also, conditional moment tests can be used to detect misspecification.

3.2 Quasi-Maximum Likelihood Estimation

The HEAVY model is parameterised with a finite-dimensional $(\delta \times 1)$ parameter vector $\theta \in \Theta \subset \mathbb{R}^\delta$. Decompose $\theta = (\theta'_H, \theta'_M)'$ where the $(\delta_H \times 1)$ vector θ_H and $(\delta_M \times 1)$ vector θ_M denote the parameter vectors of the HEAVY-P and HEAVY-V equations, respectively. Let $\theta_0 = (\theta'_{H,0}, \theta'_{M,0})'$ denote the true parameter vector. The log-likelihood for the t -th observation will be denoted by $l_{H,t}(\theta_H)$ and $l_{M,t}(\theta_M)$. Inference for the HEAVY model will be based on QMLE of the following two log-likelihood functions

$$l_{H,t}(\theta_H) = c_H - \frac{1}{2} (\log |H_t| + \text{tr}(H_t^{-1} P_t)), \quad l_{M,t}(\theta_M) = c_M - \frac{n}{2} (\log |M_t| + \text{tr}(M_t^{-1} V_t)),$$

where c_H and c_M are constants with respect to θ_H and θ_M ; see, respectively, (C.2) and (C.1) in Appendix C. Thus the distinction between the densities of ε_t and η_t is of no consequence for QMLE of the model parameters. Engle and Gallo (2006) argue similarly for the Gamma density where the shape parameter is of no consequence when estimating the scale parameter by QMLE.

We assume the initial values, H_0 and M_0 , are known and are positive semidefinite. We also assume that θ_H and θ_M are variation free in the sense of Engle et al. (1983), which allows for equation-by-equation estimation. This assumption is not essential and is only used to simplify estimation and inference. The QML estimator is $\hat{\theta} = (\hat{\theta}'_H, \hat{\theta}'_M)'$ where

$$\hat{\theta}_H = \arg \max_{\theta_H \in \Theta} L_H(\theta_H), \quad \hat{\theta}_M = \arg \max_{\theta_M \in \Theta} L_M(\theta_M),$$

$$\text{and } L_H(\theta_H) = \sum_{t=1}^T l_{H,t}(\theta_H), \quad L_M(\theta_M) = \sum_{t=1}^T l_{M,t}(\theta_M).$$

For the BEKK model, Comte and Lieberman (2003) show strong consistency of QMLE

by verifying the conditions given in Jeantheau (1998). Hafner and Preminger (2009) show similar results for the VEC model which nests the BEKK model, and their results also apply to integrated processes. An important condition to establish strong consistency is for the model to admit a strictly stationary and ergodic solution, which we assume for the HEAVY model.

Before discussing the asymptotic distribution of $\widehat{\theta}$, we first give results on the score vector in the following proposition. It will be convenient to consider the score for each equation separately.

Proposition 3 (i) *The score vectors, $S_{H,t}(\theta_H) = \frac{\partial l_{H,t}(\theta_H)}{\partial \theta'_H}$ and $S_{M,t}(\theta_M) = \frac{\partial l_{M,t}(\theta_M)}{\partial \theta'_M}$ of dimensions $(1 \times \delta_H)$ and $(1 \times \delta_M)$, respectively, are given by*

$$S_{H,t}(\theta_H) = \frac{\partial l_{H,t}(\theta_H)}{\partial \theta'_H} = \frac{1}{2} [(vec(P_t))' - (vec(H_t))'] (H_t^{-1} \otimes H_t^{-1}) \frac{\partial vec(H_t)}{\partial \theta'_H}, \quad (16)$$

$$S_{M,t}(\theta_M) = \frac{\partial l_{M,t}(\theta_M)}{\partial \theta'_M} = \frac{1}{2} [(vec(V_t))' - (vec(M_t))'] (M_t^{-1} \otimes M_t^{-1}) \frac{\partial vec(M_t)}{\partial \theta'_M}. \quad (17)$$

(ii) *Under $E_{t-1}[\varepsilon_t] = I_k$ and $E_{t-1}[\eta_t] = I_k$, the score vectors evaluated at the true parameter value are a martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$.*

The scores have a similar structure to those of GARCH models (e.g. Bollerslev and Wooldridge (1992)). In analogy with generalised least squares, the terms in square brackets can be considered "errors", while $(H_t^{-1} \otimes H_t^{-1})$ and $(M_t^{-1} \otimes M_t^{-1})$ are weights and the derivatives $\frac{\partial vec(H_t)}{\partial \theta'_H}$ and $\frac{\partial vec(M_t)}{\partial \theta'_M}$ are instruments which are orthogonal to the errors at the maximum likelihood estimator, which is a condition for consistency.

To discuss the asymptotic distribution of the QML estimator, $\widehat{\theta}$, we define the $(1 \times \delta)$ combined score vector $S_t(\theta) = (S_{H,t}(\theta_H), S_{M,t}(\theta_M))$. Having established that the scores

are a martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$, it can be shown under certain regularity conditions (e.g. Comte and Lieberman (2003)) that

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1} \mathcal{J} \mathcal{I}^{-1}),$$

where

$$\mathcal{J} = \text{E} [S_t(\theta)' S_t(\theta)] = \text{E} \begin{bmatrix} \frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H} \frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'} & \frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H} \frac{\partial l_{M,t}(\theta_M)}{\partial \theta_M'} \\ \frac{\partial l_{M,t}(\theta_M)}{\partial \theta_M} \frac{\partial l_{H,t}(\theta_H)}{\partial \theta_H'} & \frac{\partial l_{M,t}(\theta_M)}{\partial \theta_M} \frac{\partial l_{M,t}(\theta_M)}{\partial \theta_M'} \end{bmatrix},$$

$$\mathcal{I} = -\text{E} \left[\frac{\partial S_t(\theta)}{\partial \theta} \right] = -\text{E} \begin{bmatrix} \frac{\partial^2 l_{H,t}(\theta_H)}{\partial \theta_H \partial \theta_H'} & 0 \\ 0 & \frac{\partial^2 l_{M,t}(\theta_M)}{\partial \theta_M \partial \theta_M'} \end{bmatrix}.$$

The block diagonality of the Hessian, $\mathcal{I}$, is due to the assumption that θ_H and θ_M are variation free, which implies that equation-by-equation standard errors are correct for the HEAVY system. With covariance targeting, a two-step estimation procedure is adopted and in this case the score vector will no longer be a martingale difference sequence, but it will have mean zero at the true parameter value. Also, the Hessian will not be block diagonal due to accounting for the accumulation of estimation error from the first step. We formalise inference in the case of covariance targeting in the following subsection.

3.3 Two-Step Estimation Under Covariance Targeting

With covariance targeting, the parameter vectors θ_H and θ_M are decomposed into $\theta_H = (\omega_H', \tilde{\theta}_H')'$ and $\theta_M = (\omega_M', \tilde{\theta}_M')'$ and are to be estimated in two steps. The unconditional moments, ω_H and ω_M , will be estimated in the first step by a moment estimator

$$\hat{\omega}_H = T^{-1} \sum_{t=1}^T p_t, \quad \hat{\omega}_M = T^{-1} \sum_{t=1}^T v_t,$$

and then $\tilde{\theta}_H$ and $\tilde{\theta}_M$ will be estimated by QMLE in the second step. The asymptotics of the QML estimator in this case is a direct application of two-step GMM estimation discussed in Newey and McFadden (1994). Define $\tilde{l}_{H,t}(\omega_H, \omega_M, \tilde{\theta}_H)$ and $\tilde{l}_{M,t}(\omega_M, \tilde{\theta}_M)$ to be the t -th observation log-likelihoods for the covariance targeting HEAVY model. Two-step estimation gives the following $(1 \times \delta)$ vector of moment conditions

$$\tilde{S}_t(\tilde{\theta}) = \left((p_t - \omega_H)', \frac{\partial \tilde{l}_{H,t}}{\partial \tilde{\theta}_H'}, (v_t - \omega_M)', \frac{\partial \tilde{l}_{M,t}}{\partial \tilde{\theta}_M'} \right)', \quad \tilde{\theta} = (\omega_H', \tilde{\theta}_H', \omega_M', \tilde{\theta}_M')',$$

which is no longer martingale difference sequence with respect to $\mathcal{F}_{t-1}^{HF}$. In this case

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1} \mathcal{J} (\mathcal{I}^{-1})'),$$

where

$$\mathcal{J} = \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{S}_t(\tilde{\theta}) \right],$$

$$\mathcal{I} = -\text{E} \left[\frac{\partial \tilde{S}_t(\tilde{\theta})}{\partial \tilde{\theta}} \right] = -\text{E} \begin{bmatrix} -I_{k^*} & \frac{\partial^2 \tilde{l}_{H,t}}{\partial \omega_H \partial \tilde{\theta}_H'} & 0 & 0 \\ 0 & \frac{\partial^2 \tilde{l}_{H,t}}{\partial \tilde{\theta}_H \partial \tilde{\theta}_H'} & 0 & 0 \\ 0 & \frac{\partial^2 \tilde{l}_{H,t}}{\partial \omega_M \partial \tilde{\theta}_H'} & -I_{k^*} & \frac{\partial^2 \tilde{l}_{M,t}}{\partial \omega_M \partial \tilde{\theta}_M'} \\ 0 & 0 & 0 & \frac{\partial^2 \tilde{l}_{M,t}}{\partial \tilde{\theta}_M \partial \tilde{\theta}_M'} \end{bmatrix}.$$

In implementation we use a HAC estimator, e.g. Newey and West (1987), to estimate $\mathcal{J}$. With covariance targeting, variation freeness between the parameters of the HEAVY-P and HEAVY-V equations no longer holds since κ depends on ω_M . Thus the block $\frac{\partial^2 \tilde{l}_{H,t}}{\partial \omega_M \partial \tilde{\theta}_H'}$

now appears in the Hessian to account for this dependence in the second step of estimation.

4 Model Evaluation

For out-of-sample model evaluation, we use a quasi-likelihood (QLIK) loss function of the form

$$L_{t,s}(\Sigma_{t+s}, H_{t+s|t}^a) = \log \left| H_{t+s|t}^a \right| + \text{tr}((H_{t+s|t}^a)^{-1} \Sigma_{t+s}), \quad (18)$$

where Σ_{t+s} is the actual (unobserved) covariance matrix and $H_{t+s|t}^a$ denotes its s -step forecast using model a conditional on time t information. Since Σ_{t+s} is unobservable, our analysis will be based on some proxy denoted by $\hat{\Sigma}_{t+s}$, which we take to be the realised covariance matrix, V_{t+s} . The loss function (18) evaluates the s -step predicted density from model a using the proxy $\hat{\Sigma}_{t+s}$ as data⁶, and it provides a consistent ranking of volatility models in the sense of Patton (2011) and Patton and Sheppard (2009) as it is robust to noise in the proxy $\hat{\Sigma}_{t+s}$; see also Laurent et al. (2009).

Note that even if - at time t - the true density of R_{t+1} is normal (i.e. the density of P_{t+1} is Wishart), normality will not hold under temporal aggregation unless the conditional covariance matrix is constant. Therefore the s -step density will not be normal implying that the density used for the QLIK loss function (18) is misspecified. However, the loss difference between two competing models a and b , $L_{t,s}(\Sigma_{t+s}, H_{t+s|t}^a) - L_{t,s}(\Sigma_{t+s}, H_{t+s|t}^b)$, can be interpreted as a Kullback-Leibler distance which yields a valid assessment even if both models are misspecified. Cox (1961) proposes a likelihood ratio test based on this idea, while Vuong (1989) provides the theoretical framework in the case of nested and

⁶Note that (18) is the negative of the log-likelihood of a multivariate normal density excluding the constant terms. The switched sign is due to defining (18) as a "loss" function.

non-nested models. Similar approaches are proposed for out-of-sample model selection in Amisano and Giacomini (2007) and Diks et al. (2008).

We denote the loss difference between the HEAVY and GARCH models by

$$D_{t,s} = L_{t,s}(\Sigma_{t+s}, H_{t+s|t}^{HEAVY}) - L_{t,s}(\Sigma_{t+s}, H_{t+s|t}^{GARCH}), \quad t = Q, Q+1, \dots, T-s,$$

where $L_{t,s}(\cdot)$ is given by (18), T is the size of the full sample and Q is the size of the estimation window. We assume Q is fixed so that we use a rolling-window of data to estimate the model parameters, which gives $T - Q - s + 1$ data points for out-of-sample model evaluation. The average loss is denoted by

$$\overline{D}_s = \frac{1}{T - Q - s + 1} \sum_{t=Q}^{T-s} D_{t,s}$$

which is used to test $H_0 : E[D_{t,s}] = 0$, for all s , against a two-sided alternative. Let $\overline{D}_s^*$ denote the average loss evaluated at the true parameter value, then we have

$$\sqrt{T}(\overline{D}_s - \overline{D}_s^*) \xrightarrow{d} N(0, \Lambda_s),$$

where Λ_s is the asymptotic variance of $D_{t,s}$ estimated using a HAC estimator. Significantly negative values of the test statistic indicate superior forecast performance of the HEAVY model. This predictive ability test was first introduced by Diebold and Mariano (1995), and later formalised by West (1996) and Giacomini and White (2006).

We extend this strategy in the context of multivariate volatility models by conducting separate tests for forecasts of the individual variances and also for the dependence structure of the group of assets under consideration. Consider the margins-copula decomposition of

the log-likelihood of R_t ,

$$\log f(X) = \sum_{i=1}^k \log f_i(x_i) + \log c(F_1(x_1), F_2(x_2), \dots, F_k(x_k)), \quad (19)$$

where $f(X)$ is the joint density of the returns of the k assets, $f_i(x_i)$ and $F_i(x_i)$, $i = 1, \dots, k$, are respectively the density and cumulative distribution function of asset i returns, and $c(\cdot)$ is the copula density.⁷ The normality assumption for R_t implies that $f(X)$, $f_i(x_i)$ and $c(\cdot)$ correspond to the multivariate normal density, normal density and normal copula, respectively.

We decompose the QLIK loss in (18) in a similar fashion to (19). So computing the loss in (18) based on the whole forecast matrix ($H_{t+s|t}^a$) corresponds to $\log f(X)$, while computing the loss based on a particular diagonal element of $H_{t+s|t}^a$, say $h_{ii,t+s|t}^a$, corresponds to $\log f_i(x_i)$. The latter corresponds to the loss encountered in forecasting the individual variance for asset i , and we compute it for all k assets. We compute the loss attributed to forecasting the dependence structure (summarised by the copula contribution) as the residual, i.e. corresponding to $\log f(X) - \sum_{i=1}^k \log f_i(x_i)$. Based on this QLIK loss decomposition, we conduct the predictive ability test, outlined above, separately for each margin as well as the copula. Due to the normality assumption, the copula parameter is the conditional correlation matrix of the daily returns, thus we use the terms margins-copula and variances-correlations interchangeably.

⁷Nelsen (2006) and Patton (2009) provide recent reviews of copula theory and financial applications.

5 Empirical Application

We use high-frequency data on Spyder (SPY), the S&P 500 ETF, along with some of the most liquid stocks in the Dow Jones Industrial Average (DJIA) index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). The sample period is 1/2/2001 to 31/12/2009 with a total of 2242 trading days, and the data source is the TAQ database. We choose the starting date for the sample to be after decimal pricing had been fully implemented in the NYSE, which took place on 29/1/2001.

We focus on the realised covariance matrix as our choice for V_t . In computing the realised covariance matrix, we use 5-minute returns with subsampling. We exclude the opening and closing 15 minutes of trading to control for overnight effects. For the daily return, we focus on the open-to-close returns which of course ignore overnight effects, and for consistency with the realised covariance estimator we compute the open-to-close daily returns over the same interval.⁸ Our estimation and forecast evaluation computations were repeated using the noise-robust realised kernel of Barndorff-Nielsen et al. (2011) with the results being qualitatively similar in general.⁹

The main focus of our empirical application will be on modelling and forecasting the conditional covariance matrix of a stock (BAC) and an index (S&P 500) using the scalar HEAVY model. Most of the model's features can be readily seen in this bivariate model which is analysed in Section 5.1.1. In Section 5.1.2, we analyse other pairs of assets using

⁸We also estimated some of the models using close-to-close returns. The differences in results are discussed at the end of Section 5.1.2.

⁹These are not reported in the interest of parsimony, but are available upon request.

the scalar HEAVY model. In Section 5.2, we report estimation and forecast evaluation results for the bivariate diagonal HEAVY model and highlight differences from the scalar HEAVY case. Finally, in Section 5.3 we report estimates of the scalar HEAVY model for the 10 DJIA stocks using covariance targeting.

5.1 Bivariate Scalar HEAVY Model

5.1.1 S&P 500 and Bank of America

Figure 1 contains the annualised realised volatility of SPY and BAC, their realised correlation and realised beta for BAC over the full sample. The sharp increase in volatility in 2008-2009 is associated with the turmoil in financial markets during the recent financial crisis. The increase in BAC volatility is much more pronounced especially after the collapse of Lehman Brothers in mid September 2008. BAC realised correlation with the market seems to have been relatively high during the crisis, and its realised beta increased sharply and was very volatile during this period.

In Table 1, we present the HEAVY and GARCH model estimates. We also report estimates for the GARCH-X model which is similar to (7) with p_{t-1} included on the right-hand side with coefficient D_{GX} . So the GARCH-X model nests both the HEAVY-P equation and the GARCH model. For ease of interpretation, we only report the parameter estimates for the models' *vech* representation excluding the constant terms.

The estimate of B_H implies that the elements of H_t will be smooth, although less smooth than the corresponding estimates from the GARCH model with the estimate of B_G equal to 0.934. For the HEAVY-V equation, the B_M coefficient is relatively small implying that the estimated conditional moments will be somewhat erratic. In terms of magnitude, these

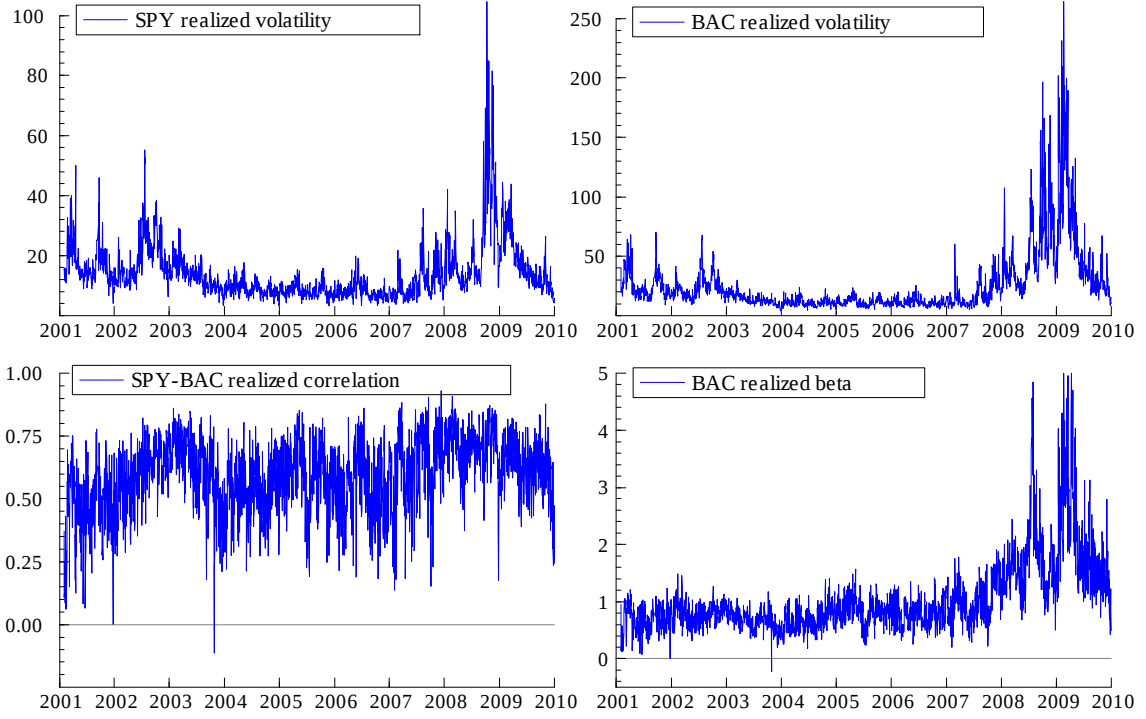


Figure 1: SPY and BAC annualised realised volatility, realised correlation and BAC realised beta.

estimates are largely in line with those from the univariate HEAVY model in Shephard and Sheppard (2010), and they also suggest a somewhat high level of persistence. Compared to the nesting GARCH-X model, there is no loss of fit when moving to HEAVY-P since the coefficient on p_{t-1} (D_{GX}) is not statistically significant. This is not the case when moving from GARCH-X to GARCH which suggests that v_{t-1} effectively crowds out p_{t-1} .

The estimates also suggest that the HEAVY model's half-life (of a deviation of the 1-step forecast of h_t from its long run) is substantially shorter than that of the GARCH model suggesting that the former's forecast responds faster to abrupt changes in the level of volatility or correlation.¹⁰

¹⁰The half-life can be easily computed from (15) by noting that the two gaps, $(h_{t+1} - \omega_H)$ and $(m_{t+1} - \omega_M)$, tend to have the same sign as our results indicate that the elements of h_t and m_t tend to be very highly correlated. Thus these two gaps can be set, without loss of generality, equal to a $(k^* \times 1)$ vector of ones.

	HEAVY-P		GARCH		GARCH-X			HEAVY-V	
	A_H	B_H	A_G	B_G	A_{GX}	B_{GX}	D_{GX}	A_M	B_M
SPY-BAC (st. error)	0.214 (0.054)	0.727 (0.068)	0.062 (0.010)	0.934 (0.011)	0.187 (0.056)	0.741 (0.068)	0.019 (0.012)	0.421 (0.033)	0.574 (0.033)
Log-likelihood decomposition (HEAVY-P versus GARCH)									
	HEAVY-P		GARCH		HEAVY-P gains				
Margin 1 (SPY)	-658		-713		55				
Margin 2 (BAC)	-1,593		-1,648		55				
Copula	815		808		7				
Joint distribution	-1,436		-1,553		117				
Predictive ability tests at different forecast horizons (days)									
	(1)	(2)	(3)	(5)	(10)	(22)			
Margin 1 (SPY)	-3.72	-3.03	-2.33	-1.23	0.84	1.87			
Margin 2 (BAC)	-3.27	-2.45	-1.70	-0.58	1.06	2.04			
Copula	-3.37	-3.22	-3.39	-3.28	-3.26	-3.85			
Joint distribution	-4.32	-3.78	-3.23	-2.33	-0.07	1.03			

Table 1: Scalar HEAVY estimation and forecast evaluation results for SPY-BAC. Top panel: parameter estimates of HEAVY, GARCH and GARCH-X with standard errors reported in parentheses. Middle panel: decomposition of the log-likelihood (excluding constant terms) at the estimated parameter values. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

The log-likelihood and its decomposition into margin and copula likelihoods in the middle panel of Table 1 indicate an improvement in fit of the HEAVY-P equation compared to the GARCH model. Note that the two models are non-nested so direct LR tests are not possible; however, we will present below the outcome of the predictive ability tests discussed in Section 4. Although non-nested, the decomposition suggests that the HEAVY-P equation improves on GARCH for both the margins and the copula. The model residuals, $\hat{\varepsilon}_t$ and $\hat{\eta}_t$, seem to be centered around the identity matrix, with the exception of two large outliers in $\hat{\eta}_t$ corresponding to the realised variances of SPY and BAC on 27/2/2007, due to the 9% fall in the Shanghai stock exchange index that day.

An interesting feature from the residual analysis is that it displays evidence of the leverage effect between the returns and the realised measure. This is shown in Figure 2. The upper-left chart shows the scatter plot of $\hat{\xi}_{1,t}$ and $\hat{\eta}_{11,t}$ which are the innovations to the daily return and realised variance of SPY, respectively.¹¹ The lower-left chart displays

¹¹ $\hat{\xi}_{1,t}$ is the first element of the vector $\hat{\xi}_t = \hat{H}_t^{-\frac{1}{2}} R_t$, and $\hat{\eta}_{11,t}$ is the (1,1) element of the matrix

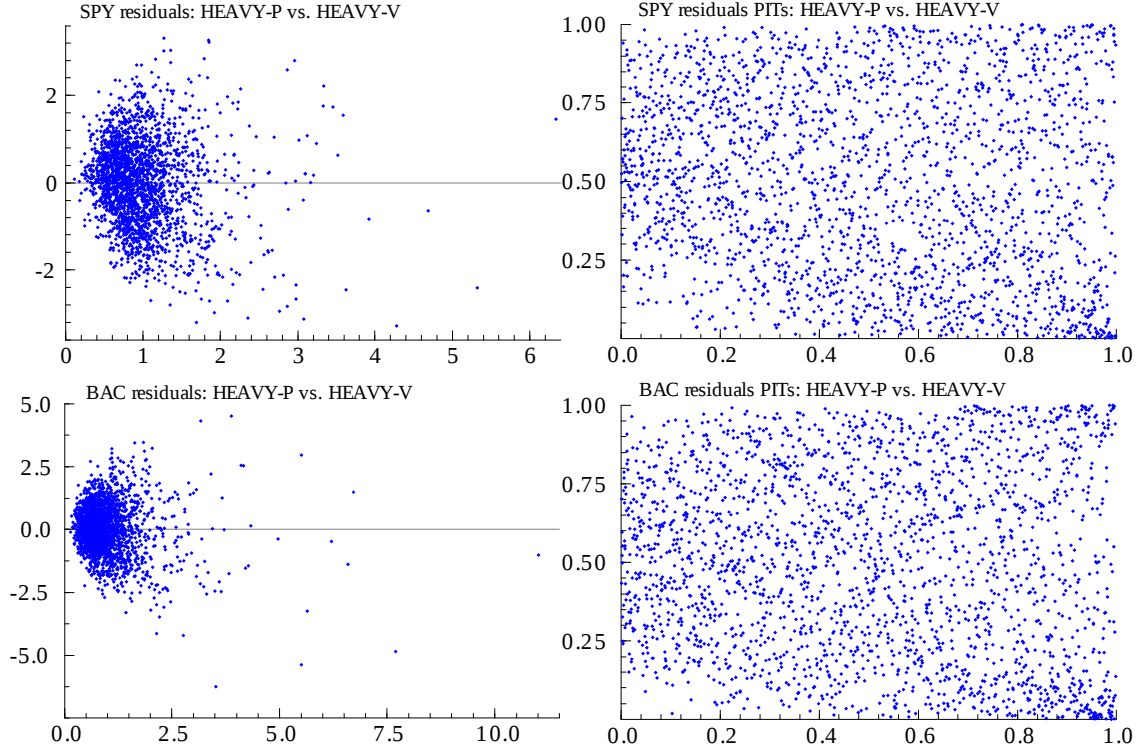


Figure 2: Left panel: scatter plots of SPY and BAC residuals in the HEAVY-P and HEAVY-V equations. Right panel: scatter plots of the residuals mapped into probability integral transforms (PITs).

the innovations to the daily return and realised variance of BAC. The right panel charts correspond to the same plots but mapped into copula space where the empirical distribution function is used to transform the innovations into probability integral transforms. The leverage effect can be seen in the right panel. For instance, large negative innovations to SPY returns tend to be associated with large positive innovations to its realised variance indicating higher volatility in response to bad news. The same applies to BAC innovations.

The bottom panel of Table 1 gives the results of the predictive ability tests. We estimate the model using a rolling-window of 1486 observations and then use the parameter estimates to obtain forecasts of H_t at horizons $s = 1, 2, 3, 5, 10, 22$ days using (14). The size of the rolling window is chosen such that our forecasts start at 3/1/2007. The reported figures

$$\hat{\eta}_t = \widehat{M}_t^{-\frac{1}{2}} V_t \widehat{M}_t^{-\frac{1}{2}}.$$

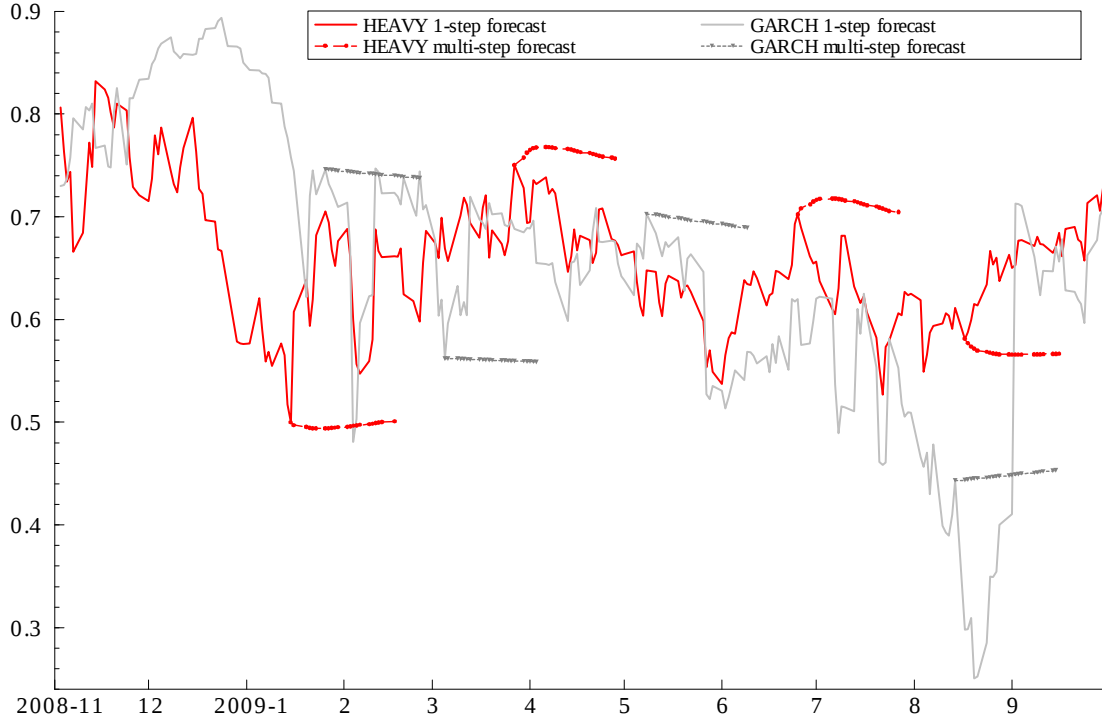


Figure 3: One-step and multi-step forecasts for the SPY-BAC conditional correlation.

are t -statistics to test equal predictive ability and significantly negative t -statistics favour the HEAVY model over the GARCH model. The results show that HEAVY outperforms GARCH especially at short forecast horizons. This is true for the whole covariance matrix forecast as well as its decomposition into margins and copula, which provides further insight into the source of forecast gains. The copula gains are maintained at longer forecast horizons indicating that the realised measure provides valuable information for forecasting the conditional correlation.

As pointed out earlier, the forecast profile of the HEAVY model is distinct from that of the GARCH(1,1) model particularly over short forecast horizons due to momentum effects. This can be seen in Figure 3 which plots the forecasts of the SPY-BAC conditional correlation (implied by the forecasts of H_t) over the period 03/11/2008 to 30/09/2009. This

is an interesting period for analysis as it marks a very volatile period during the 2007-2009 financial crisis. The solid lines are the 1-step forecasts, and at selected points we plot the forecast profile at this date for 22 days into the future. We do this only for selected peak and trough points for clarity of illustration. The momentum effects in the HEAVY model can be readily seen. Whereas the GARCH correlation forecast monotonically mean reverts, the HEAVY forecast displays some short run momentum influenced by the deviation of the realised measure from its long run before ultimately mean reverting. Interestingly, the plot also shows how the 1-step forecasts from both models diverge in some periods pointing to important differences in the information content of the realised measure and the outer product of daily returns.

It is interesting to track the model's performance in relation to the accuracy of the realised measure. For this purpose, we report in Table 2 the parameter estimates, log-likelihood gains and out-of-sample performance using various sampling intervals for the realised covariance estimator. The table also includes results when using the realised kernel as the realised measure. In general, the results indicate that when sampling between 5 and 15 minutes, the parameter estimates of the HEAVY and GARCH-X models are rather stable implying similar persistence levels, and indeed the estimates become very close when sampling at 30 minutes. At 1-minute sampling, there is substantial drop in the estimate of B_H and a moderate increase in A_H . Using the realised kernel leads to a noticeable decline in the smoothing parameters in both equations of the HEAVY model as well as the GARCH-X model. In terms of forecasting performance, the results are similar.

	HEAVY-P		GARCH-X			HEAVY-V		Joint LL
	A_H	B_H	A_{GX}	B_{GX}	D_{GX}	A_M	B_M	gain
RC 1-min	0.256	0.597	0.128	0.741	0.052	0.527	0.471	65
RC 5-min	0.214	0.727	0.187	0.741	0.019	0.421	0.574	117
RC 10-min	0.202	0.760	0.189	0.764	0.011	0.362	0.633	120
RC 15-min	0.185	0.787	0.169	0.793	0.012	0.300	0.696	116
RC 30-min	0.143	0.842	0.134	0.843	0.009	0.236	0.759	107
Realised kernel	0.213	0.677	0.194	0.689	0.015	0.508	0.488	128

Joint distribution predictive ability tests at different forecast horizons (days)						
	(1)	(2)	(3)	(5)	(10)	(22)
RC 1-min	-3.59	-3.08	-2.47	-1.82	0.30	1.50
RC 5-min	-4.32	-3.78	-3.23	-2.33	-0.08	1.03
RC 10-min	-4.08	-3.70	-3.21	-2.35	-0.23	0.99
RC 15-min	-4.26	-3.87	-3.34	-2.64	-0.26	1.11
RC 30-min	-3.81	-3.34	-2.75	-2.01	-0.13	1.12
Realised kernel	-4.25	-3.76	-3.23	-2.63	-0.46	1.00

Table 2: Scalar HEAVY estimation and forecast evaluation results for SPY-BAC using different realised measures. Top panel: scalar HEAVY and GARCH-X parameter estimates using different sampling intervals in computing the realised covariance and also using the realised kernel. Log-likelihood gains from the HEAVY model are reported in the last column. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

5.1.2 Other Asset Pairs

We estimate the scalar HEAVY model for other pairs of assets selected from the ten DJIA stocks. The pairing of the assets is chosen by selecting companies in the same sector (e.g. BAC-JPM and IBM-MSFT), where we expect more persistent correlation dynamics, and also pairs of companies in different sectors. The objective is to track the HEAVY model's performance in each case. Table 3 includes the parameter estimates for the HEAVY, GARCH and GARCH-X models. One notable feature is that the estimates do not display large variation across the different pairs. As in the SPY-BAC case, inclusion of the realised measure crowds out the outer product of returns as the coefficient D_{GX} is statistically insignificant in all cases.

The decomposition of the log-likelihood gains shows that the HEAVY model gains are obtained for each pair with respect to both margins and the copula with only one exception. The HEAVY model log-likelihood gain for the joint distribution is uniform across all pairs.

	HEAVY-P		GARCH		GARCH-X			HEAVY-V	
	A_H	B_H	A_G	B_G	A_{GX}	B_{GX}	D_{GX}	A_M	B_M
BAC - JPM	0.260	0.639	0.062	0.938	0.180	0.702	0.046	0.450	0.550
IBM - MSFT	0.179	0.762	0.051	0.941	0.135	0.794	0.026	0.309	0.676
XOM - AA	0.188	0.737	0.057	0.935	0.114	0.804	0.034	0.315	0.667
AXP - DD	0.201	0.743	0.045	0.951	0.183	0.754	0.010	0.357	0.638
GE - KO	0.220	0.727	0.039	0.957	0.205	0.740	0.007	0.344	0.651

	HEAVY-P log-likelihood gains/losses (+/-) relative to GARCH			
	Margin 1	Margin 2	Copula	Joint LL
BAC - JPM	44	58	8	110
IBM - MSFT	30	37	17	84
XOM - AA	44	31	-7	69
AXP - DD	56	49	16	121
GE - KO	47	34	9	90

	Joint distribution predictive ability tests at different forecast horizons (days)					
	(1)	(2)	(3)	(5)	(10)	(22)
BAC - JPM	-4.09	-3.41	-2.78	-1.98	0.45	1.32
IBM - MSFT	-2.92	-2.60	-2.38	-1.94	-1.56	-0.46
XOM - AA	-2.76	-2.09	-2.05	-1.76	-1.32	-0.80
AXP - DD	-3.46	-3.09	-2.84	-2.39	-1.28	0.18
GE - KO	-2.80	-2.66	-2.41	-2.21	-0.91	1.00

Table 3: Scalar HEAVY estimation and forecast evaluation results for other pairs of assets. Top panel: parameter estimates of HEAVY, GARCH and GARCH-X. BAC-JPM HEAVY-V estimates add up to 1 due to rounding. All coefficients, except D_{GX} , are statistically significant at the 5 percent significance level. Middle panel: HEAVY-P log-likelihood gains/losses relative to GARCH. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

The predictive ability test results indicate that the HEAVY model performs better than GARCH for all asset pairs, with the gains being particularly significant at short forecast horizons. We do not report the margins-copula decomposition for these tests in the interest of brevity, but they show that the HEAVY model gains are maintained for some of the margins and also for the copula of some of the pairs. In no case was the GARCH model significantly favoured at any horizon except for the BAC and JPM margins towards the end of the forecast horizon.

To investigate the sensitivity of the results to including overnight effects, we also estimated the scalar HEAVY model using close-to-close returns for SPY-BAC and the other asset pairs included in Table 3. The primary difference when using close-to-close returns is an increase in the loadings on the shock terms in both the HEAVY and GARCH models

through A_H , A_M and A_G , and particularly so for the GARCH model. The HEAVY model still provides gains for the joint and marginal log-likelihoods. The copula gains are obtained only for the pairs IBM-MSFT, AXP-DD and GE-KO. Interestingly, the predictive ability test results indicate that the HEAVY model gains for the joint log-likelihood are sustained at all horizons in most cases, which is also the case for some of the margins. The out-of-sample copula gains are significant at all horizons for the pairs IBM-MSFT and AXP-DD, only at longer horizons for SPY-BAC and BAC-JPM, and insignificant for XOM-AA and GE-KO.

5.2 Bivariate Diagonal HEAVY Model: SPY-BAC and Other Asset Pairs

We discuss the estimation and forecast evaluation results only for the diagonal HEAVY and GARCH models. We exclude the GARCH-X model results to improve presentation noting that its results are in line with those of the scalar model. The top panel of Table 4 presents estimates of the diagonal elements of the parameter matrices in (5)-(6), in order, along with those of the corresponding GARCH model. These are easier to interpret when expressed in terms of the parameters of the *vech* representation in (7)-(8), which are reported underneath. Note that if $\bar{A}_H$ is, for instance, a (2×2) diagonal matrix, then A_H will be a (3×3) diagonal matrix. The first and third diagonal elements of A_H will be the squares of the diagonal elements of $\bar{A}_H$, and the second diagonal element of A_H will be the product of the two diagonal elements of $\bar{A}_H$.

The estimates of the diagonal elements are rather similar within each parameter matrix, except for the HEAVY-V equation. Since the diagonal HEAVY model nests the scalar HEAVY model, we can test for the restriction using a Wald test. The scalar restriction is

	HEAVY-P				GARCH				HEAVY-V			
	$\bar{A}_H$	$\bar{B}_H$	$\bar{A}_G$	$\bar{B}_G$	$\bar{A}_M$	$\bar{B}_M$						
SPY-BAC (st. error)	0.447 (0.048)	0.477 (0.057)	0.858 (0.033)	0.844 (0.041)	0.238 (0.023)	0.262 (0.035)	0.968 (0.006)	0.964 (0.010)	0.632 (0.025)	0.663 (0.033)	0.768 (0.022)	0.748 (0.028)
	A_H	B_H	A_G	B_G	A_M	B_M						
Var. (SPY)	0.200	0.736	0.057	0.938	0.400	0.590						
Var. (BAC)	0.228	0.713	0.069	0.929	0.439	0.560						
Cov. eqn.	0.213	0.725	0.063	0.933	0.419	0.575						

	Log-likelihood decomposition (HEAVY-P versus GARCH)		
	HEAVY-P	GARCH	HEAVY-P gains
Marg. (SPY)	-659	-713	54
Marg. (BAC)	-1,592	-1,647	55
Copula	816	809	7
Joint dist.	-1,435	-1,552	117

	Predictive ability tests at different forecast horizons (days)					
	(1)	(2)	(3)	(5)	(10)	(22)
Marg. (SPY)	-3.86	-3.24	-2.77	-1.78	0.27	1.50
Marg. (BAC)	-3.17	-2.45	-1.79	-0.80	0.89	1.90
Copula	-3.29	-3.02	-3.04	-3.04	-2.99	-3.88
Joint dist.	-4.38	-3.79	-3.27	-2.54	-0.46	0.70

Table 4: Diagonal HEAVY estimation and forecast evaluation results for SPY-BAC. Top panel: parameter estimates of HEAVY and GARCH with standard errors reported in parentheses. Middle panel: decomposition of the log-likelihood (excluding constant terms) at the estimated parameter values. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

not rejected for both the HEAVY and GARCH models at 5 percent. The log-likelihood decomposition results are similar to the scalar model. The bottom panel of Table 4 shows that the diagonal HEAVY model provides superior forecasts with the gains being particularly significant at short forecast horizons.

We also report estimation results for the diagonal model using other pairs of assets in Table 5. For brevity, we only report parameter estimates for the *vech* representation. The parameter estimates show some variation within and across pairs. The Wald test results indicate that the scalar model restrictions are rejected at 5 percent only for the XOM-AA pair in the HEAVY-P equation, and only for the AXP-DD pair in the GARCH model. The scalar model restrictions for the HEAVY-V equation are not rejected in any of the pairs. The figures in the middle panel shows that the HEAVY model gains over GARCH in terms of the joint distribution log-likelihood are uniform across all pairs. The gains in the

	HEAVY-P		GARCH		HEAVY-V	
	A_H	B_H	A_G	B_G	A_M	B_M
Variance (BAC)	0.267	0.638	0.051	0.947	0.433	0.566
Variance (JPM)	0.256	0.634	0.074	0.925	0.473	0.526
Covariance (BAC-JPM)	0.262	0.636	0.061	0.936	0.452	0.546
Variance (IBM)	0.187	0.761	0.052	0.939	0.331	0.652
Variance (MSFT)	0.172	0.764	0.049	0.945	0.291	0.695
Covariance (IBM-MSFT)	0.180	0.763	0.050	0.942	0.311	0.673
Variance (XOM)	0.175	0.713	0.073	0.907	0.338	0.644
Variance (AA)	0.180	0.784	0.043	0.952	0.285	0.698
Covariance (XOM-AA)	0.178	0.748	0.056	0.929	0.310	0.670
Variance (AXP)	0.218	0.738	0.065	0.931	0.371	0.628
Variance (DD)	0.186	0.740	0.031	0.963	0.338	0.645
Covariance (AXP-DD)	0.201	0.739	0.045	0.947	0.354	0.636
Variance (GE)	0.211	0.750	0.042	0.956	0.354	0.645
Variance (KO)	0.283	0.610	0.037	0.957	0.331	0.653
Covariance (GE-KO)	0.245	0.676	0.039	0.956	0.342	0.649

Log-likelihood decomposition (HEAVY-P versus GARCH)			
	HEAVY-P	GARCH	HEAVY-P gains
BAC - JPM	-2,828	-2,936	108
IBM - MSFT	-2,295	-2,380	85
XOM - AA	-3,415	-3,485	71
AXP - DD	-3,155	-3,271	116
GE - KO	-2,211	-2,304	93

Predictive ability tests at different forecast horizons (days)						
	(1)	(2)	(3)	(5)	(10)	(22)
BAC - JPM	-3.78	-3.13	-2.47	-1.68	0.69	1.84
IBM - MSFT	-2.91	-2.59	-2.39	-1.94	-1.57	-0.49
XOM - AA	-2.84	-2.15	-2.14	-1.87	-1.42	-0.93
AXP - DD	-3.24	-2.94	-2.75	-2.33	-1.43	-0.24
GE - KO	-2.87	-2.71	-2.50	-2.25	-1.04	0.83

Table 5: Diagonal HEAVY parameter estimates for other pairs of assets. Top panel: parameter estimates of HEAVY and GARCH. All coefficients are statistically significant at the 5 percent significance level. Middle panel: HEAVY-P and GARCH log-likelihood (excluding constant terms) at the estimated parameter values. Bottom panel: t-statistics of the predictive ability tests for HEAVY versus GARCH.

margins and the copula - not reported for brevity - mirror the results of the corresponding scalar models; see middle panel of Table 3. The t-statistics of the predictive ability tests in the bottom panel indicate that the HEAVY model consistently outperforms the GARCH model.

5.3 Covariance Targeting Scalar HEAVY Model

In this subsection, we estimate the scalar HEAVY model including all 10 DJIA assets. We show the estimation results for both the original HEAVY specification and the covariance

targeting model given by (13)-(12). We focus on this covariance targeting specification since it is easier to handle the parameter restrictions required for covariance stationarity and positive definiteness of the target. For the GARCH model, we also estimate its covariance targeting parameterisation which has a similar structure to (12). With covariance targeting, the number of parameters to be estimated through numerical optimisation is reduced from 57 to 2 parameters per equation, where the latter are the dynamic parameters of interest.

Table 6 presents the estimates of the dynamic parameters for the HEAVY and GARCH models. The parameter estimates show some differences compared to the average estimate from bivariate models for the same assets; see Table 3. The estimates of the smoothing parameters (B_H , B_M and B_G) have all increased especially B_M , while the estimates of A_H , A_M and A_G are now smaller. The log-likelihood decomposition results show uniform gains for the HEAVY model in all margins and the copula. The copula gains seem particularly impressive. In terms of parameter estimates and the log-likelihood decomposition, the covariance targeting model (bottom panel) shows only slight differences compared to the non-targeting specification.

In Figure 4, we present summary results of the predictive ability tests for the covariance targeting scalar HEAVY and GARCH models. The figure shows the t -statistics for tests of the joint distribution and copula, as well as the minimum, maximum and median t -statistics for the ten margins. In the first three days, the HEAVY model gains are confirmed for the joint distribution, all margins and the copula. The gains of the joint distribution are maintained up to 11 days ahead, then it falls into the insignificance region before improving again towards the end of the forecast horizon. For the margins, the median t -statistics show gains up to 7 days ahead. The copula gains are maintained throughout until the end of

CHAPTER III

Scalar models						
	HEAVY-P		GARCH		HEAVY-V	
	A_H	B_H	A_G	B_G	A_M	B_M
Dynamic parameters (st. error)	0.141 (0.021)	0.792 (0.037)	0.024 (0.002)	0.973 (0.001)	0.247 (0.011)	0.744 (0.010)
Log-likelihood decomposition						
	HEAVY-P		GARCH		HEAVY-P gains	
Margin 1 (BAC)	-1,611		-1,696		85	
Margin 2 (JPM)	-1,999		-2,098		99	
Margin 3 (IBM)	-1,267		-1,323		56	
Margin 4 (MSFT)	-1,471		-1,525		54	
Margin 5 (XOM)	-1,331		-1,420		89	
Margin 6 (AA)	-2,332		-2,381		49	
Margin 7 (AXP)	-1,957		-2,034		77	
Margin 8 (DD)	-1,530		-1,595		65	
Margin 9 (GE)	-1,532		-1,590		58	
Margin 10 (KO)	-911		-956		45	
Copula	4,861		4,661		200	
Joint distribution	-11,080		-11,958		878	
Covariance targeting scalar models						
	HEAVY-P		GARCH		HEAVY-V	
	A_H	B_H	A_G	B_G	A_M	B_M
Dynamic parameters (st. error)	0.177 (0.022)	0.818 (0.023)	0.022 (0.001)	0.977 (0.001)	0.234 (0.009)	0.761 (0.010)
Log-likelihood decomposition						
	HEAVY-P		GARCH		HEAVY-P gains	
Margin 1 (BAC)	-1,616		-1,753		138	
Margin 2 (JPM)	-1,985		-2,119		133	
Margin 3 (IBM)	-1,257		-1,327		69	
Margin 4 (MSFT)	-1,464		-1,525		61	
Margin 5 (XOM)	-1,340		-1,424		84	
Margin 6 (AA)	-2,324		-2,379		55	
Margin 7 (AXP)	-1,940		-2,046		106	
Margin 8 (DD)	-1,528		-1,592		64	
Margin 9 (GE)	-1,521		-1,595		74	
Margin 10 (KO)	-911		-954		43	
Copula	4,781		4,629		151	
Joint distribution	-11,105		-12,084		978	

Table 6: Scalar HEAVY estimates for 10 DJIA assets. Top panel: parameter estimates and log-likelihood (excluding constant terms) decomposition for scalar HEAVY and GARCH without covariance targeting. Bottom panel: parameter estimates and log-likelihood (excluding constant terms) decomposition for the covariance targeting scalar HEAVY and GARCH.

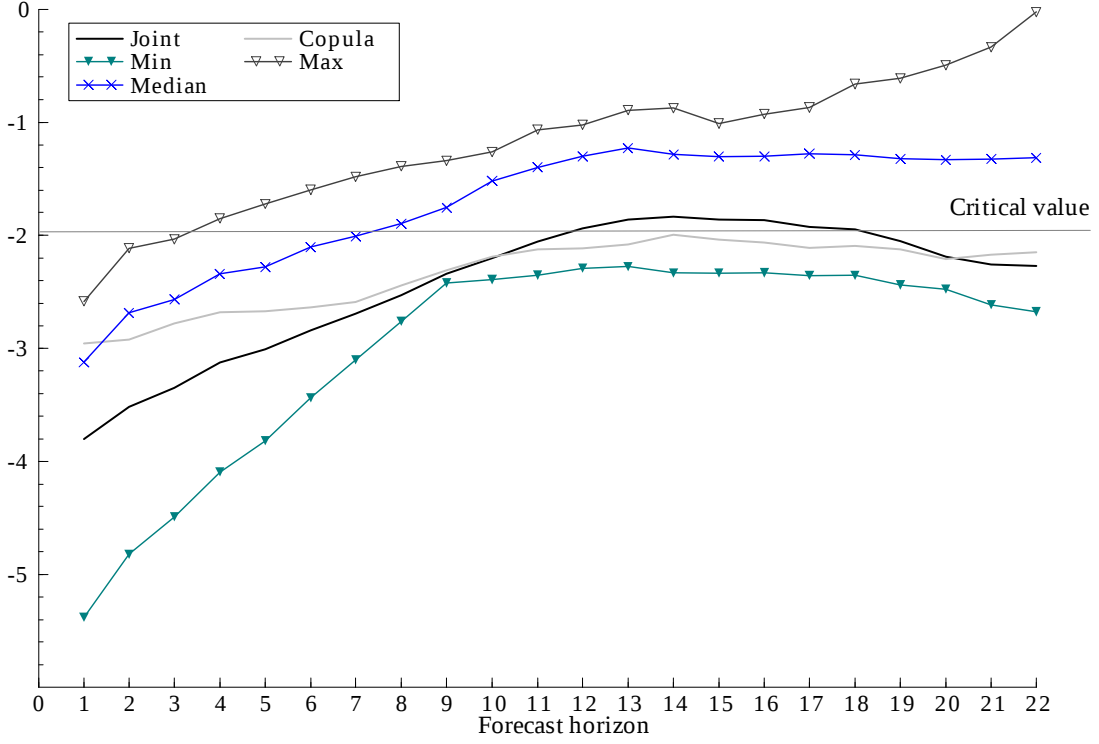


Figure 4: Predictive ability tests' t-statistics for the covariance targeting scalar HEAVY and GARCH models.

the forecast horizon, which is consistent with the substantial overall gain in the copula log-likelihood.

6 Conclusion

This paper introduces a new class of multivariate volatility models with robust performance in out-of-sample prediction of the covariance matrix for a collection of financial assets. While GARCH models - in their many variations - have proved successful in the past two decades, the increasing availability of high-frequency data provides important additional information. Utilizing this information to forecast the conditional variance of daily asset returns has already borne fruit in the univariate case as documented by several recent studies.

Our study is one of the first to document this feature in the multivariate case using a relatively large group of assets. We present our results in the framework of the multivariate HEAVY class of models. Using a linear specification, we discuss in some detail the model's dynamic properties, its covariance targeting representation, and provide closed-form forecasting formulas. We show how the profile of forecasts from HEAVY models differs from GARCH models, in particular with regard to persistence and short-run momentum effects. We also discuss QMLE of HEAVY models under the assumption of a Wishart distribution for the innovation matrices.

In an application to the S&P 500 ETF and ten stocks from the DJIA index, we compare the HEAVY and GARCH models in the challenging environment of the financial crisis. We show that forecasts from the HEAVY model dominate GARCH forecasts with the gains being particularly significant at short forecast horizons. The results seem consistent across different pairs of assets and also when using all ten DJIA stocks in a covariance targeting model. The HEAVY model's relatively short response time compared to GARCH seems to enable it to efficiently track sudden changes in asset return volatilities and correlations. With regard to the latter, our results for log-likelihood decompositions and predictive ability tests strongly suggest that high-frequency data provides timely and important information for modelling and forecasting conditional correlations.

For future research, a number of extensions could potentially add to our understanding of how best to model and forecast multivariate volatility. It would be interesting to add asymmetric terms to the HEAVY model to explicitly capture the leverage effect and see how this improves its forecast performance. It might also be beneficial to use a long-run/short-run component model in the dynamic equations to separate out transitory movements in volatility.

References

- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25, 177–190.
- Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39, 885–905.
- Andersen, T. G., T. Bollerslev, and F. X. Diebold (2007). Roughing it up: including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics* 89, 701–720.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2003). Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Andersen, T. G., T. Bollerslev, and N. Meddahi (2011). Realized volatility forecasting and market microstructure noise. *Journal of Econometrics* 160, 220–234.
- Asai, M., M. McAleer, and J. Yu (2006). Multivariate stochastic volatility: a review. *Econometric Reviews* 25, 145–175.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* 76, 1481–1536.
- Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *Journal of Econometrics* 162, 149–169.
- Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B* 64, 253–280.
- Barndorff-Nielsen, O. E. and N. Shephard (2004). Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics. *Econometrica* 72, 885–925.
- Bauer, G. H. and K. Vorkink (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics* 160, 93–101.
- Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.
- 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.
- Brownlees, C. T. and R. F. Engle (2010). Volatility, correlation and tails for systemic risk measurement. Working paper: Stern School of Business, New York University.
- Brownlees, C. T. and G. M. Gallo (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics* 8, 29–56.

- Chib, S., Y. Omori, and M. Asai (2009). Multivariate stochastic volatility. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 365–400. Springer-Verlag.
- Chiriac, R. and V. Voev (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*. Forthcoming.
- Cipollini, F., R. F. Engle, and G. M. Gallo (2007). A model for multivariate non-negative valued processes in financial econometrics. Unpublished paper: Stern School of Business, New York University.
- Comte, F. and O. Lieberman (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis* 84, 61–84.
- Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch (2008). The volatility of realized volatility. *Econometric Reviews* 17, 46–78.
- Cox, D. R. (1961). Tests of separate families of hypotheses. *Proceedings of the Berkeley Symposium* 4, 105–123.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Diks, C., V. Panchenko, and D. van Dijk (2008). Out-of-sample comparison of copula specifications in multivariate density forecasts. Working paper: Australian School of Business, University of New South Wales.
- Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R. F. and J. P. Gallo (2006). A multiple indicator model for volatility using intra daily data. *Journal of Econometrics* 131, 3–27.
- Engle, R. F., D. F. Hendry, and J. F. Richard (1983). Exogeneity. *Econometrica* 51, 277–304.
- Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Engle, R. F. and G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, pp. 475–497. Oxford: Oxford University Press.
- Engle, R. F. and J. Mezrich (1996). GARCH for groups. *Risk* 9, 36–40.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Gupta, A. K. and D. K. Nagar (2000). *Matrix Variate Distributions*. New York: Chapman & Hall.
- Hafner, C. M. (2003). Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics* 1, 26–54.
- Hafner, C. M. and A. Preminger (2009). On asymptotic theory for multivariate GARCH models. *Journal of Multivariate Analysis* 100, 2044–2054.

- Hansen, P. R., Z. A. Huang, and H. H. Shek (2011). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*. Forthcoming.
- Hansen, P. R., A. Lunde, and V. Voev (2010). Realized beta GARCH: a multivariate GARCH model with realized measures of volatility and covolatility. Working paper: CREATES, Aarhus University.
- Jeantheau, T. (1998). Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14, 70–86.
- Jin, X. and J. M. Maheu (2010). Modelling realized covariances and returns. Working paper: Department of Economics, University of Toronto.
- Laurent, S., J. V. K. Rombouts, and F. Violante (2009). On loss functions and ranking forecasting performances of multivariate volatility models. Working paper: Centre Interuniversitaire sur le Risque, les Politiques Economiques et l'Emploi (CIRPEE).
- Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: Wiley.
- Magnus, J. R. (1988). *Linear Structures*. London: Griffin.
- Nelsen, R. B. (2006). *An Introduction to Copulas* (2 ed.). New York: Springer.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2111–2245. North-Holland.
- Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Patton, A. J. (2009). Copula-based models for financial time series. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 767–785. Springer-Verlag.
- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160, 246–256.
- Patton, A. J. and K. K. Sheppard (2009). Evaluating volatility and correlation forecasts. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 801–838. Springer-Verlag.
- Shephard, N. and K. K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics* 25, 197–231.
- Srivastava, M. S. (2003). Singular Wishart and multivariate Beta distributions. *Annals of Statistics* 31, 1537–1560.
- Voev, V. (2008). Dynamic modelling of large-dimensional covariance matrices. In L. Bauwens, W. Pohlmeier, and D. Veredas (Eds.), *High Frequency Financial Econometrics*, pp. 293–312. Physica-Verlag HD.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.
- West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.

A Results from Matrix Algebra and Calculus

The following results can be found in Magnus (1988). The vec operator stacks the columns of a $(k \times k)$ matrix into a $(k^2 \times 1)$ vector. The $vech$ operator stacks the lower triangular part including the main diagonal of a $(k \times k)$ symmetric matrix into a $(k^* \times 1)$ vector, $k^* = k(k+1)/2$.

For any matrices $A(k \times l)$, $B(l \times m)$ and $C(m \times n)$, $vec(ABC) = (C' \otimes A)vec(B)$. For any $(k \times k)$ symmetric matrix A , the $(k^* \times k^2)$ elimination matrix L_k is defined such that $vech(A) = L_k vec(A)$, and the $(k^2 \times k^*)$ duplication matrix D_k is defined such that $vec(A) = D_k vech(A)$. Let Λ_{ij} be a $(k \times k)$ matrix with 1 in its (i, j) -th position and zeros elsewhere. Then L_k and D_k can be obtained using $L_k = \sum_{i \geq j} vech(\Lambda_{ij})(vec(\Lambda_{ij}))'$ and $D_k = \sum_{i > j} vec(\Lambda_{ij} + \Lambda_{ji})(vech(\Lambda_{ij}))' + \sum_{i=j} vec(\Lambda_{ij})(vech(\Lambda_{ij}))'$.

The above results can be combined to give the following result for any $(k \times k)$ matrices A and B , with B being symmetric

$$vech(ABA') = L_k vec(ABA') = L_k(A \otimes A)vec(B) = L_k(A \otimes A)D_k vech(B). \quad (A.1)$$

For a $(k \times l)$ matrix function $F(X)$ and a $(m \times n)$ matrix of variables X , the derivative of $F(X)$ with respect to X , denoted by the $(kl \times mn)$ matrix $DF(X)$, is given by

$$DF(X) = \frac{\partial vec(F(X))}{\partial (vec(X))'}.$$

Also for any $(k \times k)$ matrix A ,

$$\frac{\partial A^{-1}}{\partial A} = -((A^{-1})' \otimes A^{-1}), \frac{\partial \log |A|}{\partial A} = (vec((A^{-1})'))', \text{ and } \frac{\partial tr(AX)}{\partial X} = (vec(A'))'.$$

B Second Moments' Structure

Since the model is expressed for p_t (i.e. for the squares and cross-products of daily returns), we are able to obtain explicit expressions for the fourth moment of returns by deriving $\text{Var}[p_t]$. Similarly, by deriving $\text{Var}[v_t]$, we are able to analyse the second moment of the realised measure which gives an expression for the volatility of volatility; see Engle (2002) and Corsi et al. (2008) for a discussion of modelling the volatility of volatility using the VIX and realised volatility, respectively.

The following proposition gives the structure of the second moments of p_t and v_t , which is derived under the assumption $\mathbb{E}_{t-1}[\varepsilon_t] = \mathbb{E}_{t-1}[\eta_t] = I_k$. The expressions in (B.1)-(B.2) can be simplified further by assuming a Wishart distribution for the innovations which gives (B.3)-(B.4).

Proposition 4 (i) *Under the assumption that $\mathbb{E}_{t-1}[\varepsilon_t] = \mathbb{E}_{t-1}[\eta_t] = I_k$, the second moments of p_t and v_t are given by*

$$\text{Var}[p_t] = \mathbb{E} [Z_{H,t} \text{Var}_{t-1}[\text{vech}(\varepsilon_t)] Z'_{H,t}] + \text{Var} [Z_{H,t} \text{vech}(I_k)], \quad (\text{B.1})$$

$$\text{Var}[v_t] = \mathbb{E} [Z_{M,t} \text{Var}_{t-1}[\text{vech}(\eta_t)] Z'_{M,t}] + \text{Var} [Z_{M,t} \text{vech}(I_k)], \quad (\text{B.2})$$

where $Z_{H,t} = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}}) D_k$, $Z_{M,t} = L_k(M_t^{\frac{1}{2}} \otimes M_t^{\frac{1}{2}}) D_k$ and $\text{Var}_{t-1}[\cdot]$ denotes the variance conditional on $\mathcal{F}_{t-1}^{HF}$.

(ii) *Under the additional assumption that ε_t and η_t are i.i.d. Wishart distributed, the second moments of p_t and v_t are given by*

$$\mathbb{E}[p_t p'_t] = 2D_k^+ \mathbb{E} [(H_t \otimes H_t)] L'_k + \mathbb{E}[h_t h'_t], \quad (\text{B.3})$$

$$\mathbb{E}[v_t v_t'] = 2n^{-1} D_k^+ \mathbb{E}[(M_t \otimes M_t)] L_k' + \mathbb{E}[m_t m_t'], \quad (\text{B.4})$$

where $D_k^+ = (D_k' D_k)^{-1} D_k'$ is the Moore-Penrose inverse of D_k .

Dropping the t subscripts to avoid cluttered notation, the second moment structure of p_t given in (B.3) will have the following structure in the 2-dimensional case

$$\mathbb{E}[pp'] = \mathbb{E} \begin{pmatrix} r_1^4 & r_1^3 r_2 & r_1^2 r_2^2 \\ r_1^3 r_2 & r_1^2 r_2^2 & r_1 r_2^3 \\ r_1^2 r_2^2 & r_1 r_2^3 & r_2^4 \end{pmatrix} = \mathbb{E} \begin{pmatrix} 3h_{11}^2 & 3h_{11}h_{12} & 2h_{12}^2 + h_{11}h_{22} \\ 3h_{11}h_{21} & 2h_{21}^2 + h_{11}h_{22} & 3h_{12}h_{22} \\ 2h_{21}^2 + h_{11}h_{22} & 3h_{21}h_{22} & 3h_{22}^2 \end{pmatrix},$$

where r_1 and r_2 denote the daily returns for assets 1 and 2, respectively, and h_{ij} , $i, j = 1, 2$, are the elements of H_t . Applying a *vec* operator to (B.3) gives a similar result to (10) in Hafner (2003), which discusses the fourth moment structure of GARCH models when H_t follows a GARCH specification and daily returns are assumed to be normally distributed.

The result in (B.4) seems novel in the context of realised measures. In the univariate case, Corsi et al. (2008) estimate the volatility of realised volatility by utilizing consistent estimators of the integrated quarticity of returns, such as realised quarticity, realised quad-power quarticity and realised tri-power quarticity. In an application to S&P 500 index futures, they show that the unconditional distributions of these three measures are skewed and leptokurtic even after applying a log transformation. The three measures also exhibit clustering which prompts the authors to develop a GARCH-type model for realised volatility. Engle (2002) also discusses different models for volatility of volatility using the VIX time series.

C The Wishart Distribution: An Overview

The Wishart distribution is the matrix-variate generalisation of the χ^2 distribution. It is the sampling distribution of the sample covariance matrix for random draws from the multivariate normal distribution; see Gupta and Nagar (2000) for a detailed treatment. We begin our overview by relating the Wishart distribution to the multivariate normal and the matrix-variate normal. Let $x_1, \dots, x_n$ be independent random vectors drawn from the (centered) multivariate normal distribution where each $(k \times 1)$ vector $x_i \sim N_k(0, \Sigma)$, $i = 1, \dots, n$. Then the $(k \times n)$ matrix $X = (x_1, \dots, x_n)$ has a matrix variate normal distribution denoted as $X \sim N_{k,n}(0, \Sigma \otimes I_n)$. If $n \geq k$, then $S = XX'$ is a positive definite $(k \times k)$ matrix and follows a (centered) Wishart distribution, denoted as $S \sim W_k(n, \Sigma)$, where n is integer degrees of freedom and Σ is the scale matrix. The unconditional moments of S are given by $E(S) = n\Sigma$, and $\text{Var}(\text{vec}(S)) = 2nD_k D_k^+(\Sigma \otimes \Sigma)$.

The density of S is given by

$$W_k(n, \Sigma) = \frac{|S|^{\frac{n-k-1}{2}}}{2^{\frac{nk}{2}} \Gamma_k\left(\frac{n}{2}\right) |\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\text{tr}(\Sigma^{-1}S)\right), \quad n \geq k,$$

where $\Gamma_k\left(\frac{n}{2}\right) = \pi^{k(k-1)/4} \prod_{j=1}^k \Gamma\left(\frac{n}{2} + (1-j)/2\right)$ is the multivariate gamma function. It is also useful to define a standardised Wishart distribution such that if $S \sim SW_k(n, \Sigma)$, where SW_k denote a standardised Wishart distribution of dimension k , we have $E(S) = \Sigma$ instead of $E(S) = n\Sigma$ under the non-standardised Wishart. Note that $SW_k(n, \Sigma)$ is equivalent to $W_k(n, n^{-1}\Sigma)$. In this case, the density of S is given by

$$SW_k(n, \Sigma) = \frac{n^{\frac{nk}{2}} |S|^{\frac{n-k-1}{2}}}{2^{\frac{nk}{2}} \Gamma_k\left(\frac{n}{2}\right) |\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{n}{2}\text{tr}(\Sigma^{-1}S)\right), \quad n \geq k. \quad (\text{C.1})$$

If the $(k \times n)$ matrix X does not have full column rank (i.e. $n < k$), $S = XX'$ follows instead a singular Wishart distribution. Srivastava (2003) derives the density of the singular Wishart as

$$SINGW_k(n, \Sigma) = \frac{\pi^{(-kn+n^2)/2} |\tilde{S}|^{\frac{n-k-1}{2}}}{2^{\frac{nk}{2}} \Gamma_n(\frac{n}{2}) |\Sigma|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} \text{tr}(\Sigma^{-1}S)\right), \quad n < k, \quad (\text{C.2})$$

where $\tilde{S}$ is a diagonal matrix containing the non-zero eigenvalues of S along the main diagonal.

One of the key results on Wishart distributions is that if $S \sim W_k(n, \Sigma)$, then $ASA' \sim W_k(n, A\Sigma A')$ for any $(k \times k)$ nonsingular matrix A . Srivastava (2003) extends this result to the singular Wishart case. Based on the ranks of P_t and V_t , we use this result in specifying the distribution of the innovation matrices ε_t and η_t as discussed in Section 3.1.

D Technical Proofs

D.1 Proof of Proposition 1

By taking unconditional expectation of (7) and (8) we have

$$\omega_H = C_H + B_H\omega_H + A_H\omega_M, \quad \omega_M = C_M + B_M\omega_M + A_M\omega_M. \quad (\text{D.1})$$

By definition $\bar{\kappa} = \Omega_M^{\frac{1}{2}}\Omega_H^{-\frac{1}{2}}$, which implies $\Omega_M^{\frac{1}{2}} = \bar{\kappa}\Omega_H^{\frac{1}{2}}$, and $\Omega_M = \bar{\kappa}\Omega_H\bar{\kappa}'$. Thus $\omega_M := \text{vech}(\Omega_M) = \text{vech}(\bar{\kappa}\Omega_H\bar{\kappa}') = L_k(\bar{\kappa} \otimes \bar{\kappa})D_k\omega_H$ using (A.1) for the last equality.

Let $\kappa = L_k(\bar{\kappa} \otimes \bar{\kappa})D_k$ and substitute the last result for ω_M in the first expression in (D.1). By collecting terms we have that the intercept coefficients are given by $C_H = (I_{k^*} - B_H - A_H\kappa)\omega_H$ and $C_M = (I_{k^*} - B_M - A_M)\omega_M$, which when substituted in (7) and

(8) gives the stated result.

The proof for (13) follows by noting that $E[\tilde{v}_t] = \kappa^{-1}E[v_t] = \kappa^{-1}\omega_M = \omega_H$, where the last equality follows from above by defining $\kappa = L_k(\bar{\kappa} \otimes \bar{\kappa})D_k$. The rest follows by collecting terms and substituting for C_H in the first expression in (D.1).

D.2 Proof of Proposition 2

We start with the proof of (14). The 1-step forecast of h_t is $E_t[h_{t+1}] = h_{t+1}$, since h_{t+1} is $\mathcal{F}_t^{HF}$ -measurable. From (7), the 2-step forecast is

$$E_t[h_{t+2}] = E_t[C_H + B_H h_{t+1} + A_H v_{t+1}] = C_H + B_H h_{t+1} + A_H E_t[v_{t+1}].$$

The 3-step forecast is

$$\begin{aligned} E_t[h_{t+3}] &= E_t[C_H + B_H h_{t+2} + A_H v_{t+2}] = C_H + B_H E_t[h_{t+2}] + A_H E_t[v_{t+2}] \\ &= (I_{k^*} + B_H)C_H + B_H^2 h_{t+1} + A_H E_t[v_{t+2}] + B_H A_H E_t[v_{t+1}], \end{aligned}$$

where the last equality follows by substituting for $E_t[h_{t+2}]$ from above and collecting terms.

By forward iteration, it is straightforward to show that

$$E_t[h_{t+s}] = \sum_{i=1}^{s-1} B_H^{i-1} C_H + B_H^{s-1} h_{t+1} + \sum_{i=1}^{s-1} B_H^{i-1} A_H E_t[v_{t+s-i}]. \quad (\text{D.2})$$

Now find an expression for $E_t[v_{t+s-i}]$ in terms of m_{t+1} , which is $\mathcal{F}_t^{HF}$ -measurable. We start with the 1-step forecast of v_t which is $E_t[v_{t+1}] = m_{t+1}$ by definition. The 2-step

forecast is

$$\begin{aligned} E_t[v_{t+2}] &= E_t[E_{t+1}[v_{t+2}]] = E_t[m_{t+2}] = E_t[C_M + B_M m_{t+1} + A_M v_{t+1}] \\ &= C_M + (B_M + A_M)m_{t+1}, \end{aligned}$$

since $E_t[v_{t+1}] = m_{t+1}$. The 3-step forecast is

$$\begin{aligned} E_t[v_{t+3}] &= E_t[E_{t+1}[v_{t+3}]] = E_t[C_M + (B_M + A_M)m_{t+2}] = C_M + (B_M + A_M)E_t[m_{t+2}] \\ &= (I_{k^*} + (B_M + A_M))C_M + (B_M + A_M)^2 m_{t+1}, \end{aligned}$$

where the second equality follows by substitution from above with a 1-period forward iteration, and the last equality follows by substituting for m_{t+2} , applying the conditional expectation operator and collecting terms. By forward iteration, we have the following formula for the s -step forecast of v_t

$$E_t[v_{t+s}] = \sum_{j=1}^{s-1} (B_M + A_M)^{j-1} C_M + (B_M + A_M)^{s-1} m_{t+1}. \quad (\text{D.3})$$

Using (D.3) to substitute for $E_t[v_{t+s-i}]$ in (D.2), while adapting the summation limit by replacing s in (D.3) with $(s-i)$ gives the stated result.

The proof of (15) follows similar steps. We start by taking unconditional expectations of (7) and (8) which gives

$$\omega_H = C_H + B_H \omega_H + A_H \omega_M, \quad \omega_M = C_M + B_M \omega_M + A_M \omega_M,$$

so that the constant terms can be expressed as $C_H = \omega_H - B_H \omega_H - A_H \omega_M$ and $C_M =$

$\omega_M - B_M\omega_M - A_M\omega_M$. Substituting these expressions in (7) and (8) gives

$$\begin{aligned} h_t &= \omega_H - B_H\omega_H - A_H\omega_M + B_Hh_{t-1} + A_Hv_{t-1} \\ &= \omega_H + B_H(h_{t-1} - \omega_H) + A_H(v_{t-1} - \omega_M), \end{aligned}$$

$$\begin{aligned} m_t &= \omega_M - B_M\omega_M - A_M\omega_M + B_Mm_{t-1} + A_Mv_{t-1} \\ &= \omega_M + B_M(m_{t-1} - \omega_M) + A_M(v_{t-1} - \omega_M). \end{aligned}$$

Forward iteration of these equations as illustrated in the proof of (14) yields (15).

D.3 Proof of Proposition 3

We derive the score vector and prove that it is a martingale difference sequence only for the HEAVY-P equation. The derivation for the HEAVY-V equation is analogous. We derive the $(1 \times \delta_H)$ score vector $\frac{\partial l_{H,t}(\theta_H)}{\partial \theta'_H}$ from the log-likelihood equation which gives

$$\begin{aligned} \frac{\partial l_{H,t}(\theta_H)}{\partial \theta'_H} &= -\frac{1}{2} \frac{\partial \log |H_t|}{\partial \theta'_H} - \frac{1}{2} \frac{\partial \text{tr}(H_t^{-1} P_t)}{\partial \theta'_H} \\ &= -\frac{1}{2} \frac{\partial \log |H_t|}{\partial H_t} \frac{\partial H_t}{\partial \theta'_H} - \frac{1}{2} \frac{\partial \text{tr}(H_t^{-1} P_t)}{\partial H_t^{-1}} \frac{\partial H_t^{-1}}{\partial H_t} \frac{\partial H_t}{\partial \theta'_H} \\ &= -\frac{1}{2} (\text{vec}(H_t^{-1}))' \frac{\partial \text{vec}(H_t)}{\partial \theta'_H} + \frac{1}{2} (\text{vec}(P_t))' (H_t^{-1} \otimes H_t^{-1}) \frac{\partial \text{vec}(H_t)}{\partial \theta'_H} \\ &= \frac{1}{2} [(\text{vec}(P_t))' (H_t^{-1} \otimes H_t^{-1}) - (\text{vec}(H_t^{-1}))'] \frac{\partial \text{vec}(H_t)}{\partial \theta'_H} \\ &= \frac{1}{2} [(\text{vec}(P_t))' (H_t^{-1} \otimes H_t^{-1}) - (\text{vec}(H_t^{-1} H_t H_t^{-1}))'] \frac{\partial \text{vec}(H_t)}{\partial \theta'_H} \\ &= \frac{1}{2} [(\text{vec}(P_t))' (H_t^{-1} \otimes H_t^{-1}) - ((H_t^{-1} \otimes H_t^{-1}) \text{vec}(H_t))'] \frac{\partial \text{vec}(H_t)}{\partial \theta'_H} \\ &= \frac{1}{2} [(\text{vec}(P_t))' - (\text{vec}(H_t))'] (H_t^{-1} \otimes H_t^{-1}) \frac{\partial \text{vec}(H_t)}{\partial \theta'_H}, \end{aligned}$$

where in the second equality we used the chain rule and the matrix derivatives in the third equality are obtained using the rules stated in Appendix A.

The score vector is a martingale difference sequence such that $E_{t-1}[\frac{\partial l_{H,t}(\theta_H)}{\partial \theta'_H}] = 0$ as the conditional expectation of the term in square brackets is 0 since

$$E_{t-1}[(vec(P_t))' - (vec(H_t))'] = E_{t-1}[(vec(P_t))'] - E_{t-1}[(vec(H_t))'] = 0,$$

where we use $E_{t-1}[(vec(P_t))'] = (vec(H_t))'$, which follows directly from the conditional moment assumption $E_{t-1}[\varepsilon_t] = I_k$.

D.4 Proof of Proposition 4

For the first part of the proposition, we only show the proof for (B.1) as (B.2) follows similar arguments. We start from $p_t := vech(P_t) = vech(H_t^{\frac{1}{2}} \varepsilon_t H_t^{\frac{1}{2}}) = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}}) D_k vech(\varepsilon_t)$, where the last result follows from (A.1). Let $Z_{H,t} = L_k(H_t^{\frac{1}{2}} \otimes H_t^{\frac{1}{2}}) D_k$, which is $\mathcal{F}_{t-1}^{HF}$ -measurable. Also, let $\text{Var}_{t-1}[\cdot]$ denote the variance conditional on $\mathcal{F}_{t-1}^{HF}$. Using the variance decomposition, we obtain

$$\begin{aligned} \text{Var}[p_t] &= E[\text{Var}_{t-1}[p_t]] + \text{Var}[E_{t-1}[p_t]] \\ &= E[\text{Var}_{t-1}[Z_{H,t} vech(\varepsilon_t)]] + \text{Var}[E_{t-1}[Z_{H,t} vech(\varepsilon_t)]] \\ &= E[Z_{H,t} \text{Var}_{t-1}[vech(\varepsilon_t)] Z'_{H,t}] + \text{Var}[Z_{H,t} E_{t-1}[vech(\varepsilon_t)]] \end{aligned}$$

as $Z_{H,t}$ is $\mathcal{F}_{t-1}^{HF}$ -measurable. As $E_{t-1}[\varepsilon_t] = I_k$ by assumption, it follows that $E_{t-1}[vech(\varepsilon_t)] = vech(I_k)$ which gives (B.1).

For (B.3), $\varepsilon_t \stackrel{i.i.d.}{\sim} SINGW_k(1, I_k)$ implies $P_t | \mathcal{F}_{t-1}^{HF} \sim SINGW_k(1, H_t)$ and also im-

plies $R_t | \mathcal{F}_{t-1}^{HF} \sim N(0, H_t)$ since $P_t = R_t R_t'$. Thus $\text{Var}_{t-1}[\text{vec}(P_t)] = \text{Var}_{t-1}[(R_t \otimes R_t)] = 2D_k D_k^+(H_t \otimes H_t)$, where the second equality follows from the conditional normality of R_t by Magnus (1988, Theorem 10.2) noting that conditioning on $\mathcal{F}_{t-1}^{HF}$ enables us to treat H_t as a nonstochastic matrix. Therefore

$$\begin{aligned} \text{Var}_{t-1}[p_t] &= \text{Var}_{t-1}[L_k \text{vec}(P_t)] = L_k \text{Var}_{t-1}[\text{vec}(P_t)] L_k' \\ &= 2L_k D_k D_k^+(H_t \otimes H_t) L_k' = 2D_k^+(H_t \otimes H_t) L_k', \end{aligned}$$

where the last equality follows since $L_k D_k = I_{k^*}$ by Magnus (1988, Theorem 5.5). We obtain the unconditional second moment of p_t using the variance decomposition

$$\begin{aligned} \text{Var}[p_t] &= \text{E}[\text{Var}_{t-1}[p_t]] + \text{Var}[\text{E}_{t-1}[p_t]] = \text{E}[2D_k^+(H_t \otimes H_t) L_k'] + \text{Var}[h_t] \\ &= 2D_k^+ \text{E}[(H_t \otimes H_t)] L_k' + \text{Var}[h_t]. \end{aligned}$$

We can write $\text{Var}[p_t] = \text{E}[p_t p_t'] - \text{E}[p_t] \text{E}[p_t]'$ and $\text{Var}[h_t] = \text{E}[h_t h_t'] - \text{E}[h_t] \text{E}[h_t]'$. By noting that $\text{E}[p_t] = \text{E}[\text{E}_{t-1}[p_t]] = \text{E}[h_t]$, the last equation for $\text{Var}[p_t]$ can be simplified to give the stated result. The proof of (B.4) is similar except in the intermediate step of deriving $\text{Var}_{t-1}[\text{vec}(V_t)]$, where in this case Theorem 10.3 of Magnus (1988) directly applies since η_t has a non-singular Wishart distribution. Thus we have $\text{Var}_{t-1}[\text{vec}(V_t)] = 2n^{-1} D_k D_k^+(M_t \otimes M_t)$ and the rest of the proof follows as in (B.3).

CHAPTER IV

FLEXIBLE COVARIANCE-TARGETING VOLATILITY

MODELS USING ROTATED RETURNS

1 Introduction

Search is still ongoing for multivariate volatility models with flexible dynamics and ease of application in moderately large dimensions. Modelling and forecasting multivariate volatility is not only crucial for asset pricing and optimal portfolio allocation, but also to characterise the systemic risk profile of individual firms. Brownlees and Engle (2010) illustrate the importance of modelling and forecasting the conditional covariance matrix of asset returns, where they show that a rise in a stock's return volatility and correlation with the market magnifies its contribution to their proposed measure of systemic risk. Highly leveraged financial companies in the recent financial crisis are a case in point.

The crisis forcefully demonstrated the need for more robust models to capture and project financial risk; in particular to capture correlation dynamics. However in practice, developing new models faces the “curse of dimensionality” in reference to the - often exponential - increase in the number of model parameters as the number of assets under study grows. Reviews of the multivariate generalised autoregressive conditional heteroskedasticity (GARCH) literature are given by, for example, Bauwens et al. (2006), Engle (2009a), Francq and Zakoian (2010, Ch. 11) and Silvennoinen and Teräsvirta (2009).

The seed idea in this paper is to undertake a transformation (in particular, a rotation) of the raw returns, which enables us to easily extend the idea of variance targeting (Engle and Mezrich, 1996) to covariance targeting in multivariate models of any dimension. The transformation we propose is not novel, and is related to recent work on the orthogonal GARCH model of Alexander and Chibumba (1997) and Alexander (2001), and its extensions in van der Weide (2002), Lanne and Saikkonen (2007), Fan et al. (2008) and Boswijk and van der Weide (2011). The interest in these papers is to find orthogonal or

unconditionally uncorrelated components in the raw returns which can then be modelled individually through univariate volatility models.¹ In contrast, we utilise a closely related transformation enabling us to fit flexible multivariate models to the rotated returns using covariance targeting.

We focus on the popular BEKK (Engle and Kroner, 1995) and Dynamic Conditional Correlations (DCC) (Engle, 2002) models, and propose new parameterisations to enrich both models. We focus throughout on diagonal models, to be explained in detail below, and a related parameterisation that offers flexibility in modelling both the volatilities and correlations while economising on the number of parameters. The models we discuss are particularly attractive in terms of estimation and inference, and offers computational advantages compared to existing models.

Interest in diagonal models for the DCC process is demonstrated in a number of recent studies, where the objective is to introduce more flexible dynamics while also having parameterisations that are feasible in large dimensions. For p assets, diagonal models in the case of BEKK or DCC, when coupled with covariance targeting, will have a number of dynamic parameters equal to $2p$.² In addition to the DCC model with scalar dynamic parameters, Engle (2002) also proposed a generalisation with flexible dynamics but it is highly parameterised. Recent studies which focus on DCC with diagonal structures are, for example, Cappiello et al. (2006), Billio et al. (2006), Billio and Caporin (2009) and Hafner and Franses (2009).

¹The model of Fan et al. (2008) differs in that the estimated components are also conditionally uncorrelated. We discuss the relation of our model to orthogonal GARCH models in Section 2.6.

²We use the term ‘dynamic’ parameters to denote the parameters of the dynamic equation for the conditional covariance matrix in the BEKK model, and for the conditional correlation matrix in the DCC model. However, covariance targeting also requires the estimation of ‘static’ parameters which characterise the unconditional second moment of the returns. Estimation is typically undertaken in two stages as discussed later.

Within the class of diagonal models, we propose a novel parameterisation that may be attractive in large dimensions. We call it the common persistence (CP) model which imposes common persistence on all elements of the conditional covariance/correlation matrix. This is motivated by the empirical observation that parameter estimates of GARCH(1,1) processes tend to show similar persistence across assets, while exhibiting different levels of smoothness. In addition, the smoothness level seems to change over time; particularly it tends to decline in volatile periods. Brownlees (2010) reports similar findings in his analysis of US financial firms during the recent financial crisis. The common persistence model has only $p + 1$ dynamic parameters, and we show that it performs quite favourably in comparison to diagonal models which have $2p$ dynamic parameters.

We show that fitting multivariate volatility models to the rotated returns is essentially the same as fitting models (with different dynamic parameters, in general) to the raw returns; the rotation of the returns simply provides an easier way to do covariance targeting. This equivalence holds since the difference in the likelihood depends on the static parameters needed for the transformation, but is invariant to the type of chosen model. The usefulness of this rotation technique is illustrated using data on the S&P 500 ETF and some DJIA stocks. We analyse bivariate models as well as a moderately large system with 10 DJIA stocks.

The structure of the paper is as follows: Section 2 discusses the model and its properties. Section 3 shows how to estimate the model using a two step estimation strategy, providing a simple multivariate extension of covariance targeting. In Section 4 we apply this model to financial data to illustrate its performance in comparison to related models. Section 5 draws some conclusions.

2 Modelling Approach

2.1 The Model

First we assume the p -dimensional zero-mean time series

$$r_t, \quad t = 1, \dots, T,$$

is ergodic. The unconditional covariance of the returns is given by

$$\text{Var}[r_t] = \overline{H} = P\Lambda P',$$

using the spectral decomposition in the second equality, where P is a matrix of eigenvectors, and the eigenvalue matrix Λ is diagonal with non-negative elements $\lambda_1, \lambda_2, \dots, \lambda_p$. Throughout we assume that the eigenvalues in Λ are ordered such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ with $\lambda_p > 0$. It follows that $P^{-1} = P'$ and so $P'P = I$. Hence we can define the symmetric square root of $\overline{H}$

$$\overline{H}^{1/2} = P\Lambda^{1/2}P'.$$

Second, letting $r_t = \overline{H}^{1/2}e_t$ we can define the rotated returns

$$e_t = \overline{H}^{-1/2}r_t = P\Lambda^{-1/2}P'r_t, \quad \text{Var}[e_t] = I.$$

Then we complete the model by specifying the conditional covariance of the rotated returns

$$\text{Var}[e_t|\mathcal{F}_{t-1}] = G_t,$$

where $E[e_t|\mathcal{F}_{t-1}] = 0$. In order to ease the computational burden, we use a covariance targeting parameterisation (Engle and Mezrich, 1996, in the univariate case of variance targeting) of a BEKK-type model (Engle and Kroner, 1995) applied to e_t ,

$$G_t = (I - AA' - BB') + Ae_{t-1}e'_{t-1}A' + BG_{t-1}B', \quad G_0 = I, \quad (1)$$

where we assume

$$(I - AA' - BB') \geq 0,$$

in the sense of being positive semidefinite.

Covariance stationarity in (1) follows directly from the analysis of BEKK models by Engle and Kroner (1995) and requires the eigenvalues of $(A \otimes A) + (B \otimes B)$ to be less than one in modulus. Thus unconditionally we can rewrite (1) as

$$E[G_t] - AE[G_t]A' - BE[G_t]B' = I - AA' - BB',$$

where $E[G_t] = I$ is a solution to this equation implying $E[e_te'_t] = I$.

Let $\text{Var}[r_t|\mathcal{F}_{t-1}] = H_t$, fitting the covariance targeting BEKK model to r_t implies

$$H_t = (\bar{H} - A\bar{H}A' - B\bar{H}B') + Ar_{t-1}r'_{t-1}A' + BH_{t-1}B', \quad H_0 = \bar{H},$$

which makes estimation challenging in the case of diagonal (when A and B are diagonal) and full (when A and B are unrestricted) BEKK models since it is difficult to impose parameter restrictions to ensure that the target $(\bar{H} - A\bar{H}A' - B\bar{H}B')$ is positive semidefinite. Fitting the model to e_t instead, as in (1), circumvents this problem and allows for diagonal

and full BEKK models to be easily fitted. In the diagonal case, the parameter restrictions needed for covariance stationarity in (1) also imply a positive semidefinite target.

2.2 Dynamic Properties

The dynamic properties can be studied when the model is vectorised, so we have

$$vec(e_t e_t') = vec(G_t) + u_t, \quad u_t = vec(e_t e_t' - G_t),$$

where

$$\begin{aligned} vec(G_t) &= vec(I - AA' - BB') + (A \otimes A)vec(e_{t-1} e_{t-1}') + (B \otimes B)vec(G_{t-1}) \\ &= vec(I - AA' - BB') + \{(A \otimes A) + (B \otimes B)\}vec(G_{t-1}) + (A \otimes A)u_{t-1}, \end{aligned}$$

noting that the vector martingale difference property $E[u_t | \mathcal{F}_{t-1}] = 0$ holds. This implies u_t is a vector weak white noise sequence.

Thus $vec(G_t)$ has a covariance stationary vector autoregression representation while

$$\begin{aligned} vec(e_t e_t') &= vec(I - AA' - BB') + (A \otimes A)vec(e_{t-1} e_{t-1}') + (B \otimes B)vec(G_{t-1}) + u_t \\ &= vec(I - AA' - BB') + \{(A \otimes A) + (B \otimes B)\}vec(e_{t-1} e_{t-1}') \\ &\quad + u_t - (B \otimes B)u_{t-1}, \end{aligned}$$

is a covariance stationary vector autoregressive moving average representation.

2.3 Leading Special Cases

2.3.1 Scalar Model

The scalar model specifies $A = \alpha^{1/2}I$ and $B = \beta^{1/2}I$. In this model all elements of G_t have the same dynamic parameters and the dynamic equations are given by

$$g_{ii,t} = (1 - \alpha - \beta) + \alpha e_{i,t-1}^2 + \beta g_{ii,t-1}, \quad i = 1, \dots, p,$$

$$g_{ij,t} = \alpha e_{i,t-1} e_{j,t-1} + \beta g_{ij,t-1}, \quad i, j = 1, \dots, p, \quad i \neq j,$$

where $g_{ij,t}$ denotes the (i, j) -th element of G_t , and we assume $\alpha > 0$ and $\beta \geq 0$. Note that if $\alpha = 0$, β is unidentified and needs to be set equal to zero indicating conditional homoskedasticity in the model, so we rule out this case. To ensure covariance stationarity, we impose $\alpha + \beta < 1$.

2.3.2 Diagonal Model

In this case, A and B are assumed to be diagonal with elements $\alpha_{ii}^{1/2} > 0$ and $\beta_{ii}^{1/2} \geq 0$, respectively. This model implies variance-targeting GARCH(1,1) models for each element of G_t taking the form

$$g_{ii,t} = (1 - \alpha_{ii} - \beta_{ii}) + \alpha_{ii} e_{i,t-1}^2 + \beta_{ii} g_{ii,t-1}, \quad i = 1, \dots, p,$$

$$g_{ij,t} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} e_{i,t-1} e_{j,t-1} + \beta_{ii}^{1/2} \beta_{jj}^{1/2} g_{ij,t-1}, \quad i, j = 1, \dots, p, \quad i \neq j.$$

The cross-equation parameter restrictions between the variance and covariance equations are a feature of BEKK models. Covariance stationarity in this model is determined by the

eigenvalues of the diagonal matrix:

$$(A \otimes A) + (B \otimes B) = \begin{pmatrix} \alpha_{11}^{1/2} A + \beta_{11}^{1/2} B & 0 & \cdots & 0 \\ 0 & \alpha_{22}^{1/2} A + \beta_{22}^{1/2} B & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \alpha_{pp}^{1/2} A + \beta_{pp}^{1/2} B \end{pmatrix}.$$

Define $\lambda_{ij} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} + \beta_{ii}^{1/2} \beta_{jj}^{1/2}$, where λ_{ij} controls the persistence in the (i, j) -th element of G_t .³ To ensure covariance stationarity, we require that

$$\max \lambda_{ij} < 1, \quad i, j = 1, \dots, p. \quad (2)$$

In practice, we impose $\lambda_{ii} := \alpha_{ii} + \beta_{ii} < 1$ by reparameterisation, which is a necessary and sufficient condition for (2) to hold; see Engle and Kroner (1995). This means that in the diagonal BEKK model it suffices to impose covariance stationarity on the conditional variances.

It will be convenient to introduce heterogeneity measures for the smoothness and persistence levels of the elements of G_t . By smoothness we refer to the coefficients β_{ii} for the conditional variances, and $\beta_{ii}^{1/2} \beta_{jj}^{1/2}$ for the conditional covariances, while λ_{ij} is the measure of persistence for the (i, j) -th element of G_t .⁴ For ease of interpretation, we do this only for the dynamic parameters of the diagonal elements of G_t (i.e. the conditional variances), noting that the dynamic parameters of the conditional covariance between assets i and j

³Recall that the GARCH(1,1) model can be written as $g_{ii,t} = (1 - \alpha_{ii} - \beta_{ii}) + (\alpha_{ii} + \beta_{ii})g_{ii,t-1} + \alpha_{ii}(e_{i,t-1}^2 - g_{ii,t-1})$, where $e_{i,t-1}^2 - g_{ii,t-1}$ is a martingale difference sequence. Thus the persistence in the conditional variance depends on the autoregression coefficient $(\alpha_{ii} + \beta_{ii})$.

⁴Brownlees (2010) is interested in similar measures for the conditional variances; however, he defines the smoothness coefficient as $\alpha_{ii}/(\alpha_{ii} + \beta_{ii})$.

are linked to the dynamic parameters of their conditional variances as shown above. Let $\mu_\alpha = p^{-1} \sum_{i=1}^p \alpha_{ii}$ denote the average estimate of α_{ii} , and $\sigma_\alpha = \sqrt{p^{-1} \sum_{i=1}^p (\alpha_{ii} - \mu_\alpha)^2}$ be a corresponding measure of heterogeneity. We define similar measures for the smoothness coefficients, β_{ii} , which are μ_β and σ_β , and also for the persistence levels, λ_{ii} , which are denoted by μ_λ and σ_λ . Note that for the scalar model, $\sigma_\alpha = \sigma_\beta = \sigma_\lambda = 0$. These measures are useful for motivating the following model.

2.3.3 Common Persistence (CP) Model

In the diagonal case, $(A \otimes A) + (B \otimes B)$ will be a diagonal matrix with diagonal elements given by $\lambda_{ij} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} + \beta_{ii}^{1/2} \beta_{jj}^{1/2}$. The CP model imposes that

$$\lambda_{ij} = \lambda,$$

for all $i, j = 1, \dots, p$, which gives the dynamic equation

$$G_t = (1 - \lambda)I + Ae_{t-1}e'_{t-1}A' + \lambda G_{t-1} - AG_{t-1}A', \quad (3)$$

where A is a diagonal matrix with diagonal elements $0 < \alpha_{ii}^{1/2} < 1$, and $0 < \lambda < 1$ is a scalar parameter satisfying $\lambda > \max \alpha_{ii}$. This model has $p + 1$ dynamic parameters as opposed to $2p$ dynamic parameters in the diagonal model. It imposes common persistence on all elements of G_t through a common eigenvalue, λ , for the dynamic equation for G_t . This can be seen from the implied variance-targeting GARCH(1,1) models for each element of G_t given by

$$g_{ii,t} = (1 - \lambda) + \alpha_{ii}e_{i,t-1}^2 + (\lambda - \alpha_{ii})g_{ii,t-1}, \quad i = 1, \dots, p,$$

$$g_{ij,t} = \alpha_{ii}^{1/2} \alpha_{jj}^{1/2} e_{i,t-1} e_{j,t-1} + (\lambda - \alpha_{ii}^{1/2} \alpha_{jj}^{1/2}) g_{ij,t-1}, \quad i, j = 1, \dots, p, \quad i \neq j.$$

The condition for covariance stationarity in this model is simply that $\lambda < 1$, which also implies a positive definite target. The model allows the different elements of G_t to load freely on the lagged variances/covariances and the corresponding shocks allowing them to have different smoothness levels; however it restricts them to have common persistence through λ . In contrast to the diagonal model, here we have $\sigma_\lambda = 0$, while $\sigma_\alpha \neq 0$ which also implies $\sigma_\beta \neq 0$.

This model is motivated by the empirical observation that persistence levels in the conditional variances of asset returns are less heterogeneous compared to their smoothness levels. For instance, Brownlees (2010) studies a large cross section of U.S. financial firms during the 2007-2009 financial crises, and finds the cross-sectional variation in λ_{ii} to be negligible, while smoothness, captured by β_{ii} in our model, tends to decline with the leverage of the company. Hafner and Franses (2009) make a related observation by noting that heterogeneity in α_{ii} is greater than that in β_{ii} , and in one of their models they impose a common smoothing parameter β . We conjecture that imposing a common eigenvalue, λ , is more intuitive since assets with different α_{ii} coefficients are also likely to display varying levels of smoothness through β_{ii} . In addition, the advantage of our specification is that a single parameter, λ , controls both covariance stationarity and positive definiteness of the target regardless of the dimensionality of the system. It also preserves the correlation targeting property which is not the case in the model of Hafner and Franses (2009).

2.3.4 Orthogonal Parameter Matrices Model

Another interesting case, which we outline here but do not pursue empirically, is when A and B are made up of orthogonal vectors

$$A = (a_1, \dots, a_p)', \quad B = (b_1, \dots, b_p)'$$

and so

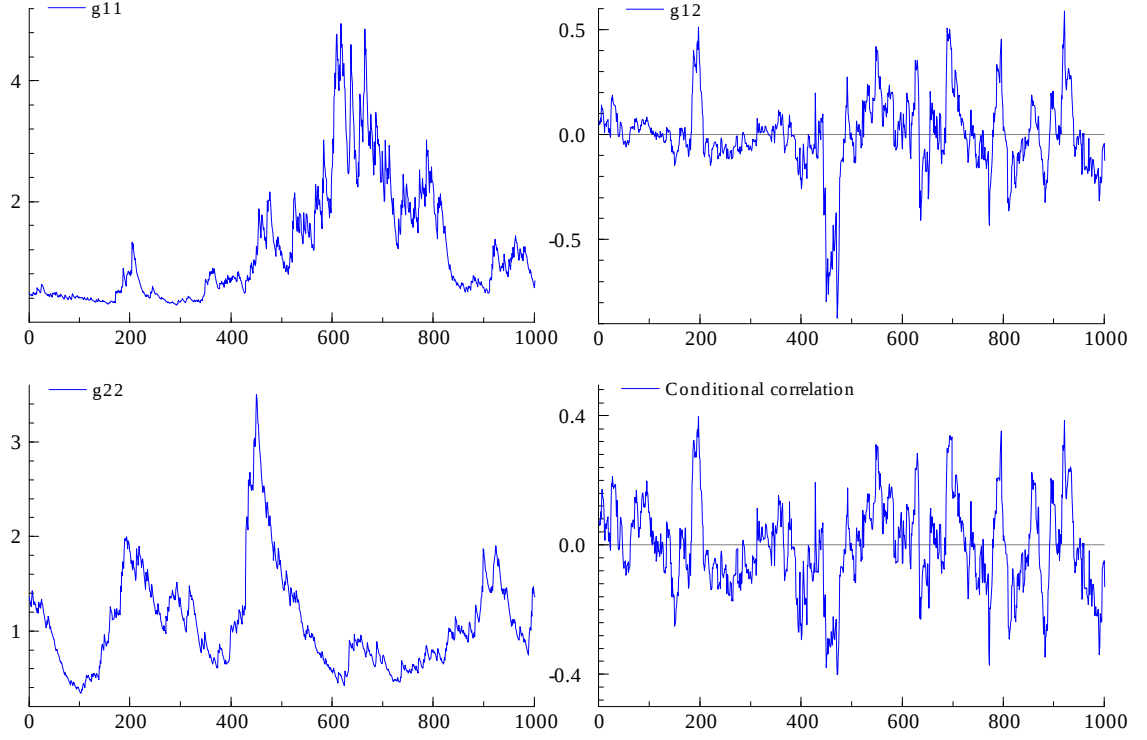
$$(AA')_{ij} = a_i' a_j = \alpha_{ij} 1_{[i=j]}, \quad (BB')_{ij} = b_i' b_j = \beta_{ij} 1_{[i=j]}, \quad i, j = 1, 2, \dots, p,$$

where $1_{[\cdot]}$ is the indicator function. Note that orthogonality of A and B implies that $I - AA' - BB'$ is diagonal. It also implies that $\text{vec}(AA') = \text{vec}(\Lambda_\alpha)$, where $\Lambda_\alpha = \text{diag}(\alpha_{11}, \dots, \alpha_{pp})$, and similarly for $\text{vec}(BB')$.

Example 1 Suppose $p = 2$ and we parameterise the orthogonal case as

$$A = \begin{pmatrix} \alpha_{11}^{1/2} & -c\alpha_{11}^{1/2} \\ c\alpha_{22}^{1/2} & \alpha_{22}^{1/2} \end{pmatrix}, \quad Ae_{t-1} = \begin{pmatrix} \alpha_{11}^{1/2} \\ c\alpha_{22}^{1/2} \end{pmatrix} e_{1,t-1} + \begin{pmatrix} -c\alpha_{11}^{1/2} \\ \alpha_{22}^{1/2} \end{pmatrix} e_{2,t-1},$$

then A is orthogonal and AA' is diagonal with first element $\alpha_{11}(1 + c^2)$, and second element $\alpha_{22}(1 + c^2)$. When $c = 0$ then A is diagonal. In this diagonal case suppose $\alpha_{11}^{1/2} = 0.275$, $\beta_{11}^{1/2} = 0.950$, $\alpha_{22}^{1/2} = 0.200$, $\beta_{22}^{1/2} = 0.980$. Figure 1 shows the sample path of G_t for 1,000 simulated observations assuming $e_{1,t}$ and $e_{2,t}$ are GARCH(1,1) processes with unconditional variance equal to 1, and persistence levels 0.995 and 0.985, respectively. Top left is $g_{11,t}$, top right is $g_{12,t}$, bottom left is $g_{22,t}$ while bottom right is the implied conditional correlation.


 Figure 1: Sample path of G_t for 1,000 simulations.

2.4 Implied BEKK Parameterisation

The model in (1) implies that

$$\begin{aligned} \text{Var}[r_t | \mathcal{F}_{t-1}] &= H_t = \bar{H}^{1/2} G_t \bar{H}^{1/2} \\ &= \bar{C}\bar{C}' + \bar{A}r_{t-1}r_{t-1}'\bar{A}' + \bar{B}H_{t-1}\bar{B}', \end{aligned}$$

where

$$\bar{A} = \bar{H}^{1/2} A \bar{H}^{-1/2}, \quad \bar{B} = \bar{H}^{1/2} B \bar{H}^{-1/2}, \quad \bar{C}\bar{C}' = \bar{H}^{1/2} (I - AA' - BB') \bar{H}^{1/2}. \quad (4)$$

So this is a particular parameterisation of the Engle and Kroner (1995) BEKK model constructed so it is relatively easy to estimate. It is also clear that this structure does not reproduce an entirely general Engle and Kroner (1995) model; rather it is a constrained version.

Example 2 *Suppose A is diagonal, then $\bar{A} = P\Lambda^{1/2}P'AP\Lambda^{-1/2}P'$ which is not symmetric in general. The same applies to $\bar{B}$ when B is diagonal.*

This means that when fitting diagonal models to e_t , this implies rather rich dynamics since the implied model for r_t will be a fully parameterised BEKK of the same order. When the asymmetric square root is used (to retrieve the standardised principal components of the data) as in the OGARCH model of Alexander (2001), a diagonal model implies $\bar{A} = P\Lambda^{1/2}A\Lambda^{-1/2}P' = PAP'$ which is diagonal, and $\bar{B}$ will also be diagonal. Thus, we prefer the symmetric square root, $P\Lambda^{1/2}P'$, since it will always give a fit that is at least as good as the fit using the asymmetric square root $P\Lambda^{1/2}$.

Example 3 *If $A = \alpha^{1/2}I$, then $\bar{A} = \alpha^{1/2}P\Lambda^{1/2}P'P\Lambda^{-1/2}P' = \alpha^{1/2}I$, and the same applies to B . Hence in the scalar case we recover the scalar BEKK model.*

It is worth noting that we use BEKK models to model the persistence in G_t , which offers an advantage over the OGARCH and GOGARCH models since the latter models assume that G_t is diagonal; these models are discussed in detail later in Section 2.6. However, we focus on fitting diagonal BEKK models which means the parameters are estimated to fit both the conditional variances and covariances. To the extent that the different elements of G_t have different dynamics, the diagonal BEKK model could potentially lead to a worse fit compared to OGARCH/GOGARCH since the former imposes cross-equation

parameter restrictions between the variance and covariance equations. The class of DCC models, which we discuss next, offers more flexibility in this regard and our empirical results indicate its superiority to both BEKK and OGARCH/GOGARCH models.

2.5 DCC Models

2.5.1 Scalar DCC Dynamics

One shortcoming of diagonal and common persistence BEKK models is that the dynamics of $g_{ij,t}$ is linked to the dynamics of $g_{ii,t}$ and $g_{jj,t}$ for all i and j through cross-equation parameter restrictions. This is partly overcome in the DCC model of Engle (2002), which allows for the speed of change in the conditional correlations to be different than that seen for the individual volatilities, and also allows for models to be fit in quite large dimensions. See the discussion in Engle (2009a).

DCC models work through first modelling the marginal conditional variances,

$$\text{Var}[r_{i,t}|\mathcal{F}_{t-1}^{r_i}] = \sigma_{i,t}^2, \quad i = 1, 2, \dots, p,$$

as univariate GARCH processes. This is an important constraint since in effect it is modelling the conditional variance using its own univariate natural filtration, $\mathcal{F}_{t-1}^{r_i}$. It then computes the standardised potentially correlated innovations

$$v_{i,t} = r_{i,t}/\sigma_{i,t}, \quad i = 1, 2, \dots, p.$$

Let $v_t = (v_{1,t}, v_{2,t}, \dots, v_{p,t})'$ and the unconditional covariance

$$\Pi_C = \text{Var}[v_t],$$

then we model

$$c_{ij,t} = \text{Corr}[v_{i,t}, v_{j,t} | \mathcal{F}_{t-1}], \quad i, j = 1, 2, \dots, p.$$

The scalar DCC model decomposes $C_t = [c_{ij,t}]$ as

$$C_t = (Q_t \circ I)^{-\frac{1}{2}} Q_t (Q_t \circ I)^{-\frac{1}{2}},$$

where $\circ$ denotes the Hadamard elementwise product, and Q_t follows a targeted scalar BEKK model

$$Q_t = (1 - \alpha - \beta) \Pi_C + \alpha v_{t-1} v'_{t-1} + \beta Q_{t-1},$$

where α and β satisfy restrictions similar to the scalar BEKK model; see Section 2.3.1. This ensures that C_t is a genuine correlation matrix.⁵ We will call this a scalar DCC model and denote it by S-DCC. The predecessor to the DCC model is the Constant Conditional Correlations (CCC) model of Bollerslev (1990) which sets $C_t = C$, where C is the unconditional correlation matrix of v_t .

2.5.2 Flexible DCC Dynamics

The more flexible BEKK-type specifications discussed above suggest similar extensions to the scalar DCC model. Note that a generalisation of the scalar DCC is already mentioned in Engle (2002) but it is not pursued empirically. Cappiello et al. (2006) propose more

⁵See Aielli (2006) for a twist on the usual DCC dynamics which has better theoretical properties.

flexible dynamics to the scalar DCC model with asymmetric effects, and they estimate a diagonal DCC model for a group of 34 assets. Billio et al. (2006) and Billio and Caporin (2009) fit a restricted diagonal DCC model to 20 assets assuming a sector-specific block structure in A and B such that each matrix has only 3 parameters. Hafner and Franses (2009) introduce a flexible diagonal DCC specification, and apply it to 39 stocks. They overcome the estimation challenge in this high dimension by using a pooled estimator based on Engle et al. (2008).

Based on the standardised returns, v_t , let

$$\Pi_C = \text{Var}[v_t] = P_C \Lambda_C P_C',$$

where P_C contains the eigenvectors and Λ_C has the eigenvalues on the main diagonal. Then we construct the rotated innovations

$$w_t = P_C \Lambda_C^{-1/2} P_C' v_t.$$

The virtue of this approach is that $\text{Var}[w_t] = I$. Then we model

$$\text{Var}[w_t | \mathcal{F}_{t-1}] = Q_t^*,$$

where

$$Q_t^* = (I - AA' - BB') + Aw_{t-1}w_{t-1}'A + BQ_{t-1}^*B, \quad Q_0^* = I.$$

As shown for the BEKK parameterisation, Q_t is given by

$$Q_t = P_C \Lambda_C^{1/2} P_C' Q_t^* P_C \Lambda_C^{1/2} P_C'.$$

In the case where $A = \alpha^{1/2}I$ and $B = \beta^{1/2}I$, we reproduce the scalar DCC model. However, the previous sections show we can simply extend this by allowing A and B to be diagonal in the recursion for Q_t^* . This added flexibility may be empirically useful, allowing some aspects of the correlation matrix to move more rapidly than others. We also consider a version similar to the CP model defined in (3).

To conclude, the scalar DCC model is a scalar BEKK model applied to the standardised residuals after fitting univariate GARCH models. The diagonal DCC model, denoted by D-DCC, is simply a diagonal BEKK model applied to the same innovations. The CP model discussed in Section 2.3.3, as a special case of diagonal models with only $p + 1$ dynamic parameters, is somewhat related to one of the proposed models in Hafner and Franses (2009). The distinction is that Hafner and Franses (2009) impose a common smoothing parameter β on the system, while we impose common persistence through λ . It is important to note that the model of Hafner and Franses (2009) loses the correlation targeting property, while our model preserves this attractive feature.

2.6 Relation to Orthogonal GARCH Models

In this subsection, we take a step back from the different specifications of our model to discuss how it generally relates to some recent propositions known in the literature as orthogonal GARCH models. In writing $r_t = \bar{H}^{1/2} e_t$ and then modelling $e_t = \bar{H}^{-1/2} r_t$, we are effectively analysing the rotated returns. A number of models have focused on linear

transformations of the form

$$r_t = Ze_t,$$

where Z is some invertible matrix. Consider the polar decomposition

$$Z = SU, \tag{5}$$

where S is a symmetric positive definite matrix, and U is an orthogonal matrix. Since $\text{Var}[e_t] = I$, we have $\text{Var}[r_t] = ZZ' = S^2$, thus S is the symmetric square root of $\text{Var}[r_t]$ given by $P\Lambda^{1/2}P'$. Therefore part of the matrix Z can be estimated using only unconditional information.

The orthogonal GARCH (OGARCH) model of Alexander and Chibumba (1997) and Alexander (2001) assumes $U = P$, hence $Z = P\Lambda^{1/2}$, which is the asymmetric square root of $\text{Var}[r_t]$. In this case e_t is a vector of the standardised principal components of r_t which are unconditionally uncorrelated by construction. Alexander (2001) assumes that these standardised principal components are also *conditionally* uncorrelated with a diagonal time-varying covariance matrix. This is a misspecification since the standardised principal components will inherit the heteroskedastic properties of the original returns. The generalised OGARCH model (GOGARCH) of van der Weide (2002) proposes $Z = P\Lambda^{1/2}U^*$, where the orthogonal link matrix, U^* , is to be estimated using conditional information. This is sought to avoid identification problems; see van der Weide (2002) for details.

Lanne and Saikkonen (2007) propose the polar decomposition in (5) and they use conditional information to estimate U under the assumption that some of the estimated components are homoskedastic which leads to dimension reduction.⁶ Fan et al. (2008)

⁶Lanne and Saikkonen (2007) focus on a reduced-factor model, while here we focus on the dynamics of

estimate U under the condition that the resulting components, e_t , are also conditionally uncorrelated. Compared to Fan et al. (2008), the models of Alexander (2001), van der Weide (2002) and Lanne and Saikkonen (2007) can all be seen as approximations since they assume that the components estimated from their models are conditionally uncorrelated, while in fact they are only unconditionally uncorrelated.

While the model of Fan et al. (2008) is conceptually appealing, a set of conditionally uncorrelated components may not exist. Thus, in practice, their model may only give components that are the *least* conditionally correlated in-sample. It is worth noting that estimating the conditionally uncorrelated components in Fan et al. (2008) requires solving an $O(p^2)$ optimisation problem which may be infeasible for large dimensions. In addition, they note that as p increases, it becomes more difficult to find factors that are conditionally uncorrelated using their proposed method. Boswijk and van der Weide (2011) adopt a closely related approach to estimate conditionally uncorrelated factors, which departs from earlier work on the GOGARCH model. It is unclear whether their approach guarantees the success of finding conditionally uncorrelated factors in large dimensions.⁷

Our model takes a different stand by directly modelling the conditional covariance matrix of e_t . Here we simply set U in (5) equal to I , which means that e_t is not going to be the only unique set of components satisfying $\text{Var}[e_t] = I$. For instance, we can post-multiply $\bar{H}^{1/2} = P\Lambda^{1/2}P'$ by an arbitrary orthogonal matrix U^* , and still have $\text{Var}[e_t] = \text{E}[e_t e_t'] = \text{E}[U^{*'} \bar{H}^{-1/2} r_t r_t' \bar{H}^{-1/2} U^*] = I$. However, uniqueness (or identifiability) of e_t is not crucial since our objective is to simplify estimation and not get unique estimates of e_t . What is important is that for any model for e_t , it is straightforward to derive the implied

the full set of returns.

⁷In all of these studies, the maximum number of assets considered in simulation experiments and for empirical analysis is 12 assets.

model for r_t as we discussed in Section 2.4.

For the models we fit, we include the OGARCH model of Alexander (2001) for comparison. This is equivalent to the following dynamic equation

$$G_t = \left(I - \tilde{A}\tilde{A}' - \tilde{B}\tilde{B}' \right) + \tilde{A}\tilde{A}' \circ (e_{t-1}e_{t-1}') + \tilde{B}\tilde{B}' \circ G_{t-1}, \quad (6)$$

where $\tilde{A}$ and $\tilde{B}$ are diagonal. Note that this equation is for the conditional covariance matrix of the standardised principal components of the returns, i.e. when using the asymmetric square root $\bar{H}^{1/2} = P\Lambda^{1/2}$. We also include results for the GOGARCH model of van der Weide (2002) but modified as proposed by Boswijk and van der Weide (2011). In this GOGARCH formulation, it is assumed that the transformation matrix, Z , is given by

$$Z = SU(\delta) = P\Lambda^{1/2}P'U(\delta), \quad (7)$$

where the orthogonal matrix $U(\delta)$ is parameterised by a $p(p-1)/2 \times 1$ vector δ , with j -th element $-180 \leq \delta_j \leq 180$ which is a rotation angle.⁸ The dynamics of the resulting $e_t = Z^{-1}r_t$ are modelled as in (6). In models of large dimension, estimating δ is generally challenging, thus we only include the GOGARCH model for comparison in our empirical analysis in the bivariate case. Note that our model imposes $U(\delta) = I$, or equivalently $\delta = 0$.

To summarise, the key feature of OGARCH and GOGARCH models is that conditionally the factors are assumed to be uncorrelated. This is not true of (1), which assumes

⁸Note that any 2×2 orthogonal matrix can be written as a rotation matrix taking the form $U(\delta) = \begin{pmatrix} \cos \delta & -\sin \delta \\ \sin \delta & \cos \delta \end{pmatrix}$ where $-180 \leq \delta \leq 180$, which is scalar in this example, is a rotation angle. A positive δ indicates counterclockwise rotation. For $p > 2$, $U(\delta)$ can be represented as the product of $p(p-1)/2$ rotation matrices each parameterised with a distinct rotation angle; see van der Weide (2002) for details.

they follow a BEKK-type model. The models are not the same even in the scalar BEKK case, hence these models are non-nested. In (1) when A and B are diagonal, the diagonal elements of G_t follow similar dynamics to the OGARCH/GOGARCH model as in (6). The models differ by the non-diagonal elements of G_t which are always assumed to be zero in the OGARCH/GOGARCH structure. This means that the marginal likelihoods for the univariate series $e_{i,t}$, $i = 1, \dots, p$, are the same (holding the parameters equal across the models) but their dependence structure will be different.

2.7 A Time-Varying-Weight Strict Factor Model Representation

Our model can also be interpreted as a time-varying-weight strict factor model. The model implies

$$\begin{aligned} \text{Var}[r_t | \mathcal{F}_{t-1}] &= H_t = \bar{H}^{1/2} G_t \bar{H}^{1/2} \\ &= P \Lambda^{1/2} P' G_t P \Lambda^{1/2} P'. \end{aligned}$$

Suppose we take the spectral decomposition of G_t at each point in time such that $G_t = P_t^G \Lambda_t^G (P_t^G)'$, where P_t^G contains the eigenvectors of G_t and the diagonal matrix Λ_t^G has the eigenvalues of G_t along its main diagonal. Then we can write

$$\begin{aligned} H_t &= P \Lambda^{1/2} P' P_t^G \Lambda_t^G (P_t^G)' P \Lambda^{1/2} P' \\ &= z_t \Lambda_t^G z_t', \end{aligned}$$

where z_t is a time-varying weight matrix. This representation is reminiscent of strict factor models where the factors are not correlated, their conditional variances are given by the

diagonal elements of the time-varying Λ_t^G , and there is no approximation error covariance since the number of factors is equal to the number of assets.

The term strict factor model is usually used to characterise a model where the idiosyncratic components of asset returns are uncorrelated as in Ross (1976), for example. Here we adapt it to describe a model where the factors are uncorrelated both conditionally and unconditionally, and the factor loadings, z_t , are time-varying.

Note that orthogonal GARCH models assume that G_t is diagonal, and in this case $\Lambda_t^G = G_t$ while $P_t^G = I$. Thus orthogonal GARCH models impose a fixed weight matrix $z_t = z$. This representation provides an additional intuition behind our model, and explains why capturing the covariance dynamics of $e_{i,t}$, $i = 1, \dots, p$, is important. Since we also consider DCC-type parameterisations, this analogy can be extended to the factor DCC model of Rangel and Engle (2011) which, if reparameterised as above, becomes a time-varying-weight strict factor model.

3 Inference

3.1 Parameter Vector

We will focus on the two part model, where the first part is

$$E[r_t] = 0, \quad \text{Var}[r_t] = \overline{H} = P\Lambda P', \quad t = 1, 2, \dots, T,$$

and the second is

$$e_t = P\Lambda^{-1/2}P'r_t, \quad E[e_t|\mathcal{F}_{t-1}] = 0, \quad \text{Var}[e_t|\mathcal{F}_{t-1}] = G_t,$$

and

$$G_t = (I - AA' - BB') + Ae_{t-1}e'_{t-1}A' + BG_{t-1}B', \quad G_0 = I.$$

Let θ_A and θ_B denote the parameters indexing A and B . The parameters in the model are

$$\theta = (\text{vech}(\overline{H})', \theta'_A, \theta'_B)' = (\theta'_{\overline{H}}, \theta'_*)'.$$

We call θ_* the ‘dynamic’ parameters and $\theta_{\overline{H}}$ the ‘static’ parameters. The true values of these parameters are denoted by $\theta_{0,*}$ and $\theta_{0,\overline{H}}$, respectively, while $\theta_0 = (\theta'_{0,\overline{H}}, \theta'_{0,*})'$. Typically the dimension of $\theta_{\overline{H}}$ is large and potentially massive if p is large since it has $O(p^2)$ elements. The dimension of θ_* is often small with only $O(p)$ parameters in the specifications we consider.

In the diagonal case, let $\theta_{*,i}$ denote the dynamic parameters which index the dynamics of the i -th series $e_{i,t}$. Thus $\theta_{*,i} = (\alpha_{ii}, \beta_{ii})$, and $\theta_* = (\theta'_{*,1}, \theta'_{*,2}, \dots, \theta'_{*,p})'$, recalling that p is the number of assets. This notation will be useful later when discussing the numerical optimisation algorithm we use for diagonal models.

3.2 Two Step Estimation

The structure of the model allows for a two-step estimation strategy to estimate θ . This approach, which dramatically eases the computational burden, was advocated in the univariate case by Engle and Mezrich (1996) and has been used for the scalar BEKK and DCC models in many studies.

In the first step we focus solely on the static parameters $\theta_{\overline{H}} = \text{vech}(\overline{H})$. By construction

$\overline{H} = \text{Var}[r_t]$, thus we use the method of moments estimator

$$\widehat{\overline{H}} = \frac{1}{T} \sum_{t=1}^T r_t r_t',$$

implying $\widehat{\theta}_{\overline{H}}$. This estimate is then decomposed into $\widehat{P}$ and $\widehat{\Lambda}$. Then we construct the time series of rotated returns

$$e_t = \widehat{P} \widehat{\Lambda}^{-1/2} \widehat{P}' r_t, \quad t = 1, 2, \dots, T.$$

The second stage estimation is based on the quasi-likelihood

$$\log L(\theta_*, \widehat{\theta}_{\overline{H}}) = \sum_{t=1}^T \log L_t(\theta_*, \widehat{\theta}_{\overline{H}}) = \text{const} - \frac{1}{2} \sum_{t=1}^T \log |G_t| - \frac{1}{2} \sum_{t=1}^T e_t' G_t^{-1} e_t, \quad (8)$$

where

$$G_t = (I - AA' - BB') + A e_{t-1} e_{t-1}' A' + B G_{t-1} B', \quad G_0 = I. \quad (9)$$

This is optimised solely over θ_* , keeping $\widehat{\theta}_{\overline{H}}$ fixed, which delivers $\widehat{\theta}_*$. If the dimensionality of the system, p , becomes very large then it may be worth switching over to use a composite likelihood (Engle et al., 2008, and Pakel et al., 2011) or the McGyver estimation method (Engle, 2009b), but we will not discuss that here.

When estimating the OGARCH model, we use $e_t = \widehat{\Lambda}^{-1/2} \widehat{P}' r_t$ in (8) while the dynamic equation (9) is replaced with (6). For GOGARCH we use $e_t = \widehat{P} \widehat{\Lambda}^{-1/2} \widehat{P}' r_t$ in the following quasi-likelihood

$$\log L(\theta_*, \widehat{\theta}_{\overline{H}}) = \sum_{t=1}^T \log L_t(\theta_*, \widehat{\theta}_{\overline{H}}) = \text{const} - \frac{1}{2} \sum_{t=1}^T \log |G_t| - \frac{1}{2} \sum_{t=1}^T e_t' U(\delta) G_t^{-1} U(\delta)' e_t, \quad (10)$$

and the dynamic equation for G_t is also given by (6). In this case, the additional $p(p-1)/2$ in δ are contained in θ_* .⁹

In terms of asymptotic theory, for fixed p and $T \rightarrow \infty$, this is simply a two step moment estimator, e.g. Newey and McFadden (1994) and Pagan (1986), where the moment conditions are given by the vector

$$\begin{aligned} m(\theta_*, \theta_{\overline{H}}) &= \sum_{t=1}^T m_t(\theta_*, \theta_{\overline{H}}), \quad m_t(\theta_*, \theta_{\overline{H}}) = \begin{pmatrix} \theta_{\overline{H}} - \text{vech}(r_t r_t') \\ \frac{\partial \log L_t(\theta_*, \theta_{\overline{H}})}{\partial \theta_*} \end{pmatrix}, \\ m(\hat{\theta}_*, \hat{\theta}_{\overline{H}}) &= 0, \end{aligned}$$

and

$$\mathbb{E} \left\{ m(\theta_*, \theta_{\overline{H}}) \Big|_{\theta_* = \theta_{0,*}; \theta_{\overline{H}} = \theta_{0,\overline{H}}} \right\} = 0,$$

at the true values. The key feature here is that the first step does not involve θ_* , which simplifies the estimation of the dynamic parameters in the second step.

The asymptotic distribution of this two step estimator has been worked over by many authors in the context of scalar BEKK models and the DCC model, so we will not discuss it in detail here. Under standard regularity conditions, as $T \rightarrow \infty$ we have

$$\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \mathcal{I}^{-1} \mathcal{J}(\mathcal{I}^{-1})')$$

⁹As noted earlier the estimation of δ is challenging when p is large. Thus we only estimate the GOGA-RCH model in the bivariate case in our empirical analysis.

where $\widehat{\theta} = (\widehat{\theta}'_*, \widehat{\theta}'_{\overline{H}})'$,

$$\mathcal{J} = \text{Var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T m_t(\theta_*, \theta_{\overline{H}}) \right], \quad \mathcal{I} = \text{E} \left[\frac{\partial m_t(\theta_*, \theta_{\overline{H}})}{\partial \theta'} \right],$$

and we use a HAC estimator, e.g. Newey and West (1987), to estimate $\mathcal{J}$.

3.3 Numerical Optimisation

General numerical optimisation routines can be used to locate $\widehat{\theta}_*$. An alternative, which we have used systematically in the estimation of diagonal models, is to employ a zig-zag algorithm based upon the structure of θ_* . We optimise

$$\log L(\theta_{*i}, \theta_{*\setminus i}, \widehat{\theta}_{\overline{H}}),$$

with respect to θ_{*i} , holding all other elements of θ_* , written as $\theta_{*\setminus i}$, at the previously best values. We then cycle over i , repeating the optimisation each time. The advantage of this is that each individual optimisation is only 2-dimensional, and we have found this method to be reliable. The theory for this estimator is discussed in Fan et al. (2007), while inference is standard as outlined in Section 3.2.

3.4 Model Comparison

We will use a quasi-likelihood criterion for r_t to compare the fit of the different models, which means we will focus on the 1-step prediction ability of the models using a Kullback-Leibler distance. Note that given the likelihood for e_t , it is straightforward to compute the likelihood for r_t since the Jacobian of the transformation is $\frac{\partial r_t}{\partial e_t} = P\Lambda^{1/2}P'$, and its determinant is $|P\Lambda^{1/2}P'| = |\Lambda^{1/2}|$, where the second equality follows from P being or-

thogonal; see Lütkepohl (1996, pp. 48). Thus for a time series of length T , we have that $\log L_r = \log L_e - \frac{T}{2} \log |\Lambda|$, where $\log L_r$ and $\log L_e$ denote the log-likelihoods for r_t and e_t , respectively. This also implies that comparisons based on models for e_t is equivalent to comparisons based on equivalent models for r_t . This is because the difference in the likelihood is independent of the dynamic parameters, and only depends on the static parameters, Λ , which are common to all the models we consider.

Let $\log L_{a,t}$ denote the t -th observation log-likelihood for r_t based on model a . To compare two models, a and b , we look at the average log-likelihood difference

$$l_{a,b} = \frac{1}{T} \sum_{t=1}^T l_{a,b,t}, \quad l_{a,b,t} = \log L_{a,t} - \log L_{b,t}. \quad (11)$$

We then test if $l_{a,b}$ is statistically significantly different than zero by computing a HAC estimator of the variance of $l_{a,b}$. This predictive ability test was first introduced by Diebold and Mariano (1995). Using a quasi-likelihood criterion is valid for non-nested and misspecified models; see Cox (1961) and Vuong (1989) for in-sample model comparison, and Amisano and Giacomini (2007) for out-of-sample model selection. For comparisons, we choose the diagonal model within each class (BEKK, DCC, OGARCH and GOGARCH) since it is the most flexible specification, and then test for equal predictive ability. We will use either OGARCH or GOGARCH in a comparison, but not both since the GOGARCH nests OGARCH and thus this test would not be appropriate.¹⁰

¹⁰If interest is in testing nested models, the approach of Giacomini and White (2006) can be adopted by using rolling-window estimation. This allows for pairwise comparisons of the predictive ability of all the four classes of models as well as the different variants under each class.

3.5 Copula and Marginal Likelihoods

It is also useful to consider the marginal log-likelihood for the i -th series

$$\log L_i = \sum_{t=1}^T \log f(r_{i,t} | \mathcal{F}_{t-1}),$$

where we have conditioned on the entire filtration, not just the natural filtration for the i -th series. The implied copula likelihood is then given by

$$\log L = \sum_{i=1}^p \log L_i.$$

Under the assumption of conditional normality, the copula parameter is the conditional correlation matrix of the returns. For the copula-margins decomposition in the CCC and DCC models, see, respectively, equation (6) in Bollerslev (1990) and equation (26) in Engle (2002).

4 Empirical Analysis

4.1 Data

We use close-to-close daily returns data on Spyder (SPY), an S&P 500 exchange traded fund, and some of the most liquid stocks in the Dow Jones Industrial Average (DJIA) index. These are: Alcoa (AA), American Express (AXP), Bank of America (BAC), Coca Cola (KO), Du Pont (DD), General Electric (GE), International Business Machines (IBM), JP Morgan (JPM), Microsoft (MSFT), and Exxon Mobil (XOM). The sample period is 1/2/2001 to 31/12/2009 and the source of the data is Yahoo!Finance, which is accessible

online. We use close prices adjusted for dividends and splits.

Our primary empirical example in Section 4.3 focuses on the pair XOM-AA, which we use to present the models' main features. In Section 4.4, we analyse stock-index dynamics by studying the pair SPY-XOM. This sheds light on the conditional correlation of a firm's stock with the overall market index, and the latter part of our sample includes the recent financial crisis. This application relates to the recent work of Brownlees and Engle (2010) and Hansen et al. (2010) where they focus on modelling systemic risk measures using conditional correlations and conditional betas, respectively. See also Noureldin et al. (2011) for a multivariate volatility model for the same group of assets which utilises high frequency data. In Section 4.5 we estimate the models using all 10 stocks from the DJIA index.

4.2 Considered Models

4.2.1 BEKK Class

We work with the rotated returns $e_t = \hat{P}\hat{\Lambda}^{-1/2}\hat{P}'r_t$, which are unconditionally uncorrelated in-sample and each has unconditional variance equal to 1. They display, of course, volatility clustering. Then we fit the covariance targeting BEKK model

$$\text{Var}[e_t|\mathcal{F}_{t-1}] = G_t = (I - AA' - BB') + Ae_{t-1}e_{t-1}'A' + BG_{t-1}B', \quad G_0 = I.$$

The dynamics are estimated using a Gaussian quasi-likelihood. We fit the following models:

- Scalar BEKK (S-BEKK). $A = \alpha^{1/2}I$ and $B = \beta^{1/2}I$.
- Diagonal BEKK (D-BEKK). $A = \text{diag}(\alpha_{11}^{1/2}, \dots, \alpha_{pp}^{1/2})$ and $B = \text{diag}(\beta_{11}^{1/2}, \dots, \beta_{pp}^{1/2})$.
- Diagonal BEKK with common persistence (D-BEKK-CP). $A = \text{diag}(\alpha_{11}^{1/2}, \dots, \alpha_{pp}^{1/2})$

and λ is the common persistence parameter.

For comparison, we also report results for these three specifications when applied to OGARCH-type and GOGARCH-type models, where in the latter models it is assumed that G_t is diagonal.¹¹ The diagonal OGARCH and GOGARCH models (with unconstrained $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$) correspond to the models of Alexander (2001) and Boswijk and van der Weide (2011), respectively, while the other specifications are novel in this context.

4.2.2 DCC Class

We first fit variance targeting univariate GARCH(1,1) models to the returns, which produces a sequence of standardised vector innovations v_t . Then we model $c_{ij,t} = \text{Corr}[v_{i,t}, v_{j,t} | \mathcal{F}_{t-1}]$. The conditional correlation matrix $C_t = [c_{ij,t}]$ is decomposed as

$$C_t = (Q_t \circ I)^{-\frac{1}{2}} Q_t (Q_t \circ I)^{-\frac{1}{2}}.$$

We first rotate v_t to generate $w_t = P_C \Lambda_C^{-1/2} P_C' v_t$, then we model $\text{Var}[w_t | \mathcal{F}_{t-1}] = Q_t^*$ and then take $Q_t = P_C \Lambda_C^{1/2} P_C' Q_t^* P_C \Lambda_C^{1/2} P_C'$. The dynamic equation for Q_t^* is

$$Q_t^* = (I - AA' - BB') + Aw_{t-1}w_{t-1}'A + BQ_{t-1}^*B, \quad Q_0^* = I,$$

which is estimated using a Gaussian quasi-likelihood. We estimate the following models:

- Constant conditional correlations (CCC). $A = B = 0$.
- Scalar DCC (S-DCC). $A = \alpha^{1/2}I$ and $B = \beta^{1/2}I$.

¹¹Note that the OGARCH model is for $e_t = \hat{\Lambda}^{-1/2} \hat{P}' r_t$, while the GOGARCH model is for $e_t = \hat{P} \hat{\Lambda}^{-1/2} \hat{P}' r_t$ and the latter's likelihood is given by (10).

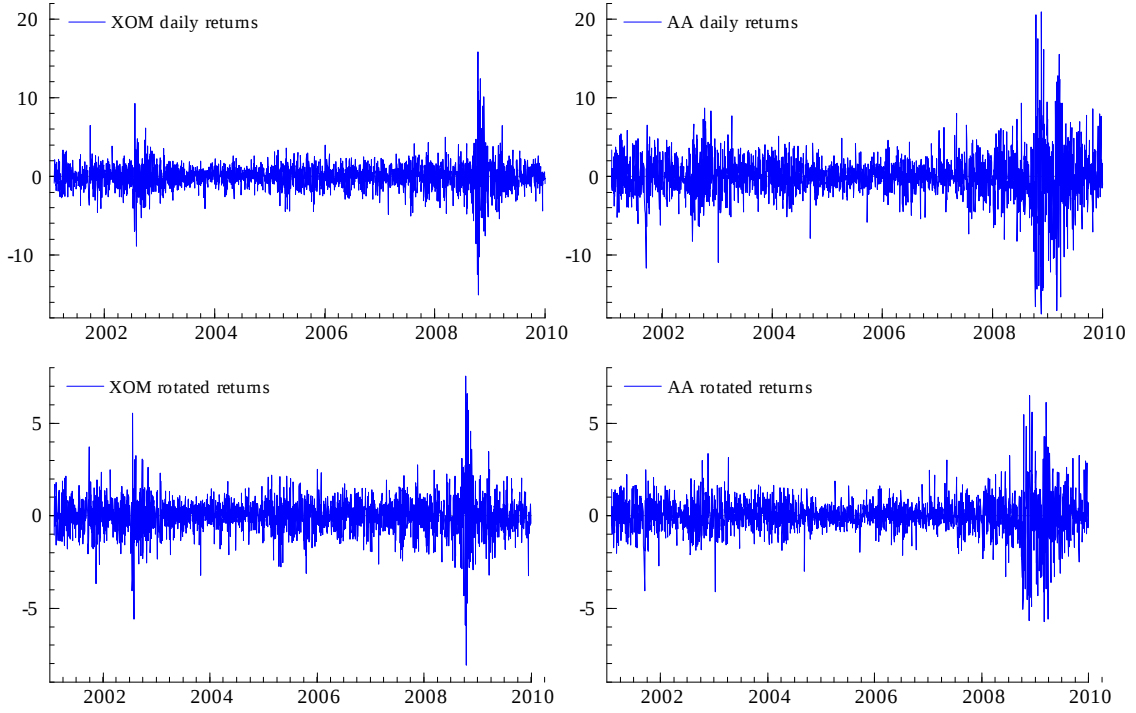


Figure 2: XOM and AA series: Top panel plots the daily returns (r_t). Bottom panel plots the rotated returns (e_t).

- Diagonal DCC (D-DCC). $A = \text{diag}(\alpha_{11}^{1/2}, \dots, \alpha_{pp}^{1/2})$ and $B = \text{diag}(\beta_{11}^{1/2}, \dots, \beta_{pp}^{1/2})$.
- Diagonal DCC with common persistence (D-DCC-CP). $A = \text{diag}(\alpha_{11}^{1/2}, \dots, \alpha_{pp}^{1/2})$ and λ is the common persistence parameter.

4.3 Analyzing the Pair XOM-AA

4.3.1 BEKK, OGARCH and GOGARCH Models

We will start out with a detailed bivariate example: the daily returns of Exxon Mobil (XOM) and Alcoa (AA). Figure 2 provides summary of the series. The daily returns are in the upper panel while the rotated returns are in the lower panel. The unconditional covariance matrix of the returns is given in Table 1. The first eigenvector looks like a

	Returns				GARCH(1,1) innovations			
	Covariance		Eigenvectors		Covariance		Eigenvectors	
XOM	3.069	2.918	0.394	-0.919	0.981	0.470	0.706	-0.708
AA	2.918	8.633	0.919	0.394	0.470	0.983	0.708	0.706
Eigenvalues	–	–	9.882	1.820	–	–	1.452	0.512

Table 1: Left-hand side is the unconditional covariance of returns, together with their eigenvalues and (normalised) eigenvectors. On the right-hand side is the unconditional covariance of the innovations from univariate variance targeting GARCH(1,1) models.

market factor, while the second is a long/short portfolio.

The parameter estimates of the BEKK, OGARCH and GOGARCH models are given in Table 2 together with the associated log-likelihood values for the (unrotated) returns evaluated at $(\hat{\theta}_*, \hat{\theta}_H)$. The joint log-likelihood is decomposed to indicate the performance in terms of the margins and the copula. In the BEKK class, the D-BEKK model provides a moderate improvement in fit compared to S-BEKK. This is due to the diagonal parameters freely fitting each conditional variance. The effects are quite considerable since α_1 and α_2 are an order of magnitude different than the S-BEKK's α , so that XOM's conditional variance dynamics are much more responsive to its own shock, while the estimates for the conditional variance of AA are smoother. Of course, these estimates also fit the conditional covariance dynamics given the cross-equation parameter restrictions of the diagonal BEKK model.

	BEKK			OGARCH			GOGARCH		
	Scalar	Diagonal	CP	Scalar	Diagonal	CP	Scalar	Diagonal	CP
α	0.050	—	—	0.066	—	—	0.063	—	—
β	0.943	—	—	0.924	—	—	0.927	—	—
α_{11}	—	0.072	0.055	—	0.060	0.076	—	0.088	0.067
α_{22}	—	0.036	0.044	—	0.082	0.059	—	0.041	0.055
β_{11}	—	0.909	—	—	0.934	—	—	0.884	—
β_{22}	—	0.960	—	—	0.887	—	—	0.955	—
λ	—	—	0.993	—	—	0.989	—	—	0.991
δ	—	—	—	—	—	—	0.015	-0.008	0.030
<i>LL decomposition</i>									
Margin (XOM)	-4,034	-4,030	-4,032	-4,028	-4,031	-4,027	-4,029	-4,030	-4,028
Margin (AA)	-5,098	-5,098	-5,099	-5,099	-5,126	-5,102	-5,100	-5,097	-5,099
Copula	284	288	284	233	270	235	258	266	258
Total LL	-8,848	-8,840	-8,847	-8,894	-8,887	-8,893	-8,870	-8,861	-8,869

Table 2: Dataset: XOM and AA daily returns 1/2/2001-31/12/2009. Top Panel: parameter estimates of the scalar, diagonal, and common persistence (CP) parameterisations for the BEKK, OGARCH and GOGARCH models. α and β are the parameters of the scalar models, while $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, are those of the diagonal models. For CP, λ (the common persistence parameter) and α_{ii} for each asset are reported. δ is the rotation angle in the bivariate GOGARCH model. All parameters, except for δ , are statistically significant at 5 percent. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

The parameters of the implied BEKK model for r_t , given by (4), are

$$\bar{A} = \bar{H}^{1/2} A \bar{H}^{-1/2} = \begin{pmatrix} 0.275 & -0.019 \\ 0.034 & 0.182 \end{pmatrix}, \quad \bar{B} = \bar{H}^{1/2} B \bar{H}^{-1/2} = \begin{pmatrix} 0.951 & 0.007 \\ -0.012 & 0.983 \end{pmatrix},$$

indicating that a diagonal model for the rotated returns implies a full BEKK model for the unrotated returns. Recall that this follows from specifying $\bar{H}^{1/2}$ as the symmetric square root using the spectral decomposition. The D-BEKK-CP model estimates imply roughly the same level of persistence in the elements of G_t as the S-BEKK and D-BEKK models. The picture for OGARCH and GOGARCH is rather similar but indicating a slightly lower level of persistence.

Interestingly, the GOGARCH model's estimated rotation angle is very close to zero and statistically insignificant. This implies that $U(\delta) \approx I$, making the e_t series from the GOGARCH model very close to those from the BEKK model; see (7). The primary difference between the two models is that GOGARCH assumes that $g_{12,t}$ is zero, which is reflected in BEKK's superior copula fit.

The BEKK models provide an important increase in the likelihood compared to OGARCH and GOGARCH. The increase in the log-likelihood in BEKK models is primarily due to an increase in the copula fit, implying that capturing the conditional correlations in the rotated returns (which is not the case in OGARCH and GOGARCH) does improve the modelling of the conditional correlations of the unrotated returns. There is a small loss in fit in the first margin (XOM) when using the BEKK model, however this is more than compensated through capturing the conditional correlation dynamics with BEKK models

providing an overall gain in fit.

4.3.2 DCC Models

Table 3 gives estimates of the CCC and DCC models. When estimating the variance targeting GARCH(1,1) models for the margins, we first standardise the returns of XOM and AA by their respective unconditional variances, fit variance targeting GARCH(1,1) models for these standardised returns and report the log-likelihood for the original returns as the marginal log-likelihood. The estimates suggest different dynamics for the two series, which can already be inferred from the improvement offered by the diagonal models in Table 2. Not surprisingly, the fit for the margins in this case is better than all the BEKK, OGARCH and GOGARCH models. For CCC the unconditional correlation of the standardised returns is 0.480. We use the unconditional correlation to build the time-varying covariance matrix, the dynamics of which are driven only by the conditional variances in this model.

The estimates for the DCC dynamics suggest only a marginal improvement by the D-DCC and D-DCC-CP over S-DCC. With the margins fit freely, there seems to be no additional improvement from further enriching the DCC dynamics in this case. This is in contrast to the BEKK model results, but it is perhaps unsurprising since there is a single conditional correlation to model in this case. As we show later, in higher dimensions the gains from the further flexibility of the D-DCC and D-DCC-CP models can be substantial. Overall the estimates suggest that the conditional correlation matrix is quite persistent. The log-likelihood decomposition results indicate a rather significant improvement in the overall fit compared to the BEKK, OGARCH and GOGARCH models, especially in comparison to OGARCH.

	DCC			
	CCC	Scalar	Diagonal	CP
<i>Variance parameters</i>				
Margin (XOM)		(0.084, 0.901)		
Margin (AA)		(0.048, 0.948)		
<i>Correlation parameters</i>				
CCC	0.480	–	–	–
α	–	0.015	–	–
β	–	0.977	–	–
α_{11}	–	–	0.006	0.005
α_{22}	–	–	0.037	0.054
β_{11}	–	–	0.993	–
β_{22}	–	–	0.960	–
λ	–	–	–	0.992
<i>LL decomposition</i>				
Margin (XOM)	-4,026	-4,026	-4,026	-4,026
Margin (AA)	-5,096	-5,096	-5,096	-5,096
Copula	293	306	307	307
Total LL	-8,829	-8,816	-8,815	-8,815

Table 3: Dataset: XOM and AA daily returns 1/2/2001-31/12/2009. Parameter estimates of the constant conditional correlations (CCC), and scalar, diagonal and common persistence (CP) parameterisations for the DCC model. Top panel: estimates of the variance targeting GARCH(1,1) models for the margins. Middle panel: estimates of the correlation parameters: α and β are the parameters of S-DCC, while $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, are those of D-DCC. For CP, λ (the common persistence parameter) and α_{ii} for each asset are reported. All parameters are statistically significant at the 5 percent level of significance. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

Figure 3 plots the conditional correlations from the diagonal models which provided the best fit in each model class. The D-DCC conditional correlation is the most persistent and lies within a tighter range. It appears to be generally lower than the conditional correlation from the D-BEKK and D-OGARCH model, with the exception of the year 2005 where D-OGARCH conditional correlation was noticeably lower. This observation is perhaps most evident during the latter part of the financial crisis, roughly starting 2009, with the difference in the implied correlation level being rather significant at times during this period.

We apply the 1-step predictive ability test outlined in Section 3.4 to the D-BEKK, D-OGARCH and D-DCC models which are the most flexible in each class. Comparing D-BEKK to D-OGARCH gives a t -statistic of 2.81 which is statistically significant at 1 percent, indicating that D-BEKK provides superior 1-step forecasts. Comparing D-DCC to

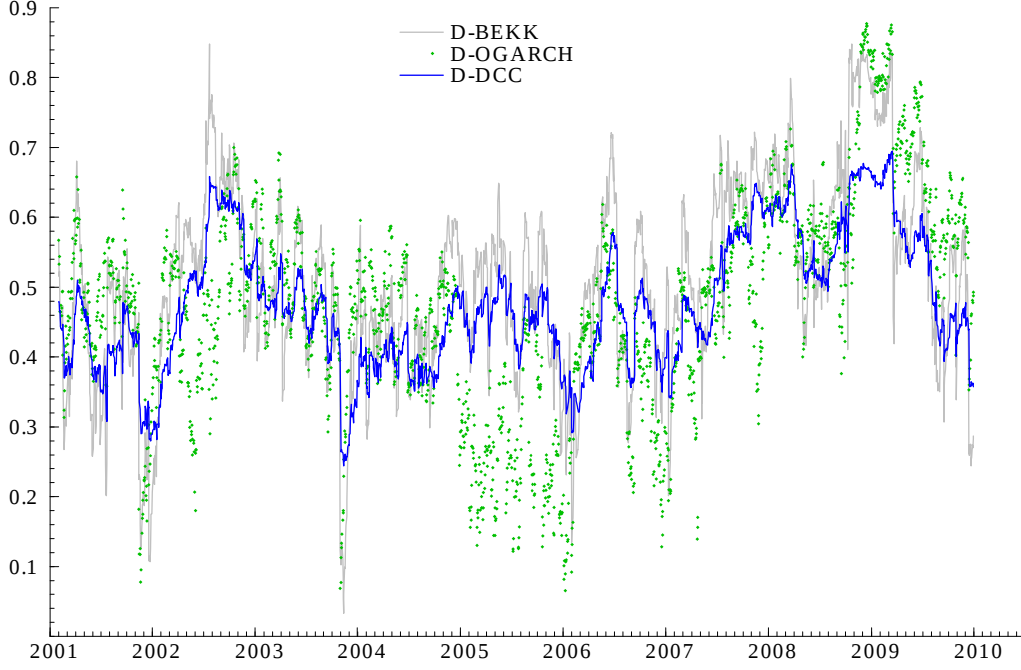


Figure 3: Conditional correlations from the diagonal BEKK, OGARCH and DCC models.

D-BEKK and D-OGARCH gives t -statistics equal to 2.24 and 3.74, respectively, indicating that D-DCC outperforms both models out of sample. These results are, of course, in line with the substantial in-sample gains shown by the DCC models.

4.4 Index-Stock Dynamics: SPY-XOM

The results for SPY-XOM are reported in Table 4. Moving from S-BEKK to D-BEKK leads to a modest improvement in fit for the first margin and the copula. This is also the case in OGARCH with gains only in the first margin. GOGARCH provides considerable gain compared to OGARCH, particularly in the copula fit, with a statistically significant estimate of the rotation angle at about -130 degrees. Both DCC and BEKK models improve significantly over OGARCH and GOGARCH, with DCC providing some gain over BEKK in both margins and the copula.

	BEKK			OGARCH			GOGARCH			DCC		
	S	D	CP	S	D	CP	S	D	CP	S	D	CP
Marginal parameters												
Margin (SPY)	-	-	-	-	-	-	-	-	-	-	(0.077, 0.918)	
Margin (XOM)	-	-	-	-	-	-	-	-	-	-	(0.084, 0.901)	
Dynamic parameters												
α	0.062	-	-	0.083	-	-	0.072	-	-	0.035	-	-
β	0.931	-	-	0.903	-	-	0.921	-	-	0.945	-	-
α_{11}	-	0.064	0.077	-	0.088	0.094	-	0.087	0.054	-	0.095	0.158
α_{22}	-	0.070	0.054	-	0.079	0.066	-	0.074	0.090	-	0.016	0.009
β_{11}	-	0.932	-	-	0.900	-	-	0.876	-	-	0.904	-
β_{22}	-	0.911	-	-	0.899	-	-	0.921	-	-	0.972	-
λ	-	-	0.992	-	-	0.986	-	-	0.991	-	-	0.977
δ	-	-	-	-	-	-	-129.996	-129.999	-129.998	-	-	-
LL decomposition												
Margin (SPY)	-3,316	-3,313	-3,313	-3,324	-3,318	-3,321	-3,314	-3,384	-3,330	-3,311	-3,311	-3,311
Margin (XOM)	-4,030	-4,031	-4,032	-4,026	-4,029	-4,029	-4,029	-4,032	-4,030	-4,026	-4,026	-4,026
Copula	609	616	611	506	506	509	574	664	597	620	621	623
Total LL	-6,737	-6,727	-6,734	-6,844	-6,840	-6,841	-6,769	-6,752	-6,764	-6,717	-6,716	-6,715

Table 4: Dataset: SPY and XOM daily returns 1/2/2001-31/12/2009. Parameter estimates of the scalar (S), diagonal (D) and common persistence (CP) models. Top panel: Marginal parameter estimates are of the variance targeting GARCH(1,1) models for the DCC margins. Dynamic parameters are estimates of the BEKK, OGARCH and GOGARCH models, and the correlation dynamics for DCC. λ is the common persistence parameter, while δ is the rotation angle in the bivariate GOGARCH model. All parameters are statistically significant at 5 percent. Bottom panel: Log-likelihood decomposition at the estimated parameter values.

In terms of predictive ability, the D-BEKK model provides superior 1-step forecasts compared to OGARCH models with a t -statistic of 4.48. The D-DCC also significantly improves over D-OGARCH with a t -statistic of 4.77; however, its improvement over D-BEKK is statistically insignificant with a t -statistic of 1.12. Again this mirrors the in-sample results of the three models.

Figure 4 shows the conditional volatilities, correlation and beta from D-OGARCH and D-DCC for SPY-XOM. The conditional variances from the two models seem quite similar, except for the SPY conditional volatility during 2005-2007 where the difference is mainly one of scale. The path of the conditional correlations is also somewhat similar although the D-OGARCH model attains more spikes. The interesting difference in this figure is the rather different profile for the conditional beta. From 2005 to mid 2007, the D-DCC model implies a conditional beta that is consistently larger and typically greater than 1, and it seems to have moderated gradually during the financial crisis.

4.5 Ten Dimensional Example

We now analyse all 10 stocks from the DJIA index. The first two eigenvectors, corresponding to the two largest eigenvalues of the unconditional covariance matrix of the returns, are reported in Table 5. The first eigenvector looks roughly like a market factor and the second is a market portfolio that is short (long) in financial stocks (BAC, JPM and AXP) and long (short) in the other stocks.¹² The two largest eigenvalues are 35.93 and 6.85, and they account for 73 percent of the total variation in the returns, where total variation is measured by the trace of $\overline{H}$.

Table 6 shows the estimated parameters for the scalar, diagonal and common persis-

¹²Note that when the eigenvalues are distinct, the normalised eigenvectors are unique up to sign.

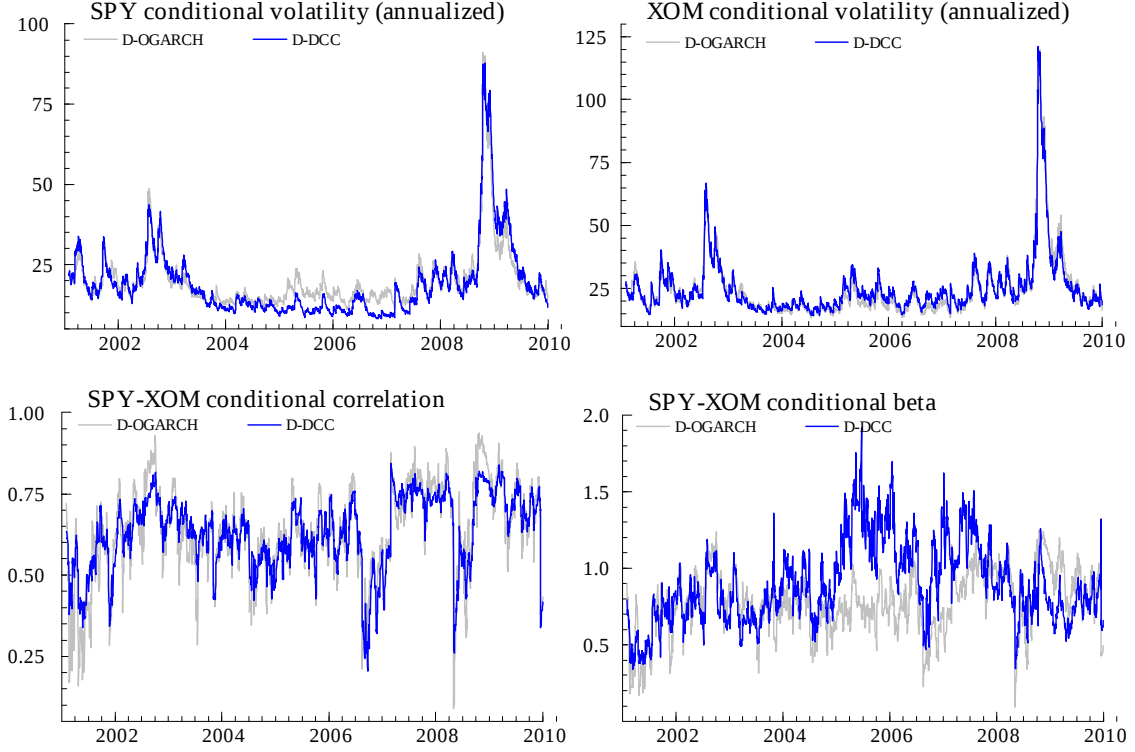


Figure 4: SPY-XOM conditional variances, correlation and beta from the diagonal OGARCH and DCC models.

	BAC	JPM	IBM	MSFT	XOM	AA	AXP	DD	GE	KO
Eigenvector 1	0.505	0.439	0.182	0.218	0.180	0.360	0.392	0.242	0.292	0.106
Eigenvector 2	-0.584	-0.288	0.177	0.259	0.255	0.582	-0.033	0.223	0.077	0.129

Table 5: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. The first two (normalised) eigenvectors correspond to the two largest eigenvalues of the unconditional covariance matrix of the returns.

tence models. The latter are an interesting alternative in moderately large dimensions since they have only $p + 1$ dynamic parameters compared to $2p$ parameters in the diagonal models. Moving from the scalar to the diagonal models seems to pay off with a considerable improvement in overall fit in D-BEKK, and less so for D-OGARCH. The BEKK models provide a significant overall gain in the log-likelihood over OGARCH all due to improving the copula fit. Note that the BEKK loses in the margins to OGARCH as the BEKK parameters provide a fit to both the variance and covariance elements of G_t .

Of course, the DCC models provide the best fit since the margins are freely estimated.

The overall gain compared to BEKK and OGARCH is quite impressive, and the DCC gains are uniform across all margins and the copula. Unlike BEKK and OGARCH cases, moving from S-DCC to D-DCC does not improve the copula fit massively. In this moderately large dimension, the favourable performance of the CP model is evident, particularly in the BEKK and OGARCH cases. In both cases, the diagonal specifications significantly improve the overall fit (mostly due to the copula contribution) and when fitting the CP model the deterioration in fit is rather slight. To a lesser extent, this is also the case in the DCC models.

Given that both the scalar and CP specifications are nested in the diagonal model, we can use a likelihood ratio (LR) test. The scalar model imposes $2p$ restrictions on the diagonal model, and according to the LR test, the reduction in fit is statistically significant at 5 percent in all three cases. The CP model imposes $p(p+1)/2$ restrictions on the diagonal model and according to the LR test, the loss in fit when moving from D-BEKK to D-BEKK-CP is statistically significant at 5 percent, while this is not the case in the OGARCH and DCC models.

This is an interesting result since the number of dynamic parameters in the CP model is $p+1$ compared to $2p$ dynamic parameters in the diagonal model. This could be due to the differences in the heterogeneity in the persistence and smoothness levels among the parameters of the diagonal models. For instance, in D-BEKK the heterogeneity in the parameters is given by $\sigma_\alpha = 0.014$ and $\sigma_\beta = 0.023$, while the corresponding measures in D-DCC are $\sigma_\alpha = 0.004$ and $\sigma_\beta = 0.020$. Since both are lower, especially $\sigma_\alpha = 0.004$, it is expected that imposing a common persistence level in the case of DCC may not substantially affect the empirical fit, and this is what the LR ratio test result suggests.

The picture from the overall log-likelihood analysis is confirmed by the predictive ability

	BEKK			OGARCH			DCC		
	S	D	CP	S	D	CP	S	D	CP
<i>Dynamic parameters</i>									
α	0.020	–	–	0.045	–	–	0.007	–	–
β	0.978	–	–	0.952	–	–	0.980	–	–
$\min \alpha_{ii}$	–	0.009	0.010	–	0.027	0.025	–	0.004	0.002
$\max \alpha_{ii}$	–	0.054	0.054	–	0.097	0.095	–	0.017	0.015
$\min \beta_{ii}$	–	0.905	–	–	0.869	–	–	0.932	–
$\max \beta_{ii}$	–	0.989	–	–	0.967	–	–	0.991	–
λ	–	–	0.998	–	–	0.996	–	–	0.987
<i>LL decomposition</i>									
Margin (BAC)	-4,496	-4,355	-4,373	-4,416	-4,351	-4,361	-4,350	-4,350	-4,350
Margin (JPM)	-4,769	-4,719	-4,734	-4,706	-4,695	-4,700	-4,671	-4,671	-4,671
Margin (IBM)	-4,058	-4,085	-4,092	-4,025	-4,025	-4,023	-4,011	-4,011	-4,011
Margin (MSFT)	-4,449	-4,488	-4,482	-4,438	-4,431	-4,433	-4,424	-4,424	-4,424
Margin (XOM)	-4,090	-4,067	-4,084	-4,040	-4,032	-4,035	-4,026	-4,026	-4,026
Margin (AA)	-5,115	-5,130	-5,132	-5,097	-5,097	-5,096	-5,096	-5,096	-5,096
Margin (AXP)	-4,665	-4,648	-4,652	-4,620	-4,705	-4,654	-4,599	-4,599	-4,599
Margin (DD)	-4,249	-4,310	-4,299	-4,231	-4,247	-4,232	-4,228	-4,228	-4,228
Margin (GE)	-4,291	-4,299	-4,300	-4,263	-4,327	-4,314	-4,257	-4,257	-4,257
Margin (KO)	-3,556	-3,558	-3,562	-3,528	-3,542	-3,542	-3,520	-3,520	-3,520
Copula	4,640	4,860	4,807	3,888	4,040	3,963	4,919	4,946	4,939
Total LL	-39,098	-38,798	-38,904	-39,475	-39,413	-39,426	-38,263	-38,236	-38,244

Table 6: Dataset: 10 DJIA stocks daily returns 1/2/2001-31/12/2009. Parameter estimates of the scalar (S), diagonal (D), and common persistence (CP) models. Top panel: estimates of the dynamic parameters. α and β are the parameters of the scalar models, while $(\alpha_{ii}, \beta_{ii})$, $i = 1, 2$, are those of the diagonal models. For CP, only λ (the common persistence parameter) and α_{ii} are reported. All parameters are statistically significant at the 5 percent level of significance. Lower panel: Log-likelihood decomposition at the estimated parameter values.

tests for the diagonal models. Compared to the D-OGARCH specifications, D-BEKK produces superior 1-step forecasts with a statistically significant t -statistic equal to 3.49. The D-DCC model outperforms both D-BEKK and D-OGARCH with statistically significant t -statistics equal to 2.66 and 7.09, respectively.

5 Conclusion

This paper advocates a rotation technique for raw returns which leads to easy-to-fit multivariate volatility models via covariance targeting. We discuss the similarities and differences between our approach and the recent orthogonal GARCH models. In particular, while the early contributions to the OGARCH literature assumed, for simplicity, that

the estimated orthogonal components are also conditionally uncorrelated, we observe that this is only an approximation since the rotated returns will inherit the conditionally heteroskedastic properties of the unrotated returns. Therefore, we advocate using the popular BEKK and DCC models to study the dynamics of the conditional covariance matrix of the rotated returns. We also discuss a distinct extension of the diagonal BEKK and DCC models, and draw parallels to the OGARCH model of Alexander (2001) and the GOGARCH model of van der Weide (2002).

We show that fitting a diagonal BEKK model to the rotated returns implies a full BEKK specification for the unrotated returns further highlighting the modelling flexibility our approach offers. Estimation and inference is also computationally attractive, thanks to the convenient form of covariance targeting with a long-run identity matrix. Using two-step estimation, we end up estimating only $O(p)$ parameters with numerical optimisation which offers advantages in moderately large dimensions.

Indeed using our approach leads to notable 1-step prediction gains compared to OGARCH and GOGARCH. Capturing the dynamics of the covariances of the rotated returns does improve the prediction of the conditional correlation. Given their flexibility, the DCC suite of models performs best in the 10 dimensional example we study. Interestingly, our newly proposed common persistence model performs quite favourably in comparison to the diagonal model while being more tightly parameterised.

References

- Aielli, G. P. (2006). Consistent estimation of large scale dynamic conditional correlations. Unpublished paper: Department of Statistics, University of Florence.
- Alexander, C. (2001). Orthogonal GARCH. In *Mastering Risk*, Volume 2, pp. 21–38. Prentice Hall.
- Alexander, C. O. and A. M. Chibumba (1997). Multivariate orthogonal factor GARCH. Mimeo: University of Sussex, UK.

- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25, 177–190.
- Bauwens, L., S. Laurent, and J. V. K. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–109.
- Billio, M. and M. Caporin (2009). A generalized dynamic conditional correlation model for portfolio risk evaluation. *Mathematics and Computers in Simulation* 79, 2566–2578.
- Billio, M., M. Caporin, and M. Gobbo (2006). Flexible dynamic conditional correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters* 2, 123–130.
- Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics* 72, 498–505.
- Boswijk, H. P. and R. van der Weide (2011). Method of moments estimation of GO-GARCH models. *Journal of Econometrics*. Forthcoming.
- Brownlees, C. T. (2010). On the relation between firm characteristics and volatility dynamics with an application to the 2007-2009 financial crisis. Working paper: Stern School of Business, New York University.
- Brownlees, C. T. and R. F. Engle (2010). Volatility, correlation and tails for systemic risk measurement. Working paper: Stern School of Business, New York University.
- Cappiello, L., R. F. Engle, and K. Sheppard (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics* 4, 537–572.
- Cox, D. R. (1961). Tests of separate families of hypotheses. *Proceedings of the Berkeley Symposium* 4, 105–123.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Engle, R. F. (2002). Dynamic conditional correlation - a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339–350.
- Engle, R. F. (2009a). *Anticipating Correlations*. Princeton University Press.
- Engle, R. F. (2009b). High dimensional dynamic correlations. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: Papers in Honour of David F Hendry*, pp. 122–148. Oxford University Press.
- 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., N. Shephard, and K. K. Sheppard (2008). Fitting vast dimensional time-varying covariance models. Unpublished paper: Oxford-Man Institute, University of Oxford.
- Fan, J., M. Wang, and Q. Yao (2008). Modelling multivariate volatilities via conditionally uncorrelated components. *Journal of the Royal Statistical Society, Series B* 70, 679–702.

- Fan, Y., S. Pastorello, and E. Renault (2007). Maximization by parts in extremum estimation. Unpublished paper.
- Francq, C. and J. M. Zakoian (2010). *GARCH Models*. Chichester: Wiley.
- Giacomini, R. and H. White (2006). Tests of conditional predictive ability. *Econometrica* 74, 1545–1578.
- Hafner, C. M. and P. H. Franses (2009). A generalized dynamic conditional correlation model: simulation and application to many assets. *Econometric Reviews* 28, 612–631.
- Hansen, P. R., A. Lunde, and V. Voev (2010). Realized beta GARCH: a multivariate GARCH model with realized measures of volatility and covolatility. Working paper: CREATES, Aarhus University.
- Lanne, M. and P. Saikkonen (2007). A multivariate generalized orthogonal factor GARCH model. *Journal of Business and Economic Statistics* 25, 61–75.
- Lütkepohl, H. (1996). *Handbook of Matrices*. Chichester: Wiley.
- Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2111–2245. North-Holland.
- Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Noureldin, D., N. Shephard, and K. Sheppard (2011). Multivariate high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics*. Forthcoming.
- Pagan, A. (1986). Two stage and related estimators and their applications. *Review of Economic Studies* 53, 517–538.
- Pakel, C., N. Shephard, and K. Sheppard (2011). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica* 21, 307–329.
- Rangel, J. G. and R. F. Engle (2011). The factor-spline-GARCH model for high- and low-frequency correlations. Working paper.
- Ross, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.
- Silvennoinen, A. and T. Teräsvirta (2009). Multivariate GARCH models. In T. G. Andersen, R. A. Davis, J. P. Kreiss, and T. Mikosch (Eds.), *Handbook of Financial Time Series*, pp. 201–229. Springer-Verlag.
- van der Weide, R. (2002). GO-GARCH: a multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17, 549–564.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.