

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
FACULTÉ DES SCIENCES ÉCONOMIQUES ET DE GESTION
DÉPARTEMENT DES SCIENCES ÉCONOMIQUES
CENTER FOR OPERATIONS RESEARCH AND ECONOMETRICS



Component dynamic models for realized covariance matrices

Manuela Braione

*Thèse présentée en vue de l'obtention du grade de
docteur en sciences économiques et de gestion*

Composition of the jury:

Supervisors:

Prof. Luc Bauwens (Université catholique de Louvain)
Prof. Christian Hafner (Université catholique de Louvain)

President of the jury:

Prof. Sébastien Van Bellegem (Université catholique de Louvain)

Other jury members:

Prof. Giuseppe Storti (University of Salerno)
Prof. Lyudmila Grigoryeva (University of Konstanz)
Prof. Kris Boudt (Department of Business, Vrije Universiteit Brussel, Belgium)

Louvain-la-Neuve, October 2016

In Loving Memory Of My Grandma

Per Nonna

Acknowledgments

This dissertation represents one of the biggest achievements of my life and for it, I am deeply grateful to many people. First of all I have to thank my supervisors, *Luc Bauwens* and *Christian Hafner* for having accepted me as their student in the first place. In particular I want to say thanks to Professor *Bauwens* for his patience, expertise and priceless guidance that made me grow and gave me strength during the whole Ph.D. Then, my unlimited gratitude inevitably goes to *Giuseppe Storti*, for being a great professor, a coauthor and mostly a mentor. He has been constantly supporting me and our work during the past few years, dedicating me time and teaching me almost all that I know. He was there at my master defense and he is here now. Sometimes I still wonder what have I done to deserve such a luck.

I would also like to thank all other members of my committee, *Sébastien Van Belleghem*, *Kris Boudt* and *Lyudmila Grigoryeva* for the time spent reading my thesis and providing constructive comments to improve it. A special thanks to *Lyudmila* for coming twice to Belgium and for being such a nice person. I honestly think her students should feel lucky.

I always felt very honored and proud to be part of CORE. It has been my second home for almost four years and it gave me the chance to extend my horizons from both a professional and personal point of view. I met incredible people there and most of them became really close friends. Even though it is impossible to acknowledge everyone, I need to pay a special tribute to some of them: to *Andrea* and *Sinem*, for being my lovely housemates; to *Nicolas* and *Pana*, for being my first and best friends in Louvain-la-Neuve; to *Ignacio*, for being my favorite Spanish teacher; to *Andras*, for being always there when needed; to *Soto* and *Keiti*, who became part of the Italian family; to *Gilles*, for all the laughs and nice discussions; to *Cyrille*, for patiently enduring my complaints; to *Elisa*, for her support and contagious joie de vivre; to *Valeria*, for always being the wise one; to *Claudio* and *Sabina*, for being such good friends, and to all the others I had the pleasure to meet and appreciate the company. I hope for you all the very best and that you achieve all your desired goals.

Apart from Ph.Ds, I want to say thanks to *Catherine*, for all the invaluable help and for the passion and patience she puts in her job everyday, and to *Leonardo*, with whom it was a pleasure and a privilege, to work with. He encouraged me to apply for the internship at the National Bank and he introduced me to valuable people. I am sure that with his endless energy and persistence he will make great things for CORE.

Last but not least, I cannot refrain from mentioning my lifetime friends, *Alessandra*

and *Antonio*, who helped me out when I needed the most and who I love as brothers. Finally, useless to say that without my family I would not even be here, so my biggest thanks goes to them, for the immense support and care. Your love is the only fixed point of my life.

Contents

1	General Introduction	1
1.1	Some preliminary concepts	6
1.1.1	What is volatility?	6
1.1.2	Quadratic variation and integrated variance	7
1.1.3	Quadratic covariation and integrated covariance	11
2	A class of multiplicative models for realized covariance matrices	17
2.1	Introduction	17
2.2	The multiplicative realized model class	19
2.2.1	Multiplicative framework	19
2.2.2	Parameterization of the short-term component	20
2.2.3	Feasible parameterizations	23
2.2.4	A dynamic attenuation approach	24
2.3	Multistep estimation approach	25
2.3.1	Covariance targeting	26
2.3.2	QML estimation	27
2.3.3	Third stage composite likelihood estimation	30
2.4	Simulation study	31
2.5	Empirical analysis	36
2.5.1	Bivariate dataset AA-XOM	37
2.5.2	High dimensional model	40
2.6	Conclusions	46
	Appendix 2.A Technical Proofs	48
	Appendix 2.B Tables	50
3	Forecasting comparison of long run component dynamic models for realized covariance matrices	53
3.1	Introduction	53
3.2	TVLR component models	56
3.2.1	Long-run additive component models for correlations	56
3.2.2	Long term multiplicative component models	60
3.3	Estimation	63
3.4	Evidence from real data	65
3.4.1	Data and forecasting scheme	65

3.4.2	Full sample results	67
3.4.3	Forecasting results	69
3.5	Conclusions	77
Appendix 3.A	Tables	79
4	Feasible inference procedures for forecasting high-dimensional realized covariance matrices	83
4.1	Introduction	83
4.2	Curse of dimensionality issue	87
4.3	An Iterative Moment based Profiling (IMP) algorithm	88
4.4	Simulation study	90
4.5	Multi-step Forecasting	95
4.5.1	A conditional bootstrap (CB) procedure	96
4.5.2	Finite sample properties	98
4.6	Empirical Applications	99
4.6.1	Small sample accuracy comparison	99
4.6.2	Forecasting performance	102
4.7	Conclusions	109
Appendix 4.A	Tables	111
Appendix 4.B	Bootstrap procedure for the cReDCC model	113
	Conclusions	115
	Appendix A The Wishart distribution	119
	Appendix B Summary of TVLR component models	121
	Appendix C Overview of competing models	123

List of Figures

1.1	Fitted conditional variance (in blue) and corresponding long term component (in red) of JP Morgan (JPM) (1/2/2001 - 31/12/2009).	3
2.1	Predicted conditional correlations of XOM-AA from diagonal CAW, RecDCC and RDCC models (27/03/2007-31/12/2009).	40
2.2	Daily returns (upper panel) and predicted conditional volatilities (lower panel) of AXP stock for Dataset A (left column) and Dataset B (right column) over the two sample periods analyzed.	41
2.3	Daily portfolio concentration and turnover for Dataset A (left side) and Dataset B (right side).	46
3.1	Beta polynomial weighting function	58
3.2	Realized series for JPM-XOM	66
3.3	Implemented rolling window setting	67
3.4	Predicted series with corresponding long term component	70
3.5	Predicted conditional correlations JPM-XOM	73
4.1	IMP Monte Carlo standard deviations	93
4.2	Bootstrapped prediction intervals	99
4.3	Maximized likelihood values	102
4.4	Estimated conditional correlation of MCD-WFC (upper panel) and of AXP-BAC (lower panel).	105
4.5	Equally weighted portfolio daily realized variance	107

List of Tables

2.1	Simulation design for the scalar Re-cDCC process	31
2.2	Simulation results for the 3-step QMLE of the GARCH and cDCC parameters	34
2.3	Simulation results for the targeted scale matrix M	34
2.4	Monte Carlo comparison results: GARCH and cDCC parameters	35
2.5	Monte Carlo comparison results: M matrix	35
2.6	AA-XOM descriptive statistics	38
2.7	Full sample estimates for the the Alcoa and Exxon Mobil dataset	39
2.8	Full sample parameter estimates	42
2.9	MCS results	44
2.10	GMV portfolio results	45
2.11	Implemented Loss functions	50
2.12	Data	51
2.13	Monte Carlo study results: comparison of estimators in low dimensions	52
3.1	List of time-varying additive component models	61
3.2	List of time-varying multiplicative component models	63
3.3	Information criteria of the models	68
3.4	Model confidence sets at 90% and 75% levels.	71
3.5	Out-of-sample hedging performance	74
3.6	Likelihood Ratio Test and Dynamic Quantile Test Results	76
3.7	Descriptive statistics of realized variances	79
3.8	QML parameter estimates	80
3.9	Robust loss functions definition	81
3.10	90% MCS p -values	81
3.11	75% MCS p -values	82
4.1	Number of parameters of MMReDCC models	88
4.2	Simulation setting	91
4.3	Simulation exercise I: summary statistics	91
4.4	Simulation exercise II: summary statistics	92
4.5	Bias results for the scale MIDAS intercept matrix	94
4.6	Properties of the estimator	95
4.7	Prediction intervals	100
4.8	Application I. In-sample comparison	103

4.9	Application II. Full sample estimates and implemented loss functions .	105
4.10	Multi-step-ahead forecast evaluation	108
4.11	GW test results across series	109
4.12	Descriptive statistics of daily realized variances used in Application I .	111
4.13	Descriptive statistics of daily realized variances used in Application II .	112
B.1	Complete list of TVLR dynamic component models	122

List of Symbols and Abbreviations

Vectors will be denoted with boldfaces (e.g. $\boldsymbol{\alpha}$) and matrices with capital letters (e.g. A). Sets and spaces will be denoted with calligraphic letters (e.g. $\mathfrak{S}$).

Symbols

argmax	Arguments of the maxima
$\mathfrak{S}_t$	Available information up to time t
$\mathbb{E}(x)$	Unconditional expected value of x
$\mathbb{E}_{t-1}(x)$	Expected value of x conditional on the available information at time $t - 1$, i.e. $\mathbb{E}(x \mathfrak{S}_{t-1})$
I_n	Identity matrix of order n
ℓ	Log-likelihood function
$\log(x)$	Natural logarithm of x
tr	Trace operator summing the elements along the main diagonal of a matrix
vech	Operator stacking the lower triangular part of a matrix into a column vector
vec	Operator stacking the columns of a matrix into a column vector
$ A $	Determinant of matrix A
$\boldsymbol{\iota}$	Column vector of elements equal to one
plim	Probability limit
$\xrightarrow{P}$	Convergence in probability
$\xrightarrow{d}$	Convergence in distribution

∂	Partial derivative
$\otimes$	Kronecker product
$\odot$	Hadamard or element by element product
Σ	Summation
$\propto$	Proportionality
■	End of proof

Abbreviations

PDS	Positive definite and symmetric (matrix)
PSD	Positive semidefinite (matrix)
QML	Quasi Maximum Likelihood
1QML	One-step Quasi Maximum Likelihood
2QML	Two-step Quasi Maximum Likelihood
3QML	Three-step Quasi Maximum Likelihood
IMP	Iterative Moment-based Profiling algorithm
BEKK	Baba Engle Kraft Kroner
CAW	Conditional Autoregressive Wishart
CCC	Constant Conditional Correlation
DCC	Dynamic Conditional Correlation
DECO	Dynamic Equicorrelation
EWMA	Equally Weighted Moving Average
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
MGARCH	Multivariate Generalized Autoregressive Conditional Heteroskedasticity
MIDAS	Mixed Data Sampling
MCS	Model Confidense Set
RV	Realized Variance
RC	Realized Covariance

Chapter 1

General Introduction

It is widely known that financial time series are characterized by complex dynamics. In particular, the recent financial and Euro-zone debt crisis has shed light on the key role played by the co-movement of financial instruments, not only within a market but also across different markets. Accurately modeling the dynamics of the commonalities between elements of a portfolio or a system is crucial for good decision making, for example with respect to asset allocations and hedging of risky investments. To this aim, a careful analysis, specification, estimation and forecasting of financial asset returns dynamics, and the construction and evaluation of financial portfolios, are essential in the toolkit of any financial planner and analyst.

Multivariate GARCH (MGARCH) models have been, and overall still are, the most widely used time series models for measuring aggregate risk as a function of the covariance matrix of the elements. Numerous parametric models have been proposed for different purposes. Leading examples are the BEKK of [Engle & Kroner \(1995\)](#), the Constant Conditional Correlation (CCC) of [Bollerslev \(1986\)](#), the Dynamic Conditional Correlation (DCC) of [Engle \(2002a\)](#) and the Dynamic Equicorrelation (DECO) of [Engle & Kelly \(2012\)](#), just to name a few. These models allow for co-volatilities and/or correlations to change over time as a function of observed daily lagged returns and of their own past realizations. The interest of the academic world in these models is explained by the fact that they are simple enough to be used in practice, but also rich in theoretical problems, some of which still unsolved.

More recently, the rapidly increasing availability of high-quality transaction data across many financial assets, allowed the construction of estimates of ex-post realized daily volatility by summing squares and cross-products of intraday high-frequency returns. This exhaustive set of data has stimulated the development of a lively field of econometric analysis with the proposition of models directly fitted to time series of realized measures. The construction of accurate *realized estimators* is a non-trivial task and constitutes the core of an influential research brunch within the field, which finds in [Andersen et al. \(2003\)](#), [Barndorff-Nielsen & Shephard \(2004\)](#), [Barndorff-Nielsen et al. \(2011\)](#), [Christensen et al. \(2010\)](#) and [Zhang \(2011\)](#) its most notable references.

As advocated by [Andersen et al. \(2003\)](#), time-series co-volatility models based on such data produce significant improvements relative to those that rely on daily data alone, because intra-daily returns not only turn out to be highly informative, but also useful for forecasting conditional co-volatilities at longer horizons. Pioneering works in this respect can be found in [Gouriéroux et al. \(2009\)](#), [Bonato et al. \(2012\)](#), [Noureldin et al. \(2012\)](#), [Golosnoy et al. \(2012\)](#), [Chiriac & Voev \(2011\)](#) and [Jin & Maheu \(2013\)](#), among many others.

Nevertheless, when multiple assets are involved, two major issues have to be faced. First and foremost, models have to be specified in order to guarantee the positive-definiteness of the conditional covariance matrix at each point in time. A natural choice in this respect is to adopt a distribution belonging to the matrix-valued Wishart family which automatically generates random positive definite symmetric (PDS) matrices without imposing additional constraints. Central and non central Wishart densities represent a milestone in this literature due to their ease of application, but it is worth mentioning that other PDS matrix distributions have been lately proven to be equally suited for modeling realized covariance matrices (see e.g. [Janus et al. \(2014\)](#) who use the matrix-Fisher distribution to account for the fat-tailedness of the realized series).

Second, since the dimension of the object to be modeled is proportional to the square of the number of assets, the number of model parameters easily becomes unmanageable even for a handful of assets. It is widely known that the so called *curse-of-dimensionality* severely constrains the practical application of many existing models and hence, beforehand, it makes opportune to be concerned with the flexibility of the proposed parameterization and the implied computational cost. A remarkable example along this line is the Conditionally Autoregressive Wishart (CAW) model of [Golosnoy et al. \(2012\)](#), where the realized covariance matrix is assumed to follow a central Wishart distribution with a time-varying conditional expectation that is inspired by the BEKK model of the MGARCH literature. Despite many interesting strength points of the model, one evident drawback is represented by its heavily parametrized structure that prevents implementation in situations of practical relevance, i.e. where the cross-sectional dimension is of several tens of assets.

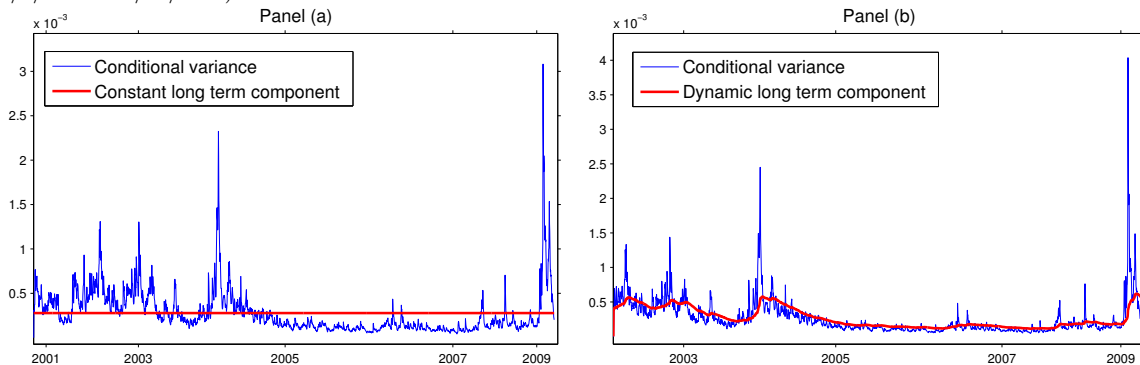
In view of such challenging background, the present work seeks to improve matters by proposing feasible realized covariance models that can be useful in financial applications.

To this aim, Chapter 2 introduces a class of models inspired by considerations of feasibility, simplicity and speed of implementation in high-dimensional environments. We call the main representative of this class the Realized consistent DCC (Re-cDCC) model, whose main ingredients are a consistent DCC-type parameterization inspired by the MGARCH model of [Aielli \(2013\)](#) coupled with a multiplicative structure that allows to extract two different co-volatility components, namely the constant long-term covariance matrix and the short term dynamics of variances and correlations. The component structure leads to several important advantages, such as a stepwise quasi-maximum likelihood estimation approach that considerably simplifies the optimization

procedure. This allows us to perform estimation for relatively large realized covariance matrices, up to fifty assets in the empirical illustration and one hundred in Monte Carlo simulations. Our ability to model matrices of such dimensions contrasts with existing competitors: [Gouriéroux et al. \(2009\)](#) work with three assets, [Bonato et al. \(2009\)](#) with four, [Golosnoy et al. \(2012\)](#) and [Jin & Maheu \(2013\)](#) with five, [Chiriac & Voev \(2011\)](#) with six, [Noureldin et al. \(2012\)](#) with ten and [Bonato et al. \(2012\)](#) work with twelve. Simplicity and ease of implementation make this model readily applicable in financial applications aimed at accurately predicting asset conditional variances, covariances and correlations as well as to manage the portfolio variability over a pre-specified horizon. These are also the key reasons for which the Re-cDCC represents a natural benchmark for the forecast performance of more sophisticated co-volatility models.

A feature of the Re-cDCC, common to many of the aforementioned models, is that it relies on the assumption of constant long term volatility. Against this conceivably limiting belief, a relatively recent branch of the literature supports the idea that the levels to which variances and correlations mean revert in the long run evolve as a function of the underlying economic conditions. To get a flavor of the intuition behind this statement, Figure 1.1 displays the JP Morgan (JPM) annualized conditional volatility spanning from February 1, 2001 to December 31, 2009, given by two different models. The time frame is large enough to observe how volatility evolves over time, either in a slow and progressive way during calmer periods or more quickly during turbulent periods. The two models differs in the way these movements are accounted for in the estimation of the long term volatility component: in Panel (a) they are ignored by assuming mean reversion to a constant level, while in Panel (b) the long term component is allowed to be dynamic, namely to slowly adapt to the changing economic environment.

Figure 1.1: Fitted conditional variance (in blue) and corresponding long term component (in red) of JP Morgan (JPM) (1/2/2001 - 31/12/2009).



As firstly pointed out by [Engle & Lee \(1999\)](#), this choice can have important forecasting implications, as models reverting to constant levels will produce the same long term forecast at any point in time, irrespective of the changing state variables that govern market volatility. From an economical point of view, it has been found that volatility over longer time intervals has stronger influence on those over shorter time intervals than conversely ([Zumbach & Lynch \(2001\)](#), [Müller et al. \(1997\)](#)). Intuitively,

this can be explained by noticing that for short term traders the level of long term volatility matters because it determines the expected future size of trends and risk. Then, on the one hand, short term traders react to changes in long term volatility by revising their trading behavior and so causing short term volatility. On the other hand, the level of short term volatility does not affect the trading strategies of long term investors, generating the so called “cascade effect” that has inspired, for example, the Heterogeneous Autoregressive (HAR) model of [Corsi \(2009\)](#).

For these reasons, and in light of the recent financial events, attention towards models that include a time-varying long run component has sensibly increased.

Since [Engle & Lee \(1999\)](#) introduced a GARCH model with additive long run and short run dynamic components, several others have proposed related component models for volatility: [Engle & Rangel \(2008\)](#) specify a semi-parametric model that allows for time-variation in the unconditional level of stock market volatility, while [Engle et al. \(2008\)](#) use a parametric specification involving the Mixed Data Sampling (MIDAS) scheme to extract a slowly moving secular component around which daily volatility moves. More precisely, the short run component is a GARCH specification based on daily squared returns, while the long run component is driven by realized volatilities computed over a monthly, quarterly or biannual basis. The idea has been later extended to dynamic correlations in the DCC-MIDAS of [Colacito et al. \(2011\)](#) and in the multiplicative-DCC of [Bauwens et al. \(2013\)](#).

The first time-varying long run component model directly built for realized covariance matrices is the MIDAS-CAW of [Golosnoy et al. \(2012\)](#), designed to capture long run fluctuations in the levels of covariances by means of a parametric MIDAS filter of past realized covariance realizations. From the empirical results discussed in that paper, it is clear that this class of models has several strength points. First, it is shown that long-term dynamic models for realized covariance matrices can be applied for portfolio optimization or asset allocation strategies, with clear improvements over simpler benchmarks. Also, they are found to provide more accurate multi-steps ahead predictions at medium-term horizon than models relying on daily data alone.

Unfortunately, the MIDAS structure coupled with the full BEKK recursion renders the MIDAS-CAW estimation difficult already in very small dimensional frameworks, thus suggesting that alternative specifications could be proposed to drastically lessen computational burden.

In this spirit, Chapter 3 progresses over Chapter 2 by extending the basic Re-cDCC framework to account for the stylized facts listed above yet preserving computational tractability in moderately large dimensional frameworks. Specifically, we build on the previous contribution by decomposing the conditional expectation of the realized covariance matrix into a slowly evolving process and a mean-reverting short run one. Different model structures featuring long and short term dynamics are proposed and the impact of parametric as well as nonparametric secular components are examined. The one-step ahead predictive accuracy of the models is evaluated in several economic and financial applications, such as portfolio allocation and Value at Risk (VaR) prediction. Given

the complexity of the proposed models, we do not push the analysis to cross-sectional systems larger than ten assets. As in [Golosnoy et al. \(2012\)](#), empirical results support the evidence that including a dynamic long term component is beneficial for enhancing predictive performance already at short-term horizons. Furthermore, our findings suggest that, among those investigated, the most powerful predictive model is the Multiplicative MIDAS Realized DCC (MMReDCC), whose structure obtains by combining a DCC-type recursion for the short term component with the simplest realized MIDAS filter for the secular trend.

In light of these encouraging results, we devote the last chapter of the thesis to extend the range of applicability of the MMReDCC model along two directions: estimation for high-dimensional systems and multi-step forecasting.

The first part of the chapter deals with the application of the model to large dimensional portfolios such as those encountered in Chapter 2. Differently from the Re-cDCC case, where parameter proliferation could be avoided by targeting the constant long term covariance matrix with the unconditional covariance, the MMReDCC presents an additional challenge given by the latent time-varying trend component which cannot be directly linked to any observed quantity. In order to surpass this issue, we implement a feasible estimation procedure, the *Iterative Moment-based Profiling* (IMP) algorithm, which profiles out the nuisance parameters in the MIDAS filter and iteratively maximizes the likelihood in terms of the other parameters of interest. The IMP drastically reduces the dimension of the optimization problem to be solved and allows to fit the model to realized covariance matrices of the order of fifty assets.

For its properties, it can be considered a *switching algorithm* in the sense discussed by [Boswijk \(1995\)](#) and [Cubadda et al. \(2015\)](#) since the maximization of the overall likelihood is obtained by switching between optimizations over different blocks of parameters, but with an additional targeting step. A comprehensive simulation study assesses the finite sample properties of the estimator, showing that it delivers unbiased estimates under alternative sets of starting values and initial conditions.

Finally, the last part of Chapter 4 is dedicated to the proposition of a resampling based procedure for the generation of multi-steps-ahead forecasts from the model, where the choice of a distribution-free approach based on residual bootstrap makes it robust to possible distribution misspecification. In an application to forty-six assets, the model estimated by the IMP method is used to generate forecasts of the realized covariance matrix up to twenty days ahead, and compared to existing benchmarks not accounting for short and long term co-volatility dynamics. It emerges that over calm periods, simpler model specifications tend to be preferred especially at the shortest horizons, while during more volatile periods, such as the one including the 2007-2008 financial crisis, accounting for time-varying long term dynamics in the conditional covariance process generates convincingly superior forecasts. The latter are particularly significant at the longest horizons.

Overall, modeling high-frequency financial time-series is far from being obvious. The research pursued in this thesis tries to contribute to the existing literature but

covers only a limited portion of the several unexplored topics that this field is offering. Clearly, extensions are feasible along several directions. These are briefly mentioned in the general conclusions, which offer some final comments and suggestions for possible future research venues.

1.1 Some preliminary concepts

Despite not being the research topic of interest, that of *realized measure* is a key concept of the present thesis and deserves some attention. Thus, before delving into the core of our work, we find opportune, for the reader's convenience, to report some useful notions that constitute the theoretical framework we will be working with.

At the outset, we briefly review the concept of volatility, its main features and the common issues that practitioners face in order to measure it. Consequently, we present the most commonly used ex-post volatility estimator in the univariate setting, the so called Realized Variance (RV) (Andersen et al. 2003). Then, we extend the concept to the case when there are multiple assets and illustrate the main features of the two realized covariance estimators that will be used in the remainder of this thesis, namely the Realized Covariance (RC) (Barndorff-Nielsen & Shephard 2004) and the CholCov estimator (Boudt et al. 2014).

Both in the univariate and in the multivariate case, we will introduce the basic, continuous-time, theoretical framework and the underlying assumptions on which our analysis relies.

1.1.1 What is volatility?

A price series or an economic indicator that is continuously subject to sudden and remarkable changes is usually said to be volatile. Although volatility is, in principle, a simple and intuitive concept, we should be aware that its statistical treatment is not simple and problem-free. This contributes to create substantial difficulties in many areas of financial modeling.

Unlike many other market parameters which can be directly observed, volatility is indeed a latent variable, and then it has to be estimated by referring to some observable proxy. This is not straightforward, as we cannot say that it is necessarily stochastic or that it conforms to any mathematical model; all we know is that the evolution of volatility is uncertain.

In financial time series analysis, we refer to volatility as a risk measure that specifies the level of uncertainty associated to a given financial asset or to a given investment portfolio. It is deeply related to the concept of variability in financial returns series and it has been proven to be characterized by some stylized features:

- *volatility clustering*: meaning that it tends to be steadily high in some periods and

steadily low in some others, or as first noticed by Mandelbrot (1963), that “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes”. A quantitative manifestation of this fact is that, while returns themselves are uncorrelated, absolute returns or their squares display a positive, significant and slowly decaying autocorrelation function;

- *long memory behavior*: changes in volatility typically have a very long-lasting impact on its subsequent evolution, indicating a high level of persistence;
- *asymmetric dynamics*: this is a nontrivial problem in financial time series analysis since the volatility level and dynamics are highly sensitive to the sign of past returns. The so called leverage effect (Black, 1972) suggests that current volatility and past asset returns are negatively cross-correlated, in the sense that we expect that a past negative return will lead to a positive variation of volatility higher than the variation that would have been generated by a positive return of the same magnitude.

Hence, measuring and forecasting volatility of financial returns is of fundamental importance in many financial applications such as risk management, hedging, option pricing and, what is more in line with our work, portfolio management and selection.

Over the last three decades, the above discussed stylized facts have stimulated the research on stochastic dynamic modeling of volatility. As previously mentioned, we have recently assisted to the rise of a growing interest in direct modeling of volatility measures built from intra-daily information. A prerequisite for these models is the availability of accurate ex-post volatility measures to be used, as observed data, for model identification, estimation and diagnostics.

This has led many academic researchers and practitioners to focus their research on ex-post non-parametric estimation of the daily variance and covariance of asset returns using information on intra-daily price quotations. The non-parametric procedures are generally free from explicit functional form assumptions regarding the expected volatility and afford estimates of notional volatility (as defined in [Aït-Sahalia & Hansen \(2009\)](#)) that are flexible yet consistent. The non-parametric approaches that utilize the ex-post returns in extracting measures of the notional volatility include the recently-popularized *realized volatility measures* for fixed-length time intervals.

All these estimators are derived under the assumption that the log-transformed price process follows, in continuous time, a time homogeneous diffusion process. Realized measures use high frequency data to estimate the parameters of interest in the stochastic structure of this process which is the *integrated variance* of stock prices and its multivariate counterpart, known as *integrated covariance*.

1.1.2 Quadratic variation and integrated variance

Consider a logarithmic price process X of a generic i -th asset observed at equally-spaced discrete points in time over a given horizon τ , where the observation times are assumed to be strictly increasing. The realizations of X are observed at the observation

times $\{0, 1/J, 2/J, \dots, \tau - (1/J), \tau\}$ where J is the intraday sampling frequency and $J, \tau > 0$.

The observed log price process is assumed to be driven by the latent efficient log price process Y , which is modeled as a Brownian semimartingale ($Y \in BSM$) to rule out arbitrage opportunities. This is expressed as:

$$dY(t) = \mu(t)dt + \sigma(t)dW(t), \quad 0 \leq t \leq \tau, \quad (1.1)$$

where $\mu(t)$ is a predictable locally bounded drift process, $\sigma(t)$ is a càdlàg volatility process (not necessarily independent of W), and W is a standard Brownian motion. Hence, the processes $\mu(t)$ and $\sigma(t)$ signify the instantaneous conditional mean and volatility of the return. The relation between the real price and the observed price can be explained by

$$X_t = Y_t + \epsilon_t, \quad 0 < t \leq \tau, \quad (1.2)$$

where ϵ_t is the additive noise component satisfying the conditions

- (i) $\mathbb{E}(\epsilon_t) = 0$
- (ii) $\text{Var}(\epsilon_t) < \infty$.

Also, it holds that the error term is independent from the price process Y . The intraday return is then denoted as

$$r(t, t - 1/J) = X_t - X_{t-1/J} \quad (1.3)$$

and its *quadratic variation* $QV(t, t + \tau)$ over the same horizon is

$$QV(t, t + \tau) = \int_t^{t+\tau} \sigma^2(u)du. \quad (1.4)$$

To facilitate the presentation, we can normalize the unit time interval, or $\tau = 1$, to represent one trading day, such that we can refer to t as the daily sampling. Equation (1.4) shows that innovations to the mean component $\mu(t)$ do not affect the sample path variation of the return. The intuition is that the mean term, $\mu(t)dt$, is of lower order in terms of second order properties than the diffusive innovations, $\sigma(t)dW(t)$. Thus, when cumulated across many high-frequency returns over a short time interval they can effectively be neglected. The diffusive sample path variation over $[t, t + 1]$ is also known as the *integrated variance* $IV(t, t + 1)$,

$$IV(t, t + 1) = \int_t^{t+1} \sigma^2(u)du. \quad (1.5)$$

Equations (1.4)-(1.5) show that, under the simplifying assumptions of absence of microstructure noise and measurement error, the quadratic and integrated variation coincide. This is however no longer true for more general return processes like, for example,

if the underlying continuous time process in Eq.(1.1) contains jumps. In this case, the quadratic return variation process over the horizon τ , is the sum of the diffusive integrated variance and the cumulative squared jumps; for further discussion along these lines see e.g. Andersen, Bollerslev, Diebold & Labys (2001).

The Realized Variance (RV) estimator

The simplest and most widely used estimator of the integrated variance is given by the Realized Variance (RV). The leading reference on the topic is the paper by Andersen et al. (2003), where results are achieved in the case $X_t = Y_t$, i.e. assuming no measurement error.

Under this assumption, the return quadratic variation can be approximated arbitrarily well by the corresponding cumulative squared intra-daily return process. Consider a partition $\{t + \frac{j}{J}, j = 1, \dots, J\}$ of the $[t, t + 1]$ interval. Then the realized variance (RV)¹ of the logarithmic price process is

$$RV(t, t + 1) = \sum_{j=1}^J r \left((t + 1) - \frac{j}{J}, (t + 1) - \frac{j-1}{J} \right)^2. \quad (1.6)$$

Semimartingale theory ensures that the realized variance measure converges in probability to the return quadratic variation QV defined in Eq.(1.4), when the sampling frequency J increases:

$$RV(t, t + 1) \xrightarrow{P} QV(t, t + 1).$$

In other words, by summing sufficiently finely sampled high-frequency returns, it is possible to construct ex-post realized volatility measures for the integrated latent volatilities that are asymptotically free of measurement error. Although the squared return over the interval horizon provides an unbiased estimate for the realized integrated volatility, it is an extremely noisy estimator, and predictable variation in the true latent volatility process is typically dwarfed by measurement error. In addition, for longer horizons any conditional mean dependence will tend to contaminate this variance measure. In contrast, as the length of the return horizon decreases the impact of the drift term vanishes, so that the mean is effectively annihilated.

Building on the results in Jacod & Protter (1998), Barndorff-Nielsen & Shephard (2002) derived the asymptotic distribution of the realized variance as follows:

$$\left(\frac{RV(t, t + 1) - QV(t, t + 1)}{\sqrt{2IQ(t, t + 1)/J}} \right) \xrightarrow{d} N(0, 1), \quad (1.7)$$

where $IQ(t, t + 1) = \int_t^{t+1} \sigma_s^4 ds$ is the so-called integrated quarticity of the process,

¹Some authors refer to the square root of RV as the *Realized Volatility* estimator.

which is a random unknown quantity consistently estimated by the Realized Quarticity, $RQ(t, t+1) = (\frac{J}{3}) \sum_{j=1}^J r(t - \frac{j}{J}, t - \frac{(j-1)}{J})^4$. This conclusion is heavily dependent on the absence of jumps in the price process. Importantly, in the presence of jumps the conditional distribution of the returns in Eq.(1.3) is no longer Gaussian. As such, the corresponding empirical distribution of the standardized returns speaks directly to the relevance of allowing for jumps in the underlying continuous time process when analyzing the returns over longer horizons.

So far we have not yet been explicit about the relationship between realized volatility, quadratic variation and conditional volatility, that is the object of econometric interest. What is important to stress is that when the drift and volatility coefficients are constant, the conditional (and unconditional) return variance equals the quadratic variation of the log price process. On the contrary, when volatility is stochastic as in Eq.(1.1), we must distinguish clearly between conditional variance, representing the (ex-ante) expected size of future squared return innovations over a certain period, and the quadratic variation, reflecting the actual (ex-post) realization of return variation, over the corresponding horizon. Hence, the one concept of a priori expectations is opposed to the one of subsequent actual realizations of return volatility. Under the ideal conditions discussed above, the realized volatility captures the latter, but not the former. Nevertheless, as already mentioned, well calibrated forecasts (i.e., conditional expectations) of return volatility can be obtained by exploiting useful information from time series of past realized volatilities. Since RV is approximately unbiased for the corresponding unobserved QV , the realized volatility measure is the natural benchmark against which to gauge the performance of volatility forecasts.

In the very recent past there have been various elaborations to this literature. Two are worth mentioning. First, more effort has been put toward the study of the total quadratic variation in the asset return process, which may be attributable to both a continuous diffusion and discrete jump process. [Barndorff-Nielsen & Shephard \(2004\)](#) and [Andersen et al. \(2007\)](#) develop methods for measuring the contribution to QV from both sources. [Andersen et al. \(2007\)](#) propose a formal test for identifying when significant contributions from jump activity occur. By separating total volatility into jump and diffusion components, they find that almost all of predictability in volatility stems from the persistent diffusive component.

Second, there has been interest in studying the impact of market microstructure effects on the estimates of realized variance. This causes the estimator of the QV to become biased. As noted by [Aït-Sahalia et al. \(2005\)](#), Realized Variance properties change by considering that in most empirically realistic situations, the errors are often nontrivial. Under the same additive model given in Eq.(1.2), they analyze RV under market microstructure noise (with the assumption that the number of observations in $[t, t + \tau]$ goes to infinity and the maximum distance in time between two consecutive observations goes to zero) and prove that RV is not a reliable estimator for the true variation of the returns, since it has a positive bias whose magnitude increases linearly with the sample size. They argue that RV relates to the wrong quantity and that the problem gets worse when observations are sampled more frequently, since market microstructure

noise totally swamps the variance of the price signal.

Consequently, the recommendation in the literature has been to use a realized variance estimator constructed by summing squared log-returns sampled at some lower frequency: 10, 15, 30 minutes typically;² see Andersen & Bollerslev (1998), Andersen, Bollerslev, Diebold & Labys (2001) and Bandi & Russell (2008). Alternatively, other realized estimators have been proposed: examples are the Kernel-based estimators of Hansen & Lunde (2006) and Barndorff-Nielsen et al. (2008), the Two-scales realized volatility of Zhang et al. (2006), the Realized Range Variance of Christensen & Podolskij (2007) and the Scaled Realized Range of Martens & Van Dijk (2007), to name a few.

1.1.3 Quadratic covariation and integrated covariance

In the financial econometrics literature the covariance between returns on different assets is widely called co-volatility or cross-volatility and, as its univariate counterpart, is a latent variable. The concepts we introduce in the following can be considered as a generalization of those explained in more details in Section 1.1.2. As in the univariate case, we focus on the continuous time price generating process, assuming again $\tau = 1$.

Let n define a number of stocks whose log price processes are denoted by $X_t^{(i)}$ for $i = 1, \dots, n$ and let $\mathbf{X}_t = (X_t^{(1)}, X_t^{(2)}, \dots, X_t^{(n)})'$ denote the joint process at time t . The latent efficient price process $\mathbf{Y}(t, t+1)$ is again modeled as a *BSM*:

$$\mathbf{Y}(t, t+1) = \int_t^{t+1} \boldsymbol{\mu}(u) du + \int_t^{t+1} \boldsymbol{\Theta}(u) d\mathbf{W}(u)$$

where $\boldsymbol{\mu}(u)$ is a n -dimensional vector of elements which are predictable locally bounded drifts, $\Omega(u) = \boldsymbol{\Theta}(u)\boldsymbol{\Theta}'(u)$ is the instantaneous or spot co-volatility process with elements that are all càdlàg over a fixed time interval set to vary, as before, between t and $t+1$ and $\mathbf{W}$ is a vector of independent standard Brownian motions.

As previously done, we maintain the assumption that the observed prices are the true efficient prices, so that $\mathbf{X}_t = \mathbf{Y}_t$. Under this scenario, the quadratic covariation matrix of the multivariate price process is defined as

$$QC(t, t+1) = \lim_{J \rightarrow \infty} \sum_{j=1}^J \left(\mathbf{X}_{(t+1)-\frac{j}{J}} - \mathbf{X}_{(t+1)-\frac{(j-1)}{J}} \right) \left(\mathbf{X}_{(t+1)-\frac{j}{J}} - \mathbf{X}_{(t+1)-\frac{(j-1)}{J}} \right)' \quad (1.8)$$

where the limit is finite and well defined with probability one. Following advances in the realized co-volatility literature, we are interested in modeling and forecasting the daily ex-post covariation which almost surely equals the daily integrated covariance.

² Formally, this so called *sparse sampling* is implemented by sampling on a subgrid (selected arbitrarily or optimally) of the full grid containing all of the observation points $\{1/J, 2/J, \dots, \tau\}$. However it comes at the cost of a less precise estimate of the integrated variance and, in general, not advisable since it implies an efficiency loss that, depending on the amount of discarded data, can be sizable.

For the price process in Eq.(1.1.3), the *Integrated Covariance (IC)* matrix for day t simply equals

$$IC(t, t+1) = \int_t^{t+1} \Omega(u) du. \quad (1.9)$$

It can be considered as the multivariate counterpart of the integrated variance examined previously and it plays a central role in the construction of the realized covariance models that will be covered in the following chapters.

The basic Realized Covariance estimator

Barndorff-Nielsen & Shephard (2004) introduce the first generation of estimators of integrated covariance, namely the so called *Realized Covariance (RC)* henceforth, which is defined as follows.

Suppose we can observe the log price processes at discrete time points over the interval $[t, t+1]$ constituted by $j = 1, \dots, J$ intra-daily observations, and define $M = \sup_j \left| \left(t + \frac{j}{J} \right) - \left(t + \frac{(j-1)}{J} \right) \right|$, where for $J \rightarrow \infty$, $M \rightarrow 0$.

Then the *RC* estimator at time t corresponds to the finite sum given in the definition of quadratic covariation in Eq.(1.8), as follows:

$$RC(t, t+1) = \sum_{j=1}^J \left(\mathbf{X}_{(t+1)-\frac{j}{J}} - \mathbf{X}_{(t+1)-\frac{(j-1)}{J}} \right) \left(\mathbf{X}_{(t+1)-\frac{j}{J}} - \mathbf{X}_{(t+1)-\frac{(j-1)}{J}} \right)' \quad (1.10)$$

The realized covariance matrices are symmetric by construction and, for $n < J$, positive definite almost surely. Also, under the assumptions of observations simultaneously observed and absence of microstructure noise, the realized covariance is a consistent estimator for $IC(t, t+1)$ and its estimation error converges to a mixed normal distribution with $\sqrt{J}$ rate; see, for example, Jacod & Protter (1998), Barndorff-Nielsen & Shephard (2002) and Mykland & Zhang (2009).

However, in the real world things are not so simple. The asset prices are not simultaneously observed and practical calculation of the estimator in Eq.(1.10) requires transforming the observed price data to an evenly spaced multivariate time series by adopting some interpolation rule. The general idea about synchronization is to resample the original tick-by-tick data with respect to a pre-specified grid of trading times that can be equally or un-equally spaced. This choice influences the properties of the realized covariance estimator as it both affects the number of returns on which *RC* is computed as well as the level of noise.

A widespread and simple solution is to adopt the *previous-tick* interpolation. This consists in subsampling the data by considering the prices that immediately precede (or are at) the pre-specified sampling points, which are equally-spaced by definition. It can

be defined by the formula

$$X_h^{(i)} = X^{(i)}(h) = X^{(i)} \left(\max \left\{ \left(t + \frac{j}{J} \right)^{(i)} : \left(t + \frac{j}{J} \right)^{(i)} \leq \frac{h}{H} \right\} \right), \quad (1.11)$$

where H denotes the number of subsampled intervals over the whole period $[t, t + 1]$ and h is the given sampling point.³

Alternatively, [Barndorff-Nielsen et al. \(2011\)](#) propose the *refresh time* sampling, a more advanced technique that deals with non-synchronicity by considering the specific asset price that takes place only when all assets have been traded. Specifically, refresh time can be interpreted as the time taken for all the prices to be refreshed. The final refresh time's sample size is determined by the degree of non-synchronicity and by the sample size of the n assets. Intuitively, the main drawback of this approach is that it is highly dependent on the relatively illiquid assets, which can lead some bias in the estimation process.

So, moving from an univariate to a multivariate framework, beyond the usual problem of microstructural noise affecting the data, an additional relevant problem stems from the asynchronized nature of price observations for multiple assets. Nonsynchronicity in the trading process of different assets is of fundamental importance whenever high-frequency data are used to estimate their covariances. While, theoretically, realized covariances in the absence of noise are unbiased and consistent, this is no longer true when nonsynchronicity is present.

Nevertheless, it must be said that the widespread use of the RC estimator in practical applications is related to the possibility of controlling its bias by choosing an appropriate frequency of discretization. This applies when assets that are very frequently traded, so that we are confident that the effect of nonsynchronicity is rather mild at the chosen frequency. What is also common practice in order to avoid the noise induced by measuring the overnight volatility as the squared overnight return, is to assume that investors open a position at the beginning of the trading day and close it at the end of the same trading day. This implies using open-to-close data only and measuring the volatility over the daily trading session.

The *CholCov* estimator

Among the integrated covariance estimators existing in the literature, we briefly recall the so called *CholCov* estimator recently proposed by [Boudt et al. \(2014\)](#), as it will be used in several empirical applications in the remainder of the thesis to check the robustness of the results achieved using the standard RC of Eq.(1.10).⁴

³For a generic interval $[t, t + \tau]$, the formula becomes $X^{(i)}(h) = X^{(i)} \left(\max \left\{ \left(t + \frac{j}{J} \right)^{(i)} : \left(t + \frac{j}{J} \right)^{(i)} \leq \frac{h\tau}{H} \right\} \right)$.

⁴It is worth to mention that other interesting estimators not discussed for sake of space are the Cumulative Covariance of [Hayashi & Yoshida \(2008\)](#), the Multivariate Realized Kernel of [Barndorff-](#)

This estimator of the ex-post covariation of log-prices can take into account both asynchronicity and microstructure noise, with the nice feature of being positive semidefinite by construction.

As already stressed, the problem encountered in the presence of non-synchronous observations is that estimation of the off-diagonal elements of $IC(t, t + 1)$ in Eq.(1.9) is not possible unless a method for synchronizing the series is adopted. Two main approaches are popular: one is to synchronize all observations jointly, at the cost of letting the resulting observation frequency be determined by the least liquid asset (refresh-time approach); the second is to estimate the off-diagonal elements of the matrix separately using synchronized pairs of price processes (composite approach), but the result is not ensured to be positive semidefinite (PSD). The novelty of the *CholCov* estimator is that it obtains through a Cholesky factorization of the spot covariance matrix into symmetric square matrices, which enables to estimate the different parameters sequentially using as many observations as possible. To be more precise, $\Omega(u)$ is decomposed as

$$\Omega(u) = \Upsilon(u)G(u)\Upsilon(u)', \quad (1.12)$$

where $\Upsilon(u)$ is a lower diagonal matrix with ones on the diagonal, and $G(u)$ a diagonal matrix. Namely,

$$\Upsilon = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \Upsilon_{21} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ \Upsilon_{n1} & \Upsilon_{n2} & \dots & 1 \end{bmatrix}, \quad G = \begin{bmatrix} G_{11} & 0 & \dots & 0 \\ 0 & G_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & G_{nn} \end{bmatrix}. \quad (1.13)$$

The links between the elements of the Υ and G matrices to the elements of Ω are given by the following relations:

$$G_{kk} = \Omega_{kk} - \sum_{m=1}^{k-1} \Upsilon_{km}^2 G_{mm},$$

$$\Upsilon_{kl} = \frac{1}{G_{ll}} \left(\Omega_{kl} - \sum_{m=1}^{k-1} \Upsilon_{km} \Upsilon_{lm} G_{mm} \right),$$

for $k = 2, \dots, n$ with $G_{11} = \Omega_{11}$.

Sequential estimation is possible because the (kl) -th element of the covariance matrix only depends on G_{mm} and Υ_{pq} with $m, p, q \leq \max(k, l)$.

The Cholesky decomposition thus allows to exploit the heterogeneity in trading intensities and to estimate the different parameters sequentially, using as much data as possible. The decomposition holds for any PSD matrix, which is by definition true for

Nielsen et al. (2011), the Modulated Realized Covariance of Christensen et al. (2010) and the Blocking and Regularization estimator of Hautsch et al. (2012). Not all of them have the same properties of the Cholcov, which makes the latter an appealing choice.

the covariance matrix, and the recombined results will also be PSD. However, ensuring positive semidefiniteness comes at a cost. It is not possible to synchronize just pairs of data, but is needed to sample over a growing set of series. Namely, while the diagonal elements are estimated sampling over just the single series, it is required to sample over more series for the off-diagonal estimates. For example, for the (kl) element all series in $\{1, \dots, l, k\}$ are synchronized. To make optimal use of the data, it is therefore crucial to order the series in terms of decreasing liquidity.

The liquidity criterion proposed by the authors is the squared *duration sorting* method based on the following formula:

$$\sum_{j=0}^{J_i} \left(\Delta_j^{(i)} \right)^2,$$

where it is accounted for both the number of each i -th asset intraday available returns (J_i) and also the degree to which these returns are spread out over the day (i.e. the asset's duration $\Delta^{(i)}$). Hence, the first series is the most liquid, and the last one is the least liquid. This approach is different, for example, from the *observation count sorting* proposed in [Hautsch et al. \(2012\)](#).

The basic *CholCov* estimation algorithm in the presence of asynchronicity is made up of three main recursive steps and the authors demonstrate that its computational intensity is fully compensated by an efficiency gain if compared to other robust to asynchronicity estimators. The final estimator is obtained by recombining together the achieved matrices:

$$Cholcov = \hat{\Upsilon} \hat{G} \hat{\Upsilon}',$$

with $\hat{\Upsilon}$ and $\hat{G}$ the corresponding estimates for Υ and G .

Finally, to deal with microstructure noise, when present, they advice to refine the procedure by employing noise-robust integrated variance and beta estimators implemented after the price-series have been synchronized.

Simulation results demonstrate that the resulting estimates are accurate and better conditioned than full-dimensional estimators (like the Modulated Realized Covariance of [Christensen et al. \(2010\)](#) and the Multivariate Kernel of [Barndorff-Nielsen et al. \(2011\)](#)) or their composite versions (e.g. [Lunde et al. \(2016\)](#)).

Realized covariance matrices built using this estimator will be employed in the empirical applications presented in Chapter 2 and Chapter 4.

Chapter 2

A class of multiplicative models for realized covariance matrices¹

Abstract This chapter introduces a class of multiplicative dynamic models for realized covariance matrices. The multiplicative structure enables consistent three-step estimation of the parameters, starting by covariance targeting of a scale matrix. The dynamics of conditional variances and correlations are inspired by specifications akin to the consistent dynamic conditional correlation model of the multivariate GARCH literature, and estimation is performed by the quasi-maximum likelihood method, assuming a Wishart conditional distribution. Simulations show that in finite samples a three-step estimator has smaller bias and root mean squared error than a one-step estimator when the cross-sectional dimension increases. An empirical application illustrates the flexibility of the model in a low-dimensional setting, and another one illustrates its effectiveness and practical usefulness in high dimensional portfolio allocation strategies.

2.1 Introduction

The availability of high-frequency data to construct measures of daily variances of financial returns, and covariances between these, allows researchers to model time series of realized covariance matrices. One interest of these models is that they can be used for forecasting, which is of use in financial applications such as hedging, option pricing, risk management and portfolio allocation. Another potential interest of models for realized covariance matrices is that they allow researchers to study the macroeconomic and financial determinants of the changes in multivariate volatility. GARCH models can be used for the same purposes - see for example [Engle & Rangel \(2008\)](#) - but since they rely on daily observed returns, in principle they provide less precise estimates and forecasts of variances and covariances than measures based on intraday data.

¹This chapter is joint work with Luc Bauwens and Giuseppe Storti.

Modeling a covariance matrix is a challenging task for two reasons. First, since the dimension of the object to be modeled is proportional to the square of the number of assets, the number of parameters easily becomes unmanageable even for a handful of assets. This suggests that models should be parsimoniously parameterized to avoid the parameter proliferation issue. Second, models should be congruent with the property that covariance matrices are positive definite. Several dynamic models for realized covariance matrices use the Wishart distribution, which is a natural choice since the support of the Wishart is the set of positive definite symmetric (PDS) matrices. We also use this distribution as a basis for estimation.

The main contribution of this chapter is to propose a class of dynamic models for realized covariance matrices, which are computationally tractable and readily applicable in large dimensional settings. Its main features are a multiplicative structure and a parameterization inspired by the consistent dynamic conditional correlation (cDCC) model of the multivariate GARCH literature (see [Engle \(2002a\)](#) and [Aielli \(2013\)](#)). We call this type of models realized consistent DCC (Re-cDCC). Jointly with the conditional Wishart assumption, these features lead to two important advantages.

First, the multiplicative structure allows to separate the parameters of the constant long-term covariance matrix from those of the dynamic equations of variances and correlations, such that it is possible to target (i.e. estimate consistently) the constant scale matrix term by the sample average of the observed covariance matrices. After this, the parameters of the dynamic variances and correlations can be estimated in two additional steps, the first one for the realized variances, and the second one for the realized correlation matrix. This extends over similar models making use of a less flexible BEKK-type parameterization, see [Golosnoy et al. \(2012\)](#) and [Noureldin et al. \(2012\)](#).

Second, due to the properties of the Wishart distribution, the estimators of the parameters of the dynamic processes of the variances and correlations at each step have a quasi-maximum likelihood (QML) interpretation. Estimation of all the parameters, including the constant scale matrix, can be performed by full QML in a single step, if the cross-sectional dimension is small (at most fifteen).

The stepwise estimation procedure allows us to estimate the models for a relatively large order of the realized covariance matrices, up to fifty in the empirical illustration and one hundred in simulations. Our ability to model matrices of such dimensions contrasts with existing models that also rely on the Wishart assumption: [Gouriéroux et al. \(2009\)](#) work with three assets, [Bonato et al. \(2009\)](#) with four, [Golosnoy et al. \(2012\)](#) and [Jin & Maheu \(2013\)](#) with five, [Noureldin et al. \(2012\)](#) with ten, and [Bonato et al. \(2012\)](#) with twelve. It also surpasses the dimension used in different modeling approaches for realized covariance matrices: [Chiriac & Voev \(2011\)](#) work with six assets and [Bauer & Vorkink \(2011\)](#) with five.

In order to assess the merits of the Re-cDCC models in practice, we undertake a comprehensive out-of-sample forecasting study using series of daily realized covariance matrices of different cross-sectional dimensions and spanning different time periods.

In a first study, we illustrate the models on a dataset composed of two stocks, where the correlation process is parameterized by scalar and non-scalar cDCC processes. We show that the Re-cDCC models perform quite favourably in comparison to benchmarks both in- and out-of-sample, and that the in-sample fit of the scalar model is no worse than the considered alternatives. In a second study, we focus on the scalar specification and evaluate the model forecasting accuracy using two datasets spanning different time periods. An application of the model confidence set (MCS) of Hansen et al. (2011) reveals that the Re-cDCC scalar model leads to statistically significant gains in one-step ahead predictions, irrespective of the chosen loss function or of the forecasting period considered. Moreover, applying the model to optimal portfolio allocation, we find that during the turbulent period of the 2008 financial crisis it improves the portfolio return and achieves the lowest standard deviation among the considered set of competing models, yet maintaining the smallest concentration. These results demonstrate that the flexibility of the Re-cDCC induced by capturing the dynamics of realized variances and correlations through a cDCC structure, which is missing in similar models featuring BEKK or DECO-type (Engle & Kelly (2012)) specifications, improves the prediction of the conditional covariance matrix of returns.

Section 2.2 introduces the proposed class of Re-cDCC models for realized covariance matrices. Section 2.3 covers the three-step QML estimation method applicable to these models. Section 2.4 presents the results of a simulation study to investigate the potential efficiency loss of the proposed three-step estimator with respect to a one-step estimator (in low dimension) and the bias and efficiency properties of the three-step estimator. Empirical results are provided in Section 2.5 and concluding remarks are presented in the Section 2.6.

2.2 The multiplicative realized model class

2.2.1 Multiplicative framework

The framework recalled in the following is the same already discussed in Section 1.1.3. Let C_t be a $n \times n$ positive-definite symmetric (PDS) unbiased estimator of the latent integrated covariance matrix of the daily logarithmic return vector $r_t = (r_{1,t}, \dots, r_{n,t})'$ of n assets. For the moment we will not be explicit about the specific realized covariance estimator we chose, as this is not crucial at this stage and will be made clear later on.

We assume that conditional on past information $\mathfrak{S}_{t-1}$, C_t follows a n -dimensional central Wishart distribution:

$$C_t | \mathfrak{S}_{t-1} \sim W_n(\nu, S_t / \nu), \quad \forall t = 1, \dots, T, \quad (2.1)$$

where ν ($> n - 1$) is the degrees of freedom parameter and S_t is the PDS conditional expectation of C_t , denoted by $\mathbb{E}(C_t | \mathfrak{S}_{t-1}) := \mathbb{E}_{t-1}(C_t)$. By the properties of the Wishart

distribution, under the assumption of absence of microstructure noise and other biases (see Section 1.1.3), S_t represents the conditional covariance matrix of returns, so that the i, j -th element of S_t is the conditional covariance between daily returns on assets i and j , $\text{Cov}(r_{i,t}, r_{j,t} | \mathfrak{S}_{t-1})$, for $i, j = 1, \dots, n$.

The unconditional expectation of C_t is assumed to be a constant matrix:

$$\mathbb{E}(S_t) = M = LL', \quad (2.2)$$

where the Cholesky factorization is used in the second equality, such that L (L') is a $n \times n$ lower (upper) triangular scale matrix of parameters. By applying a multiplicative decomposition, S_t can be rewritten as a product of positive-definite matrices, namely

$$S_t = LH_tL' \quad (2.3)$$

where H_t must satisfy the property that $\mathbb{E}(H_t) = I_n$, I_n being the $n \times n$ identity matrix. It immediately follows that the conditional covariance S_t can be interpreted as the result of the interaction of two components, namely a mean reverting matrix-valued (short-term) process H_t and a constant long-term level M around which H_t fluctuates from day to day. A similar structure has been proposed by Golosnoy et al. (2012) and, in the context of MGARCH models, by Bauwens et al. (2013). The main difference is that we use a constant matrix L for C_t while they use a time-varying matrix computed either parametrically or as a nonparametric function of time. We discuss the main pros and cons of their approach in the remainder of this section.

2.2.2 Parameterization of the short-term component

Our idea is to separately modeling the elements of the matrix H_t , in a way that permits some flexibility and does not prevent to estimate the model parameters in large dimensions. Similarly to the approach developed in the multivariate GARCH literature, we resort to parametric specifications, according to dynamic processes that depend on the available information set.

Modeling volatilities and correlations

There is a substantial literature on how to model the dynamics of the conditional covariance matrix of returns in a stationary framework. This means modeling the fluctuations of S_t around M , which, in our case, translates into modeling the dynamics of the short-term component H_t around its long run level I_n .

Inspired by the dynamic conditional correlation (DCC) model of the multivariate GARCH literature, we use the representation of a covariance matrix in terms of the corresponding diagonal matrix of standard deviations and the corresponding correlation matrix (see Tse & Tsui (2002), Engle (2002a) and the *consistent* or *corrected* version

proposed by [Aielli \(2013\)](#)). Therefore, the H_t component in Eq.(2.3) is further decomposed as

$$H_t = D_t R_t D_t, \quad (2.4)$$

where $D_t = \{\text{diag}(H_t)\}^{1/2}$ is the diagonal matrix with i -th diagonal entry equal to $\sqrt{H_{ii,t}}$ and R_t is the corresponding correlation matrix. This type of decomposition enables us to separately specify the dynamic equation of each short-term conditional variance and of the short-term conditional correlation matrix. In the sequel, when the difference between the components is made clear, we shall simply refer to $H_{ii,t}$ as the conditional variance of asset i and to R_t as the conditional correlation matrix of the assets. The specifications used for their temporal dynamics are explained in the following.

To begin with, we opt for separate univariate specifications for the conditional variances which exclude interaction (spillover) terms, since this simplifies inference and allows for a stepwise estimation approach as explained in the next section. We opt for a GARCH-type recursion of the form:

$$H_{ii,t} = (1 - \gamma_i - \delta_i) + \gamma_i C_{ii,t-1}^* + \delta_i H_{ii,t-1}, \quad (2.5)$$

where $C_{ii,t-1}^*$ is a diagonal element of

$$C_{t-1}^* = L^{-1} C_{t-1} L'^{-1}. \quad (2.6)$$

The matrix C_t^* is the observed realized covariance matrix transformed to have its unconditional expectation equal to the identity matrix of order n , so that $C_{ii,t-1}^*$ is interpretable as the short-term realized variance of asset i . This is used in Eq.(2.5) as the factor that drives the volatility dynamics, instead of the less precise squared daily asset return used in GARCH equations, but like in the realized GARCH model of [Shephard & Sheppard \(2010\)](#). Restrictions on the parameters are that $\gamma_i > 0, \delta_i > 0$ and $\gamma_i + \delta_i < 1$. Notice that the constant term in Eq. (2.5) implies that $\mathbb{E}(H_{ii,t}) = 1$, in conformity with the restriction $\mathbb{E}(H_t) = I_n$.

With respect to proposals available in the literature, such as the ARFIMA model used in [Andersen et al. \(2003\)](#) and the HAR model of [Corsi \(2009\)](#), the choice of a GARCH(1,1)-type recursion is convenient as it is parsimonious in parameters and permits to impose easily the restriction $\mathbb{E}(H_{ii,t}) = 1$. Richer specifications can be considered, with more lags or nonlinear effects.

The next stage entails defining the dynamic process of the conditional correlation matrix R_t . The most general formulation we consider consists in the following equations:

$$Q_t = (I_n - AA' - BB') + AC_{t-1}^Q A' + BQ_{t-1}B' \quad (2.7)$$

$$R_t = \{Q_t\}^{-1/2} Q_t \{Q_t\}^{-1/2} \quad (2.8)$$

where $\{A, B\}$ are $n \times n$ matrices of parameters and Q_t is a PDS matrix representing the correlation driving process at time t , with fixed initial value $Q_0 = I_n$. To shorten notations, $\{Q_t\}^{1/2}$ denotes the diagonal matrix $\text{diag}(Q_{ii,t}^{1/2}), i = 1, 2, \dots, n$. The relation

in Eq.(2.8) is necessary to transform Q_t into a correlation matrix since its diagonal elements are not necessarily equal to 1. Notice that in order to obtain a *consistent* (in the sense of Aielli (2013)²) DCC-type specification, C_t^Q has to be defined as

$$C_t^Q = \{Q_t\}^{1/2} D_t^{-1} C_t^* D_t^{-1} \{Q_t\}^{1/2}, \quad (2.9)$$

with C_t^* defined as in Eq.(2.6) and D_t denoting the diagonal matrix of short-term conditional standard deviations $\sqrt{H_{ii,t}}$. The matrix $D_t^{-1} C_t^* D_t^{-1}$ represents the short-term realized covariance matrix purged from its short-term conditional variances. The pre- and post-multiplication by $\{Q_t\}^{1/2}$ is necessary to achieve equality between the unconditional expectations of C_t^Q and Q_t , by taking expectations on both sides of Eq.(2.7) and recalling the equality $\mathbb{E}(Q_t) = I_n$.

Equations (2.4) – (2.9) define the most general formulation of the Realized consistent DCC (Re-cDCC) model, which we consider as the main representative of the multiplicative realized class. It is probably useful to outline that other members of the same family can be derived just by changing the dynamic process for the correlation matrix R_t . For example, in the spirit of the Dynamic Equicorrelation (DECO) model of Engle & Kelly (2012), the Re-cDECO model is obtained by averaging all the pairwise DCC correlations across the day in order to get a new, equicorrelated matrix R_t^{DECO} , according to the following equations

$$R_t^{DECO} = (1 - \rho_t) I_n + \rho_t J_n, \quad (2.10)$$

$$\rho_t = \frac{1}{n(n-1)} (\boldsymbol{\iota} R_t^{DCC} \boldsymbol{\iota}'), \quad (2.11)$$

with $\rho_t > 0$ the time t equicorrelation parameter, I_n the identity matrix of order n , J the n -dimensional matrix of ones and $\boldsymbol{\iota}$ a row vector of ones.

While sharing a similar structure, the Re-cDECO is nonnested in the Re-cDCC just like the DECO is nonnested in the DCC. The models receive the same information but use it in very different ways to get a final estimate of the daily covariance matrix of returns. Specifically, the Re-cDECO correlations between any pair of assets are related to past realizations of *all* pairs, so that the model exploits an information pooling mechanism that is lacking in the Re-cDCC. Moreover, the equicorrelation assumption keeps estimation tractable even for very large matrices as these do not have to be inverted at each time step of the procedure. As these features can be beneficial to handle vast dimensional systems, we use the Re-cDECO as a valuable competitor of the Re-cDCC in the large sample portfolio application presented in the empirical section.

²Aielli (2006) and more in details Aielli (2013) show that in the standard DCC correlation driving process proposed by Engle (2002a), where in the scalar case it is $Q_t = (1 - \alpha - \beta)S + \alpha\epsilon_{t-1}\epsilon'_{t-1} + \beta Q_{t-1}$, replacing S by the sample second moment of the estimated standardized returns ϵ_t is not a correct estimation device, as the equality $S = \mathbb{E}(\epsilon_t \epsilon'_t)$ does not hold in general, unless in the most trivial case of equal conditional correlations, i.e. $\alpha = \beta = 0$.

2.2.3 Feasible parameterizations

By imposing parametric restrictions on the dynamic matrix process Q_t accommodated in Eq.(2.7), different conditional correlation models can be obtained. Herein, three readily applicable parameterizations are illustrated: scalar, diagonal and Hadamard; for each of them we define the constraints required for the correlation process to be stationary and give advises on their application depending on the cross-sectional dimension of the dataset to be modeled. For further insights about usage and performance of more general non-scalar DCC models we refer to the recent paper by [Bauwens, Grigoryeva & Ortega \(2016\)](#).

Scalar specification The scalar specification is the most used in practical applications when parameter proliferation needs to be contained. It obtains by setting $A = \alpha^{1/2}I_n$ and $B = \beta^{1/2}I_n$, thus restricting the dynamics of all the elements of Q_t to be driven by the same parameters. The scalar version of Eq.(2.7) is written as

$$Q_t = (1 - \alpha - \beta) + \alpha C_{t-1}^Q + \beta Q_{t-1}, \quad (2.12)$$

and the constraints to be imposed on the parameters are $\alpha > 0$, $\beta \geq 0$ and $\alpha + \beta < 1$. The latter ensures the existence of $\mathbb{E}(Q_t)$. As the number of parameters implied by Eq.(2.12) is equal to two independently of the dimension of the cross-section, the scalar specification is worth using in the modeling of realized covariance matrices in dimensions up to fifty, as we show later on, and perhaps more.

Diagonal specification The diagonal specification permits asset-specific parameter heterogeneity in the dynamic structure. It implies that the parameter matrices $\{A, B\}$ are diagonal, *i.e.* with diagonal elements equal to $\alpha_{ii}^{1/2}$ and $\beta_{ii}^{1/2}$. We further impose that $\alpha_{ii}^{1/2} > 0$ and $\beta_{ii}^{1/2} \geq 0$. Denoting $\tilde{A} = \text{diag}(\alpha_{ii}^{1/2})$ and $\tilde{B} = \text{diag}(\beta_{ii}^{1/2})$ for $i = 1, \dots, n$, the diagonal model is given by

$$Q_t = (I_n - \tilde{A}\tilde{A}' - \tilde{B}\tilde{B}') + \tilde{A}C_{t-1}^Q\tilde{A}' + \tilde{B}Q_{t-1}\tilde{B}',$$

implying a number of correlation parameters to be estimated equal to $2n$, which can be already prohibitive in moderately large systems.

The restriction to be imposed for the existence of $\mathbb{E}(Q_t)$ is that the eigenvalues of the diagonal matrix $A \otimes A + B \otimes B$ be less than 1 in modulus. Necessary and sufficient condition for this to hold are $\alpha_{ii} + \beta_{ii} < 1$, which also assures that $I_n - \tilde{A}\tilde{A}' - \tilde{B}\tilde{B}'$ is positive definite.

Hadamard specification Define by $\text{math}(\cdot)$ the operator that transforms a $n^* = n(n+1)/2$ dimensional vector into a $n \times n$ symmetric matrix of parameters. Then, given α and β $n^* \times 1$ vectors of parameters, we can rewrite $\{A, B\}$ as $A := \text{math}(\alpha)$

and $B := \text{math}(\beta)$, such that the Hadamard representation of Eq.(2.7) is expressed as

$$Q_t = (\iota'\iota - \text{math}(\alpha) - \text{math}(\beta)) + \text{math}(\alpha) \odot C_{t-1}^Q + \text{math}(\beta) \odot Q_{t-1},$$

where ι is the unit vector of dimension $1 \times n$ and $\odot$ denotes element by element product. We only consider the case where $\{A, B\}$ are full rank matrices. Sufficient condition on the parameters α and β ensuring that $\mathbb{E}(Q_t)$ is finite are $|\alpha_i + \beta_i| < 1$ for $i = 1, \dots, n$ while positivity of Q_t is assured by constraining $\iota'\iota - \text{math}(\alpha) - \text{math}(\beta)$ to be positive definite. Despite the advantage of having correlation dynamics that differs for all pairs of assets, this specification exhibits a quadratic dependence on the dimension n of the process with $n(n+1)$ parameters to be estimated. Therefore, to avoid parameter proliferation, its applicability is limited to low dimensional systems.

2.2.4 A dynamic attenuation approach

Given the recent contributions to the literature by [Bollerslev et al. \(2016a\)](#) and [Bollerslev et al. \(2016b\)](#), we find opportune to briefly discuss about the possibility to include some extra features in our models. At the same time, we also mention pros and cons of pursuing such approach.

The starting point is represented by the theoretical developments in [Barndorff-Nielsen & Shephard \(2004\)](#) and [Barndorff-Nielsen et al. \(2011\)](#), already introduced in Chapter 1. According to them, estimated realized covariance matrices can be decomposed into the sum of two components: the latent true covariance matrix and a measurement error term. The latter quantity is usually present in finite samples and may be more sizeable on some days and less on others, with a direct consequence on the accuracy of the realized covariance measure. Correctly estimating the magnitude of the noise is not a trivial task especially in multivariate frameworks, where the error term is not a scalar and knowledge of the exact form of its variance-covariance matrix is required. [Barndorff-Nielsen & Shephard \(2004\)](#) show that for the Realized Covariance estimator, as well as for some other consistent estimators (e.g. the Multivariate Kernel of [Barndorff-Nielsen et al. \(2011\)](#)), the covariance matrix of the error term, let us call it Π_t , is time-varying and proportional to the *integrated quarticity*, which can be estimated consistently using the same high-frequency data.

Recently, [Bollerslev et al. \(2016a\)](#) introduced a simple and intuitive way to account for the heteroskedasticity of the errors when modeling series of realized volatilities. The intuition behind their approach is that, on days when the variance of the measurement error is small, the daily realized variance estimator provides a stronger signal for next day's volatility than on days when the variance is large. By explicitly incorporating this feature into the parameterization of the model it is possible to dynamically attenuate the parameters, and in turn improve on the forecasts. A similar reasoning holds in the multivariate setting.

The Re-cDCC model does not account for the estimation error in the series of realized covariance matrices. An improved formulation could be achieved by including

those effects in the volatilities and correlation equations, where in the latter case an appropriate transformation of Π_t , e.g. $\Gamma_t = \{\Pi_t\}^{-1/2}\Pi_t\{\Pi_t\}^{-1/2}$, should be implemented to account for the measurement errors in the correlation dynamics. Assuming the simplified scalar case, Equations (2.5) and (2.7) become as follows:

$$H_{ii,t} = (1 - \gamma_i - \delta_{i,t}) + \gamma_i C_{ii,t-1}^* + \delta_{i,t} H_{ii,t-1}, \quad (2.13)$$

$$\delta_{i,t} = \delta_i + \delta_{i,Q} \Pi_{ii,t-1}, \quad (2.14)$$

$$Q_t = (1 - \alpha - \beta_t) + \alpha C_{t-1}^Q + \beta_t \odot Q_{t-1}, \quad (2.15)$$

$$\beta_t = \beta J_n + \beta_Q \Gamma_{t-1}, \quad (2.16)$$

with $\delta_{i,Q}$ and β_Q scalar parameters, J_n the n -dimensional matrix of ones and $\odot$ the Hadamard product.

Equations (2.14) and (2.16) allow the dynamics in each conditional variance series and each of the individual elements in the pseudo-correlation matrix Q_t to depend on their own measurement error variances. Notice that negative values of $\delta_{i,Q}$ and β_Q assign lower weight to those that are measured with greater error.

Clearly, applying this *measurement error-correction* is simpler with a handful of assets but may easily become difficult in larger dimensions. Not only because it requires a higher number of parameters to be estimated (one for each univariate series plus one for the correlation matrix as a whole), but also because the properties of the dynamic evolution of the measurement error for the correlations are not well known in this case.

2.3 Multistep estimation approach

For a Re-cDCC model in the class defined in Section 2.2, the vector of parameters to be estimated is denoted by ϕ . It can be partitioned into $(\nu, \mathbf{n}'_M, \phi'_v, \phi'_c)'$, where ν is the degrees of freedom parameter of the Wishart distribution, $\mathbf{n}_M$ is the vector of unique parameters included in the constant long-term matrix M (with $n(n+1)/2$ elements), and ϕ_v, ϕ_c are the vectors of conditional variance and correlation parameters, respectively. The dimension of ϕ_v is $2n$ if the GARCH-type formulation of equation Eq.(2.5) is adopted, while that of ϕ_c depends on the specification of the Q_t process. Estimation of all the components of ϕ , if the cross-sectional dimension is relatively small, can be easily performed by full QML, *i.e.* by maximizing the sample log-likelihood function with respect to all the parameters in one step. In larger settings, this can become prohibitive, thus requiring an *ad hoc* approach. Ours is fully described in the following subsections.

2.3.1 Covariance targeting

The possibility of covariance targeting is important for the estimation of the proposed model. To understand its usefulness, consider that the Re-cDCC model, even in its scalar specification, requires the estimation of $\mathbf{n}_M$, which has a dimension proportional to the square of the number of assets. This feature can render the estimation of all the parameters, in a maximum likelihood framework, prohibitive for large values of n . Hence, it is useful to be able to estimate $\mathbf{n}_M$ in a simple and quick way, before estimating the remaining parameters in a maximum likelihood framework. For this, we resort to *targeting*, which consists in estimating the constant matrix M by a method of moment estimator, without having to estimate the other parameters. After this first step, the estimate of M is substituted for the corresponding parameter matrix in the likelihood function to be maximized in the next steps. Since the dynamic equations of variances and correlations both depend on M through C_t^* , see Eq.(2.6) and Eq.(2.9), it is desirable that the targeting is consistent, even if it is inefficient. In other words, since the QML estimators of the remaining parameters depend on the targeting estimator, they cannot be consistent if the targeting estimator is not consistent; see Pakel et al. (2014) and Pedersen & Rahbek (2014) among others.

For the Re-cDCC model, under conditions that ensure the existence of second-order moments of C_t , consistent targeting of the matrix M is straightforwardly achieved by computing

$$\hat{M} = \frac{1}{T} \sum_{t=1}^T C_t,$$

which amounts to estimate the unconditional covariance matrix of the daily returns by the sample average of the realized covariance matrices up to time T . It is worth to mention that Ledoit et al. (2003) and Fan et al. (2012), among others, noticed that when the number of assets n is a significant share of T , the sample covariance matrix, subject to large estimation errors, becomes a poor estimator of M as well as ill-conditioned. In these cases, they advice using some improved estimators built under varying assumptions that are all suitable for financial applications. They are developed in a different framework (i.e. for MGARCH models), but it could be worth trying to rationalize their use also in the realized covariance framework. For future work, it would be interesting to compare these different approaches empirically.

Overall, applying the targeting has three important implications. Firstly, $\hat{L}$, such that $\hat{L}\hat{L}' = \hat{M}$, can be substituted for L in the computation of the likelihood function without impairing the possibility of consistent estimation of the parameters of the dynamic equations of conditional variances and correlations. Secondly, along with parsimonious parameterizations of the volatility and correlation processes, it makes the model applicable for large cross-sections. Thirdly, it enables a three-step estimation approach that crucially reduces the computational burden and speeds up estimation.

In order to assess the impact of using targeting along with the three-step approach for scalar Re-cDCC models, in Section 2.4 we perform a simulation study to assess

the bias and efficiency properties of the estimator of the parameters. These properties are compared to those achieved by one-step estimation, namely by maximizing the likelihood function with respect to all the parameters jointly.

2.3.2 QML estimation

In $\phi = (\nu, \mathbf{n}'_M, \phi'_v, \phi'_c)'$, we also partition ϕ_v into $\phi_v = (\phi_v^{(1)'}, \phi_v^{(2)'}, \dots, \phi_v^{(n)'})'$, where $\phi_v^{(i)}$ is the parameter vector of the conditional variance equation specific to asset i . Using the expression of a Wishart density function, and of S_t in Eq.(2.3), the log-likelihood contribution of observation t , $\ell(C_t; \phi | \mathfrak{S}_{t-1}) := \ell_t(\phi)$, is given by

$$\begin{aligned} \ell_t(\phi) = & \frac{\nu n}{2} \log \frac{\nu}{2} + \frac{\nu - n - 1}{2} \log |C_t| - \sum_{i=1}^n \log \Gamma[(\nu + 1 - i)/2] + \\ & - \frac{\nu}{2} \log |LH_t(\phi)L'| - \frac{\nu}{2} \text{tr}\{(LH_t(\phi)L')^{-1}C_t\}, \end{aligned} \quad (2.17)$$

where the dependence of H_t on ϕ is made explicit. When targeting is used, meaning that $M = LL'$ is estimated consistently by $\hat{M} = 1/T \sum_{t=1}^T C_t$, then $\hat{L}$ is substituted for L in Eq.(2.17), otherwise, L is an unknown set of parameters that is estimated jointly with the remaining parameters (the non-zero elements of L form a set of parameters which is equivalent to $\mathbf{n}_M$).

Given our interest in applying the model to high-dimensional systems, we consider targeting as an essential part of the estimation procedure, as otherwise estimation would be not feasible. For ease of notation, in the sequel we use L in the expressions of the log-likelihood function instead of $\hat{L}$. The next proposition (proven in Appendix 2.6) provides the basis of a three-step estimation procedure.

Proposition 1 *The likelihood contribution $\ell_t(\phi)$ in Eq.(2.17) can be written as*

$$\ell_t(\phi) = \ell_{0,t}(\nu) + \ell_{v,t}(\nu, \phi_v; \hat{\mathbf{n}}_M) + \ell_{c,t}(\nu, \phi_c; \hat{\phi}_v, \hat{\mathbf{n}}_M), \quad (2.18)$$

where

$$\begin{aligned} \ell_{0,t}(\nu) = & \frac{\nu n}{2} \log \frac{\nu}{2} + \frac{\nu - n - 1}{2} \log |C_t| - \sum_{i=1}^n \log [\Gamma(\nu + 1 - i)/2], \\ \ell_{v,t}(\nu, \phi_v; \hat{\mathbf{n}}_M) = & -\nu \log(D_t) - \frac{\nu}{2} \text{tr}\{D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1}\}, \end{aligned} \quad (2.19)$$

$$\begin{aligned} \ell_{c,t}(\nu, \phi_c; \hat{\phi}_v, \hat{\mathbf{n}}_M) = & -\frac{\nu}{2} \log |R_t| - \nu \log |L| + \\ & - \frac{\nu}{2} \text{tr}\{(R_t^{-1} - I_n) D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1}\}. \end{aligned} \quad (2.20)$$

If the univariate equations for the conditional variances do not include spillover terms other than their own lagged conditional variance and their parameters are variation-free, $\ell_{v,t}$ can be written as the sum of n univariate functions:

$$\ell_{v,t}(\nu, \phi_v; \hat{\mathbf{n}}_M) = \frac{\nu}{2} \left[- \sum_{i=1}^n \log H_{ii,t} - \sum_{i=1}^n H_{ii,t}^{-1} C_{ii,t}^* \right] = \frac{\nu}{2} \sum_{i=1}^n \ell_{v,t}^{(i)}(\phi_v^{(i)}), \quad (2.21)$$

with $C_{ii,t}^*$ denoting the i -th diagonal element of the matrix $L^{-1}C_t L'^{-1}$.

The above proposition inspires the following comments:

1. The $\ell_{v,t}$ part of the log-likelihood is proportional to the shape parameter ν . This implies that it can be maximized with respect to the elements of ϕ_v without estimating ν .
2. Each function $\ell_{v,t}^{(i)}$, defined as the terms between square brackets in Eq.(2.21), only depends on the parameters $\phi_v^{(i)}$ specific to the conditional variance equation of asset i . It follows that maximization of $\ell_{v,t}$ can be achieved through n separate optimizations (under the assumptions stated in Proposition 1), which simplifies estimation. Notice that $\ell_{v,t}^{(i)}$ corresponds to the log-likelihood of a gamma distribution.
3. The $\ell_{c,t}$ part of the log-likelihood depends on the whole set of parameters in ϕ . Since it is linear in ν , it can be maximized with respect to these parameters without estimating ν .
4. As our primary interest is in the estimation of the parameters of the conditional variances and correlation processes, we consider the degree of freedom parameter ν as a nuisance parameter, which is moreover justified by the fact that the other parts of the log-likelihood depend linearly on ν . Consequently, the $\ell_{0,t}(\nu)$ part of the likelihood can be ignored.

The main advantage of these results is that we can adopt a three-step procedure to estimate the model parameters:

1. In step 1, the vector of unique parameters $\mathbf{n}_M$ of the constant matrix M are consistently estimated by a method of moment estimator and the resulting estimator $\hat{\mathbf{n}}_M$ is substituted for the corresponding parameter in the likelihood function.
2. In step 2, conditional on $\hat{\mathbf{n}}_M$, the conditional variance parameters are estimated as

$$\hat{\phi}_v^{(i)} = \operatorname{argmax}_{\phi_v^{(i)}} \sum_{t=1}^T \ell_{v,t}^{(i)}, \text{ for } i = 1, \dots, n.$$

3. In step 3, the parameters of the correlation equation are estimated as

$$\hat{\phi}_c = \operatorname{argmax}_{\phi_c} \sum_{t=1}^T \ell_{c,t}$$

conditional on $\hat{\mathbf{n}}_M$ and $\hat{\phi}_v$ obtained in the previous steps.

In order to provide inferential properties for the three-step procedure, we observe that it can be represented as a three-step moment-estimator. In the following proposition, proven in Appendix 2.6, we derive analytical expressions for the underlying moment conditions and show that their expectation is zero at the true parameter values.

Proposition 2 *The vector of moment conditions for the t -th observation is $\tilde{\mathbf{s}}_t = (\tilde{\mathbf{s}}'_{n,t}, \tilde{\mathbf{s}}'_{v,t}, \tilde{\mathbf{s}}'_{c,t})'$, with components:*

$$\tilde{\mathbf{s}}_{n,t} = \text{vech}(C_t) - \mathbf{n}_M \quad (2.22)$$

$$\tilde{\mathbf{s}}_{v,t} = \frac{\partial \ell_{v,t}(\nu, \phi_v; \hat{\mathbf{n}}_M)}{\partial \phi_v} = -\frac{\nu}{2} \frac{\partial \log |H_t|}{\partial \phi_v} - \frac{\nu}{2} \frac{\partial \text{tr}(L^{-1} C_t L'^{-1} H_t^{-1})}{\partial \phi_v} \quad (2.23)$$

$$\tilde{\mathbf{s}}_{c,t} = \frac{\partial \ell_{c,t}(\nu, \phi_c; \hat{\mathbf{n}}_M, \hat{\phi}_v)}{\partial \phi_c} = -\frac{\nu}{2} \frac{\partial \log |R_t|}{\partial \phi_c} - \frac{\nu}{2} \frac{\partial \text{tr}(R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1})}{\partial \phi_c} \quad (2.24)$$

At the true parameter values $\phi_0 = \{\mathbf{n}_{M,0}, \phi_{v,0}, \phi_{c,0}\}$, the expected value of $\tilde{\mathbf{s}}_t$ is equal to $\mathbf{0}$.

Under the stated result, the theory developed in Newey and McFadden (1994) can be applied to prove, at the price of the additional regularity conditions stated in their paper, the consistency and asymptotic normality of the three-step estimator. These hold even in cases in which the Wishart assumption is not satisfied, making our approach robust to misspecification of the distribution of the realized covariance matrices. It is worth noting that, if targeting is not used, the resulting estimator can be represented as a two step QML estimator (QMLE) with score given by $(\tilde{\mathbf{s}}'_{v,t}, \tilde{\mathbf{s}}'_{c,t})'$. In this case the estimation approach is similar to that applied, in different settings, by Engle & Russell (1998), Engle (2002b), and Engle & Gallo (2006) in the estimation of ACD and multiplicative error models with a gamma conditional distribution.

Concerning the asymptotic distribution of the estimator $\hat{\phi}$, under standard regularity conditions, it holds that

$$\sqrt{T}(\hat{\phi} - \phi_0) \xrightarrow{d} N(\mathbf{0}, l^{-1} \Xi (l^{-1})')$$

where $\Xi = \text{var} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \tilde{\mathbf{s}}_t \right]$ is estimated using a HAC estimator (e.g. Newey & West (1987)) and

$$l = -\mathbb{E} \left[\frac{\partial \tilde{\mathbf{s}}_t}{\partial \phi} \right] = -\mathbb{E} \begin{bmatrix} -I_{n^*} & \frac{\partial^2 \tilde{s}_v}{\partial n_M \partial \phi'_v} & \frac{\partial^2 \tilde{s}_c}{\partial n_M \partial \phi'_c} \\ 0 & \frac{\partial^2 \tilde{s}_v}{\partial \phi_v \partial \phi'_v} & \frac{\partial^2 \tilde{s}_c}{\partial \phi_v \partial \phi'_c} \\ 0 & 0 & \frac{\partial^2 \tilde{s}_c}{\partial \phi_c \partial \phi'_c} \end{bmatrix}, \quad (2.25)$$

with the Hessian having a block upper triangular structure. As the targeting is used,

the covariance matrix of the estimated parameters in ϕ_c has to include an adjustment to account for the efficiency loss due to the preliminary estimation of the parameters in ϕ_v and of the parameters in the constant term.

2.3.3 Third stage composite likelihood estimation

A possibly interesting option, which we outline here but do not pursue empirically, is to estimate ϕ_c by composite likelihood (CL). This approach, first introduced by Lindsay (1988) and later on popularized by Engle et al. (2008) and Pakel et al. (2014), consists, roughly speaking, in using an approximation of the joint marginal density based on lower dimensional marginal densities. The estimator that obtains leads to a loss in efficiency compared to a full quasi-likelihood approach but is shown to be potentially more robust to misspecification issues arising with high dimensional models. For this reason, the CL procedure has been extensively applied in the MGARCH literature for the scalar versions³ of the BEKK, DCC and cDCC models. In the latter case, Pakel et al. (2014) find that the CL allows for the estimation of models in dimension of many hundreds, being easy to implement and not effected as much as the full QML estimator by the incidental parameter problem.

In the realized co-volatility literature, Bollerslev et al. (2016b) use the CL approach with the aim to simplify estimation and reduce computational burden in large dimensional systems where multistep solutions are not feasible. Nevertheless, none of these papers carries out a formal investigation of the properties of realized covariance models fitted by CL. In our setting, where the likelihood function is already split in blocks, the CL could be applied to the third estimation stage in order to avoid inverting many times matrices of large dimensions. In fact, when many hundreds of assets are involved, $\ell_{c,t}(\nu, \phi_c; \hat{\phi}_v, \hat{n}_M)$ can easily result in numerically unstable R_t^{-1} matrices.

Luckily, it turns out that for the scalar Re-cDCC – and Re-cDECO – model, the Wishart assumption also enables us to use the CL method. This can be briefly described as follows.

For any square matrix C_t of order n , let $C_{AA,t}$ denote a square matrix of order n_A , extracted from C_t , which has its main diagonal elements on the main diagonal of C_t . Namely, if A stands for a subset of n_A different indexes of $\{1, 2, \dots, n\}$, $C_{AA,t}$ is the matrix that consists of the intersection of the rows and columns of C_t corresponding to the selection of indexes denoted by A .

The method relies on the following three results:

- R1: If $C_t \sim W_n(\nu, S_t/\nu)$, $C_{AA,t} \sim W_{n_A}(\nu, S_{AA,t}/\nu)$ for any selection of n_A indexes.
- R2: If $H_t = D_t R_t D_t$, $H_{AA,t} = D_{AA,t} R_{AA,t} D_{AA,t}$.
- R3: If $Q_t = (1 - \alpha - \beta) + \alpha C_{t-1}^Q + \beta Q_{t-1}$, then $Q_{AA,t} = (1 - \alpha - \beta) + \alpha C_{AA,t-1}^Q + \beta Q_{AA,t-1}$.

³There has not been, so far, formal evidence of the theoretical validity of applying CL estimation to non-scalar models. A first attempt in this direction has recently been made by Bauwens et al. (2015).

Result 1 is a property of the Wishart distribution, see [Lutkepohl \(1996\)](#). Notice that if applied with $n_A = 1$, it follows that the marginal distribution of a diagonal element of a Wishart matrix is a gamma, a result that is used to form the first step log-likelihood function $\ell_{v,t}^{(i)}$. Results 2 and 3, given that D_t is diagonal and L is lower triangular, are obvious algebraic derivations. A CL third step estimator of the parameters α and β is then defined as the maximizer of the sum of a number of Wishart marginal log-likelihoods of n_A dimensional sub-matrices corresponding to different choices of indices A . The most obvious choice is to select all unique sub-matrices of order 2, *i.e.* all the $n(n-1)/2$ correlation coefficients or pairs of assets. In this bivariate Wishart log-likelihoods, only matrices of order 2 must be inverted, and this can be efficiently programmed. Such a CL for the contribution of observation t is formed by averaging across all pairs in the sample:

$$CL_t(\nu, \phi_c; \hat{\phi}_v, \hat{n}_M) = \frac{1}{n} \sum_{h=1}^n \sum_{k < h} \ell_{hk,t} \left(\nu, \phi_c; \hat{\phi}_v^{(h)}, \hat{\phi}_v^{(k)}, \hat{n}_M^{(hk)} \right).$$

In this case the CL costs $O(n^2)$ calculations, while other choices, such as when subset of all (or some) contiguous pairs are used, can reduce the cost to $O(n(n-1)/2)$. Hence, depending on the particular sampling strategy adopted, computational gains can be rather substantial.

2.4 Simulation study

The three-step estimation approach exposed in [Section 2.3](#) is easy to implement and provides consistent parameter estimates, under appropriate conditions. However, not much is known about its behavior in finite samples, nor how it compares to the full QML estimator, hereafter denoted 1QML for brevity. We do not include ν , considered again as a nuisance parameter. A clear-cut advantage of the three-step estimator (3QMLE) over

Table 2.1: **Simulation design for the scalar Re-cDCC process**

n		2, 5, 7, 10, 15, 25, 50, 100
M	in Eq. (2.2)	$M_{i,i} = 0.1$, $M_{i,j} = 0.02$ for $i \neq j$
γ_i	in Eq. (2.5)	$\gamma_i = 0.05$, for $i = 1, \dots, n$
δ_i	in Eq. (2.5)	$\delta_i = 0.9$, for $i = 1, \dots, n$
α	in Eq. (2.12)	0.1
β	in Eq. (2.12)	0.8
ν	in Eq. (2.1)	$2n$
T		500, 1000, 2000 (+ 500 burn-in)

1QMLE is that it reduces the computational burden arising in the optimization process by splitting it in several steps. However, in medium-low cross-sectional dimensions up to ten-fifteen assets, the 1QMLE is feasible and replacing it with 3QMLE may incur a loss of efficiency. Moreover, it is also possible to define a two-step estimator (2QMLE) that adds to the targeting step a second, single, step estimator of (ϕ_c, ϕ_v) obtained

by maximizing the sample average of $\ell_{v,t}(\nu, \phi_v; \hat{\mathbf{n}}_M) + \ell_{c,t}(\nu, \phi_c; \hat{\phi}_v, \hat{\mathbf{n}}_M)$. The 2QMLE does not separate the estimation of the variance and correlation parameters and due to this it may incur a smaller efficiency loss than 3QMLE with respect to 1QMLE, the difference between 2QMLE and 1QMLE being due solely to the targeting step.

In order to compare the three estimators, we report the results of a Monte Carlo experiment, designed to compare the finite-sample bias and precision of 3QMLE, 2QMLE and 1QMLE in dimensions up to fifteen (since 1QMLE is too costly in larger dimensions). In a nutshell, we find that 3QMLE has as good properties as the other two estimators in very low dimensions ($n \leq 5$), and has better properties as soon as n exceeds 7. We also report results for the 3QMLE in dimensions up to $n = 100$.

Specifically, we assume as data generating process (DGP) a scalar Re-cDCC process, from which we simulate $K = 500$ series according to the design summarized in Table 2.1. To cover a variety of cases, eight different values of the cross-sectional dimension and three time series lengths are considered. In all cases, the estimated model corresponds to the model class to which the DGP belongs, hence it is correctly specified. The statistical properties of the estimators are evaluated in terms of relative bias (RB) and root mean squared error (RMSE) defined as

$$RB(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^K \frac{(\hat{\theta}_k - \theta_0)}{\theta_0}, \quad RMSE(\hat{\theta}) = \sqrt{\frac{1}{K} \sum_{k=1}^K (\hat{\theta}_k - \theta_0)^2},$$

for the estimator $\hat{\theta}$ of the parameter θ (an element of ϕ or M) having the true value θ_0 and estimated by $\hat{\theta}_k$ for the k -th simulated series. To synthesize results, the statistics for the elements of the matrix M are averaged across parameters: namely, at each k , RB and RMSE are averaged across diagonal (variance) and off diagonal (covariance) elements. Similarly, the results for the parameters of the univariate variance equations are reported as averages across assets.

Let first comment the finite-sample properties of the 3QMLE, reported in Table 2.2 and of the targeting estimator, reported in Table 2.3. A few conclusions can be drawn in terms of bias:

1. The average parameters of the conditional variance equations, i.e. $\bar{\gamma}$ and $\bar{\delta}$, achieve the highest biases across the panels. The bias on $\bar{\gamma}$ is always positive and ranges from the highest value of 13.1% (for $n = 5, T = 500$) to the lowest value of 0.4% (for $n = 10, T = 2000$). The bias on $\bar{\delta}$ is negative and ranges between -14.3% (for $n = 5, T = 500$) and -1% (for $n = 10, T = 2000$). The relative bias of $\bar{\gamma}$ and $\bar{\delta}$ decreases when T increases, while it remains stable with respect to n . However, it is worth to stress that the conditional variance series are estimated equation by equation and the reported estimates are averaged across a number of parameters that increases with n . The average stability of these estimates across n is an encouraging indication that, despite the increasing cross-sectional size, the first step moment targeting is not negatively affecting second stage estimation.
2. For α , the bias is negative, except in two cases, and anyway small, being under one

percent in absolute value with only one exception occurring at $n = 2, T = 500$, where $RB(\alpha) = 3\%$. The bias shrinks toward zero as T increases. This is more visible for $n \geq 7$ than for $n \leq 7$, probably because there is more information for large n about a parameter that is common to all correlation processes. There is perhaps a slight decreasing trend in the bias as n increases from $n = 2$ ($n = 5$ for $T = 500$), though the bias becomes essentially stable for $n \geq 25$.

3. The bias for β is always negative and in the majority of cases very small, ranging between the minimum of -0.3% for $n = 100, T = 2000$ and the maximum of -5.3% for $n = 2, T = 500$, the latter caused by a few outlying estimates (the lowest achieved value is almost zero). As a function of T , for each cross-sectional dimension, the relative bias is divided by about two when T is doubled. As a function of n , the relative bias decreases strongly between $n = 2$ and $n = 5$, then very little, and is stable for $n \geq 15$.
4. For the parameters of the matrix M , the average relative bias for the variances is less than 0.1% in absolute value, and less than 1% for the covariances. They are decreasing with n , this being visible more quickly for $T = 1000$ and 2000 than for $T = 500$.

Table 2.2 shows that the RMSEs of the GARCH and cDCC equations decrease when n and T increase. This is not surprising as an increase in n implies that there is more information to use, while a larger number of observations leads to more efficient estimators. The values of the RMSE are larger for $\bar{\delta}$ and β than for $\bar{\gamma}$ and α , reflecting a larger dispersion of the estimates of the former through the Monte Carlo replications. The same downward trends are observed in the RMSEs reported in Table 2.3, for the variance and covariance terms of the matrix M .

Next, we analyze the relative performance of the three estimation approaches in medium-low dimensional settings. Table 2.4 presents Monte Carlo results up to dimension 15, where entries are pairwise ratios of the RB (in absolute value) and RMSE statistics achieved by each estimator. Values smaller than unity indicate that the estimator at the numerator is less biased or more precise. Conceivably, we expect the differences between the three estimators to be minimal when only two assets are considered, and to increase with the number of series in the cross-section. The use of the 2QMLE, which can be considered as an intermediate approach between the other two, serves the purpose to disentangle the effects of targeting and multi-step estimation. Of course, 2QMLE and 3QMLE do not differ in the estimation of the constant matrix M which is done by targeting in the same way for both estimators.

The most noticeable finding emerges from the comparison of the biases of the estimators of the univariate GARCH parameters γ and δ . The table reveals that the 2QMLE is generally very close to the 3QMLE, while the 1QMLE estimates appear much more biased as n increases. For example, when $n=15$ and $T=2000$, the average biases on γ and δ estimated via 3QML are respectively only 1.4% and 11.8% of those achieved via 1QML, and 63% and 91% of those obtained via 2QML. Similarly, the RMSE ratios show that the performance of 1QMLE compared to the others deteriorates when n increases.

Table 2.2: Simulation results for the 3-step QMLE of the GARCH and cDCC parameters

The table reports the relative bias (RB) and root mean squared error (RMSE) of the 3-step QML estimators of the GARCH and cDCC parameters for $T=\{500, 1000, 2000\}$ and $n=\{2, 5, 7, 10, 15, 25, 50, 100\}$. The univariate GARCH parameters $\bar{\gamma}$ and $\bar{\delta}$ are averaged across n . DGP parameter values are listed in the top line of the table. Reported values are averaged across 500 simulations.

	T = 500				T = 1000				T = 2000			
	$\bar{\gamma} = 0.05$	$\bar{\delta} = 0.9$	$\alpha = 0.1$	$\beta = 0.8$	$\bar{\gamma} = 0.05$	$\bar{\delta} = 0.9$	$\alpha = 0.1$	$\beta = 0.8$	$\bar{\gamma} = 0.10$	$\bar{\delta} = 0.8$	$\alpha = 0.1$	$\beta = 0.8$
	n=2				n=2				n=2			
RB	0.130	-0.114	0.005	-0.055	0.032	-0.040	0.005	-0.023	0.013	-0.011	-0.002	-0.008
RMSE	0.019	0.182	0.032	0.138	0.013	0.099	0.024	0.082	0.009	0.028	0.018	0.043
	n=5				n=5				n=5			
RB	0.131	-0.143	-0.008	-0.016	0.052	-0.048	-0.003	-0.008	0.014	-0.013	-0.001	-0.004
RMSE	0.014	0.171	0.010	0.028	0.009	0.073	0.008	0.019	0.006	0.023	0.005	0.012
	n=7				n=7				n=7			
RB	0.125	-0.135	-0.004	-0.015	0.051	-0.047	-0.004	-0.007	0.013	-0.012	-0.001	-0.004
RMSE	0.012	0.151	0.007	0.021	0.007	0.067	0.005	0.013	0.005	0.020	0.003	0.009
	n=10				n=10				n=10			
RB	0.112	-0.138	-0.009	-0.014	0.042	-0.047	-0.000	-0.007	0.004	-0.010	-0.001	-0.003
RMSE	0.010	0.148	0.005	0.016	0.006	0.060	0.004	0.011	0.004	0.016	0.002	0.006
	n=15				n=15				n=15			
RB	0.109	-0.123	-0.004	-0.014	0.049	-0.047	-0.003	-0.006	0.012	-0.013	-0.002	-0.003
RMSE	0.009	0.124	0.003	0.014	0.005	0.056	0.002	0.007	0.003	0.018	0.001	0.004
	n=25				n=25				n=25			
RB	0.108	-0.122	-0.007	-0.013	0.044	-0.046	-0.002	-0.006	0.011	-0.013	-0.001	-0.003
RMSE	0.007	0.119	0.002	0.011	0.004	0.049	0.001	0.006	0.002	0.016	0.000	0.003
	n=50				n=50				n=50			
RB	0.108	-0.136	-0.007	-0.012	0.044	-0.045	-0.002	-0.006	0.012	-0.013	-0.001	-0.003
RMSE	0.007	0.127	0.002	0.010	0.003	0.045	0.000	0.005	0.052	0.870	0.100	0.803
	n=100				n=100				n=100			
RB	0.126	-0.128	-0.006	-0.013	0.044	-0.047	-0.002	-0.006	0.012	-0.012	-0.001	-0.003
RMSE	0.006	0.106	0.000	0.010	0.002	0.045	0.000	0.005	0.001	0.013	0.000	0.002

Table 2.3: Simulation results for the targeted scale matrix M

The table reports the relative bias (RB) and root mean squared error (RMSE) of the estimated targeted scale matrix M for $T=500, 1000, 2000$ and $n=2, 5, 7, 10, 15, 25, 50, 100$. $RB_{\{i,i\}}$ and $RMSE_{\{i,i\}}$ denote averaged values over diagonal terms; $RB_{\{i,j\}}$ and $RMSE_{\{i,j\}}$ denote averages over off diagonal terms. The DGP values are such that $M_{\{i,i\}}=0.1$ and $M_{\{i,j\}}=0.02$. Reported values are averaged across 500 simulations.

	T=500	T=1000	T=2000	T=500	T=1000	T=2000
	n=2			n=15		
$RB_{\{i,i\}}$	0.0006	-0.0007	-0.0005	-0.0002	-0.0001	1.20e-05
$RB_{\{i,j\}}$	0.0038	-0.0039	-0.0016	0.0009	-0.0022	-0.0007
$RMSE_{\{i,i\}}$	0.0060	0.0045	0.0031	0.0023	0.0016	0.0012
$RMSE_{\{i,j\}}$	0.0042	0.0033	0.0023	0.0017	0.0012	0.0008
	n=5			n=25		
$RB_{\{i,i\}}$	0.0008	-0.0005	-0.0002	0.0001	4.50e-05	-5.97e-05
$RB_{\{i,j\}}$	0.0069	0.0036	0.0005	0.0003	-0.0006	-0.0002
$RMSE_{\{i,i\}}$	0.0040	0.0029	0.0020	0.0018	0.0013	0.0008
$RMSE_{\{i,j\}}$	0.0029	0.0020	0.0015	0.0013	0.0009	0.0006
	n=7			n=50		
$RB_{\{i,i\}}$	-0.0002	-0.0005	-0.0002	-4.88e-05	-2.76e-05	-4.48e-05
$RB_{\{i,j\}}$	-0.0020	-0.0027	0.0016	-0.0001	0.0002	-8.63e-05
$RMSE_{\{i,i\}}$	0.0034	0.0024	0.0017	0.0012	0.0008	0.0006
$RMSE_{\{i,j\}}$	0.0024	0.0017	0.0012	0.0009	0.0006	0.0004
	n=10			n=100		
$RB_{\{i,i\}}$	0.0002	-0.0004	-0.0002	-7.16e-05	-8.71e-05	-6.13e-05
$RB_{\{i,j\}}$	0.0014	-0.0020	-0.0003	-0.0002	-0.0002	-0.0002
$RMSE_{\{i,i\}}$	0.0028	0.0020	0.0014	0.0008	0.0006	0.0004
$RMSE_{\{i,j\}}$	0.0020	0.0014	0.0010	0.0006	0.0004	0.0003

This is spectacular as soon as n exceeds 5. The effect of this distortion is also visible

Table 2.4: Monte Carlo comparison results: GARCH and cDCC parameters

The table reports ratios of RB and RMSE values for 1QML, 2QML, 3QML estimators. RB and RMSE are computed as described in Table 2.2.

n		T=500				T=1000				T=2000				
		$\bar{\gamma}$	$\bar{\delta}$	α	β	$\bar{\gamma}$	$\bar{\delta}$	α	β	$\bar{\gamma}$	$\bar{\delta}$	α	β	
2	RB	2QML/1QML	1.076	1.061	0.721	1.001	1.131	1.118	0.696	0.967	1.043	1.022	1.496	0.991
		3QML/1QML	1.028	1.034	0.433	0.982	1.048	1.096	0.687	1.038	0.936	1.019	2.240	0.966
		3QML/2QML	0.956	0.975	0.600	0.981	0.927	0.980	0.988	1.073	0.898	0.997	1.497	0.975
	RMSE	2QML/1QML	1.014	1.048	0.992	1.000	0.994	1.084	0.996	0.989	0.997	1.005	0.999	1.000
		3QML/1QML	1.007	1.040	0.994	0.999	1.007	1.067	0.998	1.107	1.009	1.020	0.998	0.998
		3QML/2QML	0.993	0.992	1.002	1.000	1.013	0.984	1.002	1.119	1.012	1.015	0.999	0.998
5	RB	2QML/1QML	0.997	0.985	1.436	0.980	1.012	1.004	1.688	0.965	0.961	0.980	2.435	0.966
		3QML/1QML	0.926	0.964	1.654	0.961	0.973	1.018	1.986	0.947	1.005	1.049	3.794	0.951
		3QML/2QML	0.929	0.979	1.152	0.980	0.962	1.014	1.177	0.981	1.046	1.071	1.558	0.985
	RMSE	2QML/1QML	0.996	0.988	0.998	0.994	1.004	1.019	1.000	0.996	0.994	0.975	0.999	0.997
		3QML/1QML	0.980	0.996	0.998	0.989	1.006	1.023	1.000	0.993	1.015	1.042	0.999	0.996
		3QML/2QML	0.985	1.008	1.000	0.995	1.002	1.004	1.000	0.998	1.021	1.069	1.000	0.999
7	RB	2QML/1QML	0.247	0.475	0.315	0.787	0.142	0.324	0.357	0.756	0.048	0.174	0.015	0.673
		3QML/1QML	0.231	0.443	0.380	0.779	0.153	0.345	0.420	0.748	0.042	0.173	0.013	0.665
		3QML/2QML	0.936	0.934	1.206	0.989	1.077	1.062	1.176	0.989	0.885	0.995	0.894	0.988
	RMSE	2QML/1QML	0.400	0.547	0.944	0.881	0.397	0.464	0.969	0.912	0.348	0.311	0.968	0.919
		3QML/1QML	0.401	0.514	0.945	0.878	0.405	0.482	0.969	0.910	0.357	0.304	0.968	0.918
		3QML/2QML	1.002	0.940	1.002	0.997	1.019	1.038	1.000	0.998	1.025	0.980	1.000	0.999
10	RB	2QML/1QML	0.146	0.491	0.690	0.746	0.076	0.280	0.008	0.720	0.005	0.078	0.089	0.655
		3QML/1QML	0.128	0.441	0.727	0.738	0.071	0.285	0.070	0.712	0.090	0.090	0.111	0.650
		3QML/2QML	0.876	0.899	1.054	0.989	0.924	1.017	0.777	0.989	1.512	1.149	1.246	0.991
	RMSE	2QML/1QML	0.257	0.565	0.940	0.849	0.219	0.405	0.916	0.874	0.150	0.140	0.874	0.870
		3QML/1QML	0.243	0.521	0.940	0.844	0.212	0.394	0.915	0.870	0.155	0.158	0.874	0.869
		3QML/2QML	0.944	0.921	1.000	0.994	0.968	0.973	0.999	0.995	1.037	1.130	1.000	0.998
15	RB	2QML/1QML	0.174	1.148	0.419	0.864	0.080	0.463	0.303	0.886	0.023	0.129	0.032	0.780
		3QML/1QML	0.124	0.857	0.451	0.853	0.055	0.400	0.339	0.872	0.014	0.118	0.051	0.771
		3QML/2QML	0.709	0.789	1.076	0.987	0.691	0.864	1.119	0.984	0.629	0.912	1.591	0.989
	RMSE	2QML/1QML	0.233	1.262	0.617	0.609	0.143	0.599	0.910	0.934	0.072	0.194	0.880	0.899
		3QML/1QML	0.198	1.027	0.617	0.603	0.119	0.523	0.911	0.926	0.071	0.180	0.881	0.895
		3QML/2QML	0.850	0.814	1.001	0.991	0.830	0.873	1.001	0.992	0.989	0.930	1.000	0.996

Table 2.5: Monte Carlo comparison results: M matrix

The table reports ratios of Frobenius norms of the 1QML estimator of M and of the targeting estimator of M (used for 2QML and 3QML estimations). Frobenius norm defined as $FN(\hat{M}) = \sqrt{\sum_{i,j=1}^n |M_{i,j} - \hat{M}_{i,j}|^2}$, where $M_{i,j}$ denotes the true value (see DGP in Table 2.1) and $\hat{M}_{i,j}$ is the element of the estimated M matrix, obtained either by 1-step estimation (1QML) or by the targeting approach. Values below (above) 1 are in favour of 1QMLE (targeting). Results based on 500 simulations.

n	T = 500	T = 1000	T = 2000
2	0.988	0.946	0.986
5	0.992	0.971	0.967
7	0.999	0.990	0.993
10	1.198	1.072	1.025
15	1.787	1.408	1.249

for the correlation parameters, but much lessened.

Our explanation of these results is that the 1QMLEs of the GARCH and cDCC

parameter are affected by the increasing number of model parameters due to the dimension of the matrix M : for example, when $n = 15$ the 1QML optimizes the likelihood function over 152 parameters (120 for M , 32 for the rest), while 2QML optimizes over 32 parameters, and 3QML achieves this by steps involving two parameters at a time after targeting. As the ratios reported in Table 2.4 between the 2QMLE and 3QMLE are close to unity, it is clear that the main source of burden for 1QMLE comes from the estimation of the parameters of the constant matrix M . By looking at Table 2.5, it is easy to realize that while in the lowest dimensions (up to $n = 7$) the 1QMLE delivers slightly more accurate estimates of M than the targeting estimator used in 3QMLE and 2QMLE, from $n > 7$ there is a clear countertrend in favor of the targeting estimator.

What is more surprising from Tables 2.4 is that, for fixed n , the relative discrepancy does not decrease with larger T . By inspection of Table 2.13 in 2.6, it appears that RB and RMSE for 1QMLE decrease with T , but at a slower rate than for the competing estimators (at $n=15$, $RB(\bar{\gamma})$ implied by the 1QMLE is reduced by 5.8 percent from $T=500$ (0.933) to $T=1000$ (0.879) and increases slightly from $T=1000$ to $T=2000$ (0.883), as opposed to the $RB(\bar{\gamma})$ of the 3QMLE which is reduced by 61 percent in the first case (from 0.127 to 0.049) and by 73 percent in the second (from 0.049 to 0.013).

This suggests that some improvements in the 1QMLE relative performance could be achieved by increasing further the number of observations in the time series; this is in accordance with the strand of literature claiming that a multistep estimator can not be asymptotically more efficient than the standard one-step estimator; see Pedersen & Rahbek (2014) for the BEKK-GARCH model, Francq et al. (2014) in the CCC-GARCH case and Francq et al. (2011) for an overview in the univariate GARCH framework.

To summarize, the simulation exercise strongly supports the use of the proposed three-step estimator in practical applications, especially of large dimension. The general message is that, under the assumption of a scalar Re-cDCC DGP, the estimator has good finite sample properties, it works no worse than the two-step approach and it favourably compares to the standard one-step QMLE, even when both are feasible.

2.5 Empirical analysis

This section illustrates the results of two empirical applications on real data. First, we analyse the model's main features on a low dimensional dataset composed of two stocks⁴, Alcoa (AA) and Exxon Mobil (XOM), traded within the same sector in the NYSE. Then, we move the attention to more heterogeneous high dimensional frameworks and investigate the usefulness of the model for one-step ahead forecasting and minimum variance allocation purposes.

In both exercises, we consider the two following competing models:

⁴The same dataset has been used in the papers by Nouredin et al. (2012) and Nouredin et al. (2014), available at <http://realized.oxford-man.ox.ac.uk>.

- the Conditional Autoregressive Wishart (CAW) model of [Golosnoy et al. \(2012\)](#), where the daily covariance process is driven by a BEKK-type recursion. Assuming a full parametrization and only one lag it is written as $S_t = CC' + AC_{t-1}A' + BS_{t-1}B'$, with C lower triangular matrix and $\{A, B\}$ $n \times n$ matrices of scalar parameters;
- the Rotated DCC (RDCC) model of [Noureldin et al. \(2014\)](#) which belongs to the rotated ARCH class of models where parameters of interest are obtained by fitting a DCC parameterization of the time varying covariance to rotated daily returns having as long-run covariance an identity matrix.

In the second application we also consider the Re-cDECO model illustrated previously and an Exponentially Weighted Moving Average (EWMA) model directly applied to the series of realized covariance matrices, *i.e.* $S_t = (1 - \lambda)C_{t-1} + \lambda S_{t-1}$, with λ parameter set equal to the value 0.94.

2.5.1 Bivariate dataset AA-XOM

The dataset comprises the series of daily realized covariance matrices of the Alcoa (AA) and Exxon Mobil (XOM) stocks, traded on the NYSE, estimated from five minute intraday returns over the period February 1, 2001 to December 31, 2009. Descriptive statistics are shown in Table 2.6. For fitting the RDCC model, we use the associated series of open-to-close daily returns, as provided by [Noureldin et al. \(2012\)](#).

For the models featuring a DCC component we adopt the three-stage estimation approach explained in Section 2.3 and fit scalar (S), diagonal (D) and Hadamard (H) specifications; in the case of the CAW, which is based on a BEKK-type recursion, estimation is done in two steps (covariance targeting is applied in the first step and then the remaining parameters are estimated via QML) and a full (F) specification is fitted in place of the Hadamard. In all cases, estimation is performed while imposing the stationarity constraints. For the F(ull)-CAW, where the matrices $\{A, B\}$ are unconstrained, *i.e.* not necessarily symmetric, covariance stationarity is achieved by imposing that the eigenvalues of the matrix $\{A + B\}$ are less than unity in modulus and A_{11}, B_{11} are restricted to be positive.

Parameter estimates and robust standard errors over the full sample period from the CAW, Re-cDCC and RDCC models are reported in Table 2.7 along with the corresponding maximized loglikelihood and AIC values. For the RDCC, the reported likelihood values refer to the unrotated returns evaluated at the estimated parameter values.

The upper panel of the table provides the parameter estimates of the univariate GARCH(1,1) processes and the correlation (or covariance) process of the scalar specifications. At first, we notice an important difference between the GARCH estimates of the S-RDCC and the S-Re-cDCC models: the γ_i estimates are much larger in the S-Re-cDCC model, and the δ_i estimates are much lower, than in the S-RDCC model.

Table 2.6: **AA-XOM descriptive statistics**

The table reports descriptive statistics of daily realized variances for AA and XOM stocks over the period February, 2001 to December, 2009.

Stock	Mean	Max.	Min.	Std.dev.	Skewness	Kurtosis
AA	4.94	160.24	0.29	8.93	7.63	92.124
XOM	2.07	115.37	0.12	4.15	13.29	288.06

Thus, the realized conditional variances are more responsive to their own shocks than the GARCH variances based on daily returns, while in the S-RDCC they are smoother. This is a consequence of the different information used by the models. In both cases, the sum of the γ_i and δ_i estimated parameters is always very close to one, with the S-RDCC model being the most persistent. This type of differences is also found for the α and β parameters of the correlation process⁵, but with a slightly lower level of persistence. Note that the estimates obtained for the S-CAW model drive the dynamics of all elements of the conditional covariance matrix and hence they are much more comparable to the estimates of the univariate GARCH processes than to those of the DCC correlation processes.

Since we perform QML estimation of the models relying on the Wishart assumption, likelihood values cannot be compared with those of the RDCC (which use a Gaussian QL), but only with models belonging to the same class. The S-Re-cDCC delivers the best fit in terms of log-likelihood value, but the AIC criterion leads to select the S-CAW as the preferred model, revealing that the four additional parameters of the former are not relevant for this dataset.

Moving to diagonal and full specifications, there is not a clear improvement in the fit. The H-RDCC slightly improves over the corresponding scalar and diagonal types, whereas for the other models there seems to be no additional benefit from enriching the correlation (or covariance) dynamics. This is further confirmed by the lower panel of Table 2.7 showing the result of a Ljung-Box Q test at lag 10 to determine the presence of excess serial correlation in the standardized residuals. It reveals that scalar and diagonal specifications clearly dominate the full or Hadamard specifications by successfully accounting for the dynamics in the variances of the stocks, while they point toward residual autocorrelation of the covariance series.

This may come as a consequence of the low cross-sectional dimension of the dataset, which only demands one conditional correlation (or covariance) to be fitted.

In order to compare the forecasting performance of the models, we perform a one-step-head prediction exercise using the diagonal specification of each model. We recursively estimate the parameters on the first 1542 daily observations of the sample (that is using the data from February 1, 2001 until March 26, 2007) and use the last 700 as the forecast sample. We compare the forecasting accuracy of the models using the QLIK loss function, which is often used in the presence of noisy proxies; see Patton &

⁵Analogously to what Noreldin et al. (2014) report, the α parameter in the S-RDCC is not significantly different from zero. Noreldin et al. (2014) test for constant conditional correlation in the scalar case imposing $\alpha = \beta = 0$ and find that the null cannot be rejected.

Table 2.7: Full sample estimates for the the Alcoa and Exxon Mobil dataset

Panels A and B show the parameter estimates, except constant terms, computed with the full sample: 1/2/2001 – 31/12/2009, of the scalar (S), diagonal (D), Hadamard (H) Re-cDCC and RDCC models, and the S, D, and F(ull) CAW. Robust standard errors are in brackets in Panel A. Panel B (where the superscript * denotes significance at the 1% level) reports estimated parameter matrices for the non-scalar specifications. Panel C provides the p -values of the Ljung-Box Q(10) test statistics for serial correlation of order ten in the standardized residuals.

PANEL A: SCALAR PARAMETER ESTIMATES									
	CAW			Re-cDCC			RDCC		
	S	D	F	S	D	H	S	D	H
γ_1				0.376 (0.032)	0.376 (0.032)	0.376 (0.032)	0.084 (0.017)	0.084 (0.017)	0.084 (0.017)
γ_2				0.298 (0.031)	0.298 (0.031)	0.298 (0.031)	0.054 (0.012)	0.054 (0.012)	0.054 (0.012)
δ_1				0.603 (0.034)	0.603 (0.034)	0.603 (0.034)	0.902 (0.015)	0.902 (0.015)	0.902 (0.015)
δ_2				0.684 (0.035)	0.684 (0.035)	0.684 (0.035)	0.942 (0.012)	0.942 (0.012)	0.942 (0.012)
α	0.314 (0.022)			0.170 (0.033)			0.023 (0.014)		
β	0.669 (0.024)			0.807 (0.043)			0.945 (0.044)		
Total LL	-3812	-3811	-3811	-3809	-3809	-3809	-7844	-7843	-7842
AIC	7627	7630	7636	7630	7634	7639	15704	15707	15708
Num. of parameters	2	4	8	6	8	10	8	10	12

PANEL B: ESTIMATED PARAMETER MATRICES						
	D-CAW		D-Re-cDCC		D-RDCC	
A	0.343*	0.276*	0.119*	0.171*	0.008	0.068
B	0.639*	0.710*	0.871*	0.829*	0.956*	0.932*
	F-CAW		H-Re-cDCC		H-RDCC	
A	0.830*	0.060*	0.181*	0.139*	0.011	0.027
	-0.150*	0.519*	0.139*	0.163	0.027	0.043
B	0.028*	0.067*	0.818*	0.860*	0.986*	0.970*
	-0.328*	-0.703*	0.860*	0.265	0.970*	0.949*

PANEL C: P-VALUES FOR THE LJUNG-BOX TEST FOR THE RESIDUALS, 10 LAGS			
	ϵ_{11}^*	ϵ_{22}^*	ϵ_{12}^*
S-CAW	0.01	0.28	0.00
D-CAW	0.08	0.14	0.00
F-CAW	0.00	0.02	0.00
S-Re-cDCC	0.25	0.28	0.22
D-Re-cDCC	0.25	0.30	0.16
H-Re-cDCC	0.02	0.34	0.06

Sheppard (2009) and Patton (2011). Specifically, if S_t^a denotes the time- t conditional forecast from model a , the corresponding QLIK loss function is defined as

$$\ell_t^a = \log |S_t^a| + (S_t^a)^{-1} \text{tr}(C_t), \quad (2.26)$$

where C_t is the the 5-minute realized covariance (RC) estimator chosen as proxy for the latent covariance matrix. As mentioned in Section 1.1.3, any estimator that fulfills the unbiasedness condition can be used as proxy. To compare two models $\{a, b\}$ over

the out-of-sample period τ , we look at the average QLIK loss difference

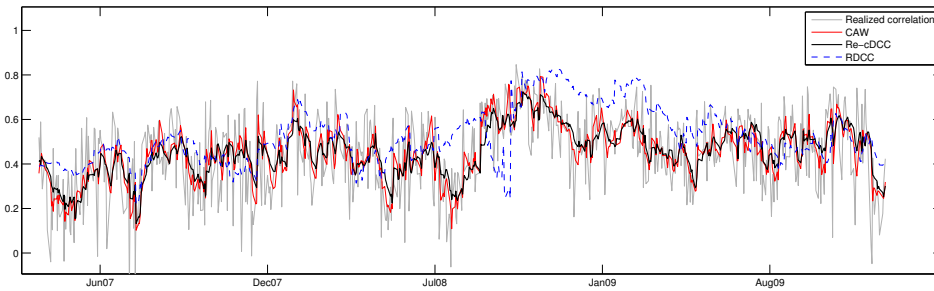
$$L_{a,b} = \frac{1}{\tau} \sum_{t=h+1}^{\tau} \ell_{a,b,t}, \quad \ell_{a,b,t} = \ell_t^a - \ell_t^b \quad (2.27)$$

and then we test the null hypothesis of equal predictive accuracy via the [Diebold & Mariano \(1995\)](#) test, where the Newey-West covariance estimator is used to compute the long run variance of $L_{a,b}$. The alternative is one-sided, so that positive (negative) values of $L_{a,b}$ correspond to superior performance of model b (a).

Comparing the D-Re-cDCC model (taken as model a) to the D-CAW and the D-RDCC, we get t -statistics of -2.86 and -6.68 , respectively, which are statistically significant at 1% and indicate that the D-Re-cDCC model strongly outperforms both competitors. Despite its higher number of parameters to be estimated than in the D-CAW model, it seems that using the D-Re-cDCC one for out-of-sample forecasting pays off by providing superior one-step ahead predictions.

Figure 2.1 plots the predicted conditional correlations from the diagonal models along with the selected proxy. Not surprisingly, the assets are highly correlated across the whole period, with the highest peaks reached during the more volatile months of the 2008 financial crisis. The Re-cDCC and CAW correlations generally exhibit a similar pattern that fits the realized correlation much better than the D-RDCC, which is not really surprising since the latter are based on daily returns only. The D-RDCC predicted correlations also appear to be more persistent than the predicted correlations obtained from the models that use the realized measures.

Figure 2.1: Predicted conditional correlations of XOM-AA from diagonal CAW, Re-cDCC and RDCC models (27/03/2007-31/12/2009).

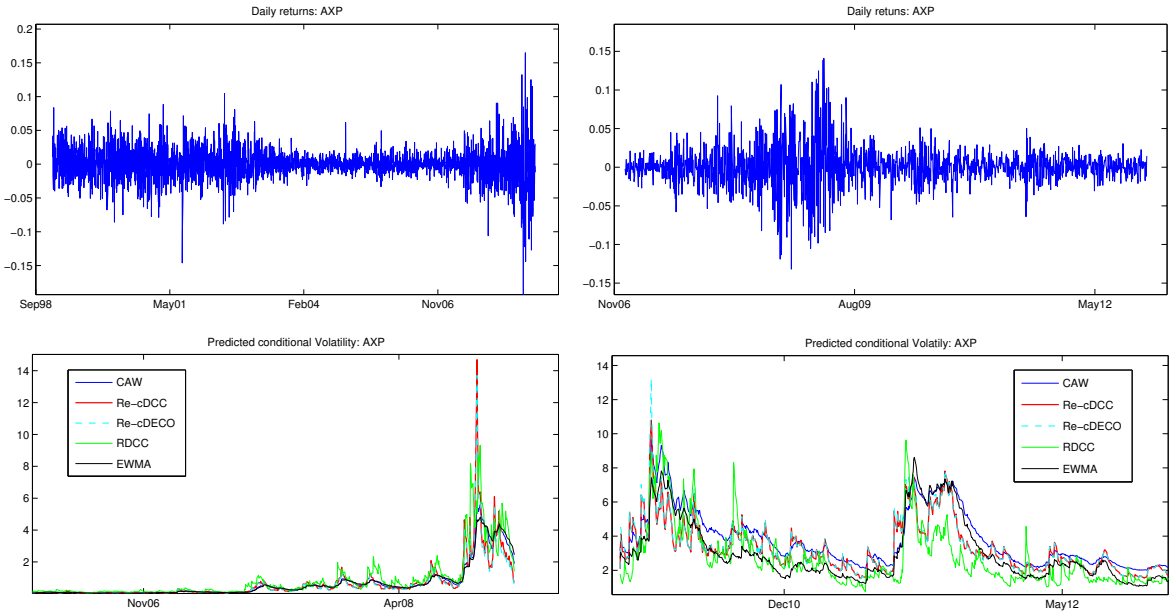


2.5.2 High dimensional model

In this subsection we push the analysis to a higher dimension. In particular, we are interested in the model forecasting ability and consequently, in its potential usefulness for portfolio allocation strategies. We conduct the same empirical application on two large-scale datasets, so as to investigate and compare the responsiveness of the model to different realized covariance estimators across different time periods. The main characteristics of the two datasets, hereafter referred to as Dataset A and Dataset B, are summarized below.

- Dataset A pertains to 46 stocks traded in the NYSE and NASDAQ. Their tickers and issue names are shown in the upper panel of Table 2.12. The sample runs from January 5, 1999 to December 31, 2008, for a total of 2483 trading days. The dataset has been cleaned from weekends, holidays and early closing days. Also, days with several consecutive missing values or constant prices have been removed and rare missing values linearly interpolated. The realized covariance matrices are constructed using the *Realized Covariance* estimator (see Barndorff-Nielsen & Shephard (2004)) based on intraday returns computed from six-minute intervals last mid-quotes. Since the trading period of the NYSE and NASDAQ is 6.5 hours per day, 65 intraday observations are available per day. Relying on the arguments of Andersen et al. (2010), we use open-to-close realized covariances.
- Dataset B pertains to 50 stocks among the largest U.S. financial institutions, with three assets (AXP, BAC, WFC) in common with Dataset A. Tickers are given in the lower panel of Table 2.12. The trading period runs from January 3, 2007 till December 13, 2012 for a total of 1499 daily observations. Data have been cleaned using the step-by-step cleaning procedure of Barndorff-Nielsen et al. (2009) whereby overnight returns have been excluded to favor estimation accuracy. The daily realized covariance matrices have been computed using the *Cholcov* estimator of Boudt et al. (2014), which is guaranteed to be positive semidefinite in the presence of asynchronicity and illiquidity (see Section 1.1.3).

Figure 2.2: Daily returns (upper panel) and predicted conditional volatilities (lower panel) of AXP stock for Dataset A (left column) and Dataset B (right column) over the two sample periods analyzed.



The upper panel of Figure 2.2 plots the series of daily returns of the AXP stock over the two sample periods. The financial crisis is included in both datasets but located at the opposite extremes of the samples, namely at the very end of Dataset A and at the beginning of Dataset B. This allows us to use a different scenario in our

Table 2.8: Full sample parameter estimates

The table shows full sample estimates (except constant terms) and robust standard errors in brackets for each dataset; $\bar{\gamma}$ and $\bar{\delta}$ are the means of the estimates of the γ_i and δ_i parameters for the different stocks (with the corresponding averages of the standard errors in brackets). Dataset A contains 46 equities from the NYSE and NASDAQ while Dataset B contains stocks of 50 among US largest institutions.

	S-CAW	S-Re-cDCC	S-Re-cDECO	S-RDCC
DATASET A: JANUARY 5, 1999 – DECEMBER 31, 2008				
$\bar{\gamma}$	-	0.348 (0.050)	0.348 (0.050)	0.110 (0.025)
$\bar{\delta}$	-	0.635 (0.054)	0.635 (0.054)	0.868 (0.028)
α	0.091 (0.005)	0.015 (0.000)	0.116 (0.028)	0.003 (0.000)
β	0.904 (0.006)	0.982 (0.001)	0.864 (0.034)	0.979 (0.004)
Total LL	433866	435747	434744	292223
AIC	-867728	-871307	-869300	-584166
Num. of parameters	2	94	94	140
DATASET B: JANUARY 3, 2007 – DECEMBER 13, 2012				
$\bar{\gamma}$	-	0.194 (0.053)	0.194 (0.053)	0.124 (0.023)
$\bar{\delta}$	-	0.794 (0.059)	0.794 (0.059)	0.871 (0.021)
α	0.055 (0.004)	0.010 (0.000)	0.041 (0.012)	0.004 (0.001)
β	0.940 (0.005)	0.978 (0.005)	0.936 (0.023)	0.924 (0.013)
Total LL	-81846	-78718	-78724	-128500
AIC	163695	157639	157652	257310
Num. of parameters	2	102	102	152

forecasting exercises and therefore constitutes an important test for the robustness of the models.

Given the large number of parameters to be estimated for non-scalar models in the considered dimensions ($n = 46$ and 50), we use scalar specifications. Full sample parameter estimates for the scalar CAW, Re-cDCC, Re-cDECO and RDCC models are reported in Table 2.8. The main difference with respect to the estimates in Table 2.7 occurs for the estimated values of the α parameter across models: it is found to be much smaller than in the case of two assets. While in the MGARCH literature, as mentioned, composite likelihood has been found to alleviate this issue, whether this applies as well to realized models is still an open question. Nevertheless, in the latter case the effect seems somehow attenuated.

As a general finding, for both datasets, the S-Re-cDCC model better fits the data than its competitors using the same data, followed by the S-Re-cDECO model. Not surprisingly, the S-CAW is too restrictive for a large number of assets, since all variance processes are constrained to share the same parameters.

To determine whether the better in-sample performance of the S-Re-cDCC goes along with forecasting gains, we perform an out-of-sample one-step ahead prediction exercise for each dataset. The parameters of each model are estimated on a fixed

rolling window and kept fixed to produce forecasts over the following ten days, which corresponds usually to two weeks of trading. Then the estimation window is shifted forward by ten observations and the models are re-estimated. This procedure is repeated until all data have been used and a set of 700 predicted covariance matrices is obtained. Predicted volatilities for the AXP stock are reproduced in the lower panels of Figure 2.2.

The first comparison of the different models is done using the Model Confidence Set (MCS) procedure of Hansen et al. (2011). This method allows us to compare the forecasts of the models based on a loss function without having to choose a reference model and leads to a subset of models that cannot be discriminated at a selected confidence level. Given that five models are in competition, this approach is preferable to performing pairwise DM tests of equal predictive accuracy. In order to make the results robust to the choice of the loss functions, in addition to the QLIK, we use three additional consistent loss functions (formulas are given in Table 2.11 of Appendix 2.6) and set the rejection levels at 10 and 25%, such that models are removed until the final model set includes the best model(s) with confidence levels of 90 and 75%. The resulting MCS p -values are given in Table 2.9. The results are similar for both datasets: the S-Re-cDCC model is included in all MCS, and is the single one at both levels in each dataset for the Stein and QLIK loss functions. For VND, a second model is in the MCS in three cases, and for the sfrob, only the S-RDCC is excluded. Actually, the latter is excluded from all MCS.

In order to compare the forecasting performances of the five models in an economically meaningful way, we also apply them to the problem of optimal portfolio allocation. We analyze the portfolio gains obtained by allocating the portfolios based on the S-Re-cDCC covariance forecasts instead of the forecasts from the chosen benchmark models. The advantage of studying global minimum variance (GMV) portfolios (in contrast to minimum variance portfolios for a given expected return) is that the changes in the ex ante optimal portfolio weights are solely determined by the forecasts of the conditional covariance matrix over the given investment horizon. The GMV portfolio weights w_t are the solution to the problem

$$\min_w \quad \frac{1}{2} \mathbf{w}_t' S_t \mathbf{w}_t, \quad s.t. \sum_{i=1}^n w_{t,i} = 1$$

where $\mathbf{w}_t$ is the vector of portfolio weights for time t chosen at time $t - 1$, S_t denotes the conditional covariance forecast from a generic model, and the only requirement is on the vector of weights to add up to unity. We further impose the portfolio weights to be nonnegative, thus ruling out short-selling operations.

We measure the portfolio performance in two ways. First, we compute the out-of-sample portfolio gross returns, defined as $r_t^P = \sum_{i=1}^n w_{t,i} r_{t,i}$, with $w_{t,i}$ and $r_{t,i}$ respectively denoting the weight and the return on asset i on that day. From the return series, we report the corresponding mean and standard deviation over the forecast period. Furthermore, we quantify the average portfolio concentration of the resulting

Table 2.9: MCS results

The table reports 90% and 75% MCS p-values for the two dataset based on 700 one-step ahead conditional covariance matrix predictions. Dataset A's forecasting sample covers the period February 7, 2006 to December 31, 2008 while Dataset B covers from March 8, 2010 to December 13, 2012. Included models highlighted in bold.

	S-CAW	S-Re-cDCC	S-Re-cDECO	S-RDCC	EWMA
DATASET A: INCLUDING THE FINANCIAL CRISIS IN THE FORECASTING PERIOD					
Panel A: 90% MCS					
sfrob	0.530	0.333	0.530	0.024	1.000
stein	0.000	1.000	0.000	0.000	0.000
Qlike	0.000	1.000	0.000	0.000	0.000
VND	1.000	0.110	0.000	0.000	0.000
Panel B: 75% MCS					
sfrob	0.537	0.343	0.537	0.025	1.000
stein	0.000	1.000	0.000	0.000	0.000
Qlike	0.000	1.000	0.000	0.000	0.000
VND	0.000	1.000	0.000	0.106	0.000
DATASET B: NOT INCLUDING THE FINANCIAL CRISIS IN THE FORECASTING PERIOD					
Panel A: 90% MCS					
sfrob	0.000	0.772	1.000	0.000	0.772
stein	0.000	1.000	0.000	0.000	0.000
Qlike	0.000	1.000	0.000	0.000	0.000
VND	0.000	1.000	0.935	0.000	0.000
Panel B: 75% MCS					
sfrob	0.000	0.780	1.000	0.000	0.780
stein	0.000	1.000	0.000	0.000	0.000
Qlike	0.000	1.000	0.000	0.000	0.000
VND	0.000	1.000	0.934	0.001	0.000

GMV portfolio weights, as large values of this measure can imply too little diversification. Concentration (CO) is measured by the Euclidean norm of the vector of portfolio weights, and averaged over the forecast period:

$$CO = \frac{1}{T} \sum_{t=1}^T \|w_t\|_2 = \frac{1}{T} \sum_{t=1}^T \left(\sum_{i=1}^n w_{t,i}^2 \right)^{1/2} \quad (2.28)$$

At each date, the measure is minimized for an equally weighted portfolio, i.e. for $w_t = (1/n)\mathbf{1}$.

Second, following [de Pooter et al. \(2008\)](#) and [Boudt et al. \(2013\)](#), we evaluate the portfolio turnover rates (for the forecast period), which can be used to assess the stability of the covariance matrix forecasts. In the moment just before rebalancing to the next period, denoted as t^- , the actual weight on each asset in the portfolio changes to $w_{i,t^-} = w_{i,t-1} \frac{1+r_{i,t}}{1+r_t^P}$, such that the total change in weights required at time t is equal to

$$TO_t = \sum_{i=1}^n |w_{t,i} - w_{i,t^-}|. \quad (2.29)$$

Table 2.10: **GMV portfolio results**

The table reports GMV portfolio summary statistics for the two datasets based on 700 one-step ahead conditional covariance matrix predictions; μ_p stands for the mean and σ_p for the standard deviations of the gross returns of the portfolio. The best model is accompanied by * or ** if its portfolio standard deviation is different than the one of the next best model at the 5% or 1% (two-sided) level, respectively. CO is defined in Eq. (2.28), Turnover in Eq. (2.29).

	S-CAW	S-Re-cDCC	S-Re-cDECO	S-RDCC	EWMA
DATASET A: CRISIS PERIOD (7/2/2006–31/12/2008)					
<i>Gross return performance</i>					
$\mu_p (e^{-03})$	-0.0774	-0.0322	-0.0650	-0.1309	-0.0772
σ_p	0.0067	0.0061*	0.0063	0.0068	0.0065
CO	0.3612	0.3338	0.3594	0.3591	0.3737
<i>Turnover</i>					
Mean	0.1203	0.1550	0.1678	0.2732	0.0984
Median	0.0907	0.1273	0.1379	0.2395	0.0703
Standard deviation	0.0862	0.0919	0.0957	0.1571	0.0826
Kurtosis	24.2441	8.1248	9.3996	10.8337	29.9466
DATASET B : POST-CRISIS PERIOD (8/3/2010–13/12/2012)					
<i>Gross return performance</i>					
μ_p	0.0323	0.0329	0.0361	0.0133	0.0315
σ_p	0.0012	0.0012	0.0013	0.0017	0.0010**
CO	0.2147	0.2118	0.2242	0.5334	0.2471
<i>Turnover</i>					
Mean	0.2321	0.2038	0.3268	0.3033	0.2887
Median	0.2011	0.1871	0.2549	0.2356	0.2648
Standard deviation	0.1280	0.0981	0.2296	0.2174	0.1293
Kurtosis	9.2625	6.1505	4.6816	8.0879	8.5586

In order to compare the turnover rates generated from the different models, we report the mean, standard deviation and kurtosis of the resulting TO_t series. The results are collected in Table 2.10.

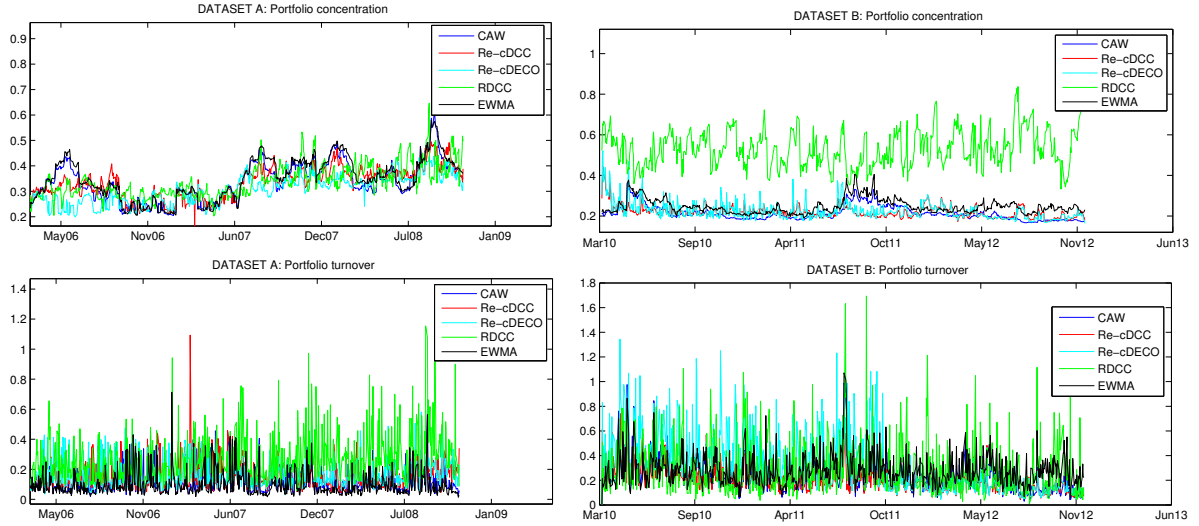
The differences between the two time periods strongly emerge by comparing the portfolio gross return performances achieved by the models. Indeed, the impact of the 2008 financial crisis is evident in the negative average portfolio returns and higher standard deviations in the upper panel of the table, in contrast to the positive returns predicted for the calmer period of 2010-2012, which is also characterized by lower average values of standard deviations.

In terms of gross return performance, the calm period shows no clear evidence of dominance of one specific model over the other, except that the S-RDCC is generating a lower mean return, a higher standard deviation and a much higher concentration than the other models, as also evident from the upper right panel of Figure 2.3. As for the turnover rates, the S-Re-cDCC compares favorably to the competitors by delivering on average more stable portfolio weights and thus lower portfolio churn, which is desirable if the transaction cost level is sizeable.

The benefits of the S-Re-cDCC allocation are more visible during the turbulent crisis period, where it generates the largest mean return and the lowest standard deviation yet maintaining the smallest concentration. The S-Re-cDECO model is less rewarding in terms of mean return but comparable on the other two measures, while the S-RDCC model performs quite poorly in terms of mean return. On the other hand, the allocations delivered by the S-CAW and the EWMA models increase volatility but lower

portfolio turnover. However, kurtosis values are largely in favor of the S-Re-cDCC, with percentage differences ranging between 16 and 268%. This denotes more stability in the S-Re-cDCC conditional covariance matrices, and overall, a higher responsiveness to the changing economic environment.

Figure 2.3: Daily portfolio concentration and turnover for Dataset A (left side) and Dataset B (right side).



2.6 Conclusions

This chapter proposes a dynamic class of models for realized covariance matrices which combine flexibility and computational tractability that make them applicable to a large number of assets. Within the different specifications nested in the proposed class, we focused on the so-called *Re-cDCC* model and explained in details its several advantages. First and foremost, it can be specified and estimated in steps, thanks to the joint use of a consistent DCC specification and of a central Wishart distribution, with each step having a QML interpretation. Then, due to the possibility of applying a preliminary covariance targeting step, parameter proliferation is kept under control so that the model can be easily fit to matrices of large dimension, where large in this context means of the order of fifty. Moreover, computational tractability and speed of implementation are preserved even in the largest case. This is already a significant progress relative to existing models relying on less flexible parameterizations.

As the procedure is expected to be less efficient than a one-step QML estimation, the cost of introducing targeting along with the multistep approach is investigated via a Monte Carlo study, revealing that the proposed three-step estimator is more precise than the standard one-step QMLE when the cross-sectional dimension is larger than ten and the time series length is up to two thousand. This favorably supports our conjecture that as the sample dimension increases, the usual 1-step estimator, even if feasible, leads to a considerable distortion in the estimated parameters especially if the

time-series is not sufficiently long. Our multistep approach proves useful to considerably reduce the bias in this situation.

Finally, the empirical forecasting performance of the model against some benchmarks is illustrated in low and large dimensional settings, in the latter case using different realized covariance estimators, with qualitatively similar results. The Re-cDCC model is found to perform best in the cases under study, thus providing evidence to its potential usefulness for all kind of financial applications beyond the one of portfolio allocation we reported.

Appendix

2.A Technical Proofs

Proof of Proposition 1: Let $\mathbf{n}_M = \text{vech}(M)$ denote the model's static parameters, and let ϕ_s and ϕ_c respectively denote the vectors of dynamic parameters indexing $\{\gamma, \delta\}$ in Eq.(2.5) and $\{A, B\}$ in Eq.(2.7). The full parameter vector is $\phi = (\nu, \mathbf{n}'_M, \phi'_s, \phi'_c)$, and the true parameter vector is denoted by ϕ_0 . Recall from Eq.(2.17) that the log-likelihood contribution of observation t , ignoring constant terms and those not depending on H_t , is given by

$$\ell_t(\phi) = -\frac{\nu}{2} \log |S_t| - \frac{\nu}{2} \text{tr} \left(S_t^{-1} C_t \right),$$

where H_t and C_t are $n \times n$ PDS matrices and L is lower triangular. The structure of the log-likelihood allows for a three-step estimation strategy to estimate ϕ , where the first step is represented by a method of moment estimator: $\hat{M} = \hat{L}\hat{L}' = 1/T \sum_{t=1}^T C_t$, which implies $\hat{n}_M$. Note that for notational ease, we use L instead of $\hat{L}$. The following two steps are obtained by decomposing the log-likelihood function as in Eq.(2.18). First, by Eq.(2.3), we have:

$$\ell_t(\phi) = -\frac{\nu}{2} \left[\log |LH_tL'| + \text{tr}(L'^{-1}H_t^{-1}L^{-1}C_t) \right],$$

and by Eq.(2.4), *i.e.* $H_t = D_t R_t D_t$,

$$\ell_t(\phi) = -\frac{\nu}{2} \left[\log |LD_t R_t D_t L'| + \text{tr}(L'^{-1}D_t^{-1}R_t^{-1}D_t^{-1}L^{-1}C_t) \right].$$

By noticing that $\log |LD_t R_t D_t L'| = 2 \log |D_t| + \log |R_t| + 2 \log |L|$ and using the trace property of invariance under cyclical rotation, we can write

$$\begin{aligned} \ell_t(\phi) = & -\frac{\nu}{2} [2 \log |D_t| + \log |R_t| + 2 \log |L|] \\ & - \frac{\nu}{2} \left[\text{tr} \left((R_t^{-1} - I_n) D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} + D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right) \right], \end{aligned}$$

where we added and subtracted the term $D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1}$ in the second term on the right hand side. Finally, reallocating elements we get

$$\begin{aligned} \ell_{v,t}(\nu, \phi_v, \hat{\mathbf{n}}_M) &= -\nu \log |D_t| - \frac{\nu}{2} \text{tr} \left(D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right), \\ \ell_{c,t}(\nu, \phi_c, \hat{\mathbf{n}}_M, \hat{\phi}_v) &= -\frac{\nu}{2} \left[\log |R_t| + \text{tr} \left((R_t^{-1} - I_n) D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right) \right] - \nu \log |L|, \end{aligned}$$

which correspond to Equations (2.19)–(2.20) ■

Proof of Proposition 2: Most steps of the following proofs rely on the matrix derivatives rules stated in [Lutkepohl \(1996\)](#). We start deriving the $(1 \times m_v)$ score vector $\partial \ell_{v,t}(\nu, \phi_v, \hat{\mathbf{n}}_M) / \partial \phi_v$ and prove that its expected value in $\phi_{v,0}$ is zero. From the $\ell_{v,t}$ log-likelihood function we have:

$$\begin{aligned} \frac{\partial \ell_{v,t}(\nu, \phi_v, \hat{\mathbf{n}}_M)}{\partial \phi_v} &= -\nu \frac{\partial \log |D_t|}{\partial \phi_v} - \frac{\nu}{2} \frac{\partial \text{tr} \left(D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right)}{\partial \phi_v} \\ &= -\frac{\nu}{2} \frac{\partial \log |H_t|}{\partial \phi_v} - \frac{\nu}{2} \frac{\partial \text{tr} \left(L^{-1} C_t L'^{-1} H_t^{-1} \right)}{\partial \phi_v} \\ &= \frac{\nu}{2} \left[-\sum_{i=1}^n \frac{\partial \log H_{ii,t}}{\partial \phi_v} - \sum_{i=1}^n L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} \left(\frac{\partial H_{ii,t}^{-1}}{\partial \phi_v} \right) \right] \\ &= \frac{\nu}{2} \left[-\sum_{i=1}^n \frac{1}{H_{ii,t}} + \sum_{i=1}^n L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} H_{ii,t}^{-2} \right], \end{aligned}$$

where we used $2 \log |D_t| = \log |H_t|$ and $D_t^{-2} = H_t^{-1}$. Using the law of iterated expectation, we obtain

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell_{v,t}(\nu, \phi_v, \hat{\mathbf{n}}_M)}{\partial \phi_v} \right] &= \frac{\nu}{2} \mathbb{E} \left\{ \mathbb{E}_{t-1} \left[-\sum_{i=1}^n \frac{1}{H_{ii,t}} + \sum_{i=1}^n L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} H_{ii,t}^{-2} \right] \right\} \\ &= \frac{\nu}{2} \mathbb{E} \left\{ -\sum_{i=1}^n \frac{1}{H_{ii,t}} + \sum_{i=1}^n \mathbb{E}_{t-1} \left(L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} \right) H_{ii,t}^{-2} \right\}. \end{aligned}$$

The expected second step score for observation t is equal to zero as, by the conditional moment assumption stated in Eq.(2.3), it holds $\mathbb{E}_{t-1} \left(L_{ii}^{-1} C_{ii,t} L'_{ii}^{-1} \right) = H_{ii,t}$ and hence

$$\mathbb{E} \left[\frac{\partial \ell_{v,t}(\nu, \phi_v, \hat{\mathbf{n}}_M)}{\partial \phi_v} \right] = 0.$$

The derivation of the $(1 \times m_c)$ score vector $\partial \ell_{c,t}(\nu, \phi_c, \hat{\mathbf{n}}_M, \hat{\phi}_v) / \partial \phi_c$ follows a similar procedure. From the $\ell_{c,t}$ likelihood function with $\mathbf{n}_M$ and ϕ_v fixed at the estimates obtained in the previous steps, we get

$$\begin{aligned} \frac{\partial \ell_{c,t}(\nu, \phi_c, \hat{\mathbf{n}}_M, \hat{\phi}_v)}{\partial \phi_c} &= -\frac{\nu}{2} \frac{\partial \log |R_t|}{\partial \phi_c} - \frac{\nu}{2} \frac{\partial \text{tr} \left(R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right)}{\partial \phi_c} \\ &= -\frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} \right) + \frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right). \end{aligned}$$

Again, by taking expectations on both sides, it gives

$$\begin{aligned} \mathbb{E} \left[\frac{\partial \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_v)}{\partial \phi_c} \right] &= -\frac{\nu}{2} \mathbb{E} \left\{ \mathbb{E}_{t-1} \left[\text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} \right) + \right. \right. \\ &\quad \left. \left. + \frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} R_t^{-1} D_t^{-1} L^{-1} C_t L'^{-1} D_t^{-1} \right) \right] \right\} \\ &= -\frac{\nu}{2} \mathbb{E} \left\{ \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} \right) + \right. \\ &\quad \left. + \frac{\nu}{2} \text{tr} \left(R_t^{-1} \frac{\partial R_t}{\partial \phi_c} R_t^{-1} D_t^{-1} L^{-1} L \mathbb{E}_{t-1}(C_t) L' L'^{-1} D_t^{-1} \right) \right\} \end{aligned}$$

where at the true parameter value $\phi_c = \phi_{c,0}$, it holds $\mathbb{E}_{t-1}(C_t) = S_t = L D_t R_t D_t L'$ and hence $\mathbb{E} \left[\frac{\partial \ell_{c,t}(\nu, \phi_c, \hat{n}_M, \hat{\phi}_v)}{\partial \phi_c} \right] = 0$ ■

2.B Tables

Table 2.11: **Implemented Loss functions**

The table gives the formulas of the chosen matrix loss functions: squared frobenius distance (sfrob), QLIKE, Stein and von Neumann Divergence (VND). S_t is the predicted conditional covariance matrix while C_t is the realized measure used as proxy; n denotes the number of assets and λ_i are the eigenvalues of $(S_t - C_t)^2$. The squared frobenius distance is a symmetric quadratic loss functions based on the forecast error, while the QLIKE is based on the log-likelihood function of the Normal distribution. The last two loss functions belong to the family of Bregman matrix divergences, which generalize the squared Euclidean distance to a class of distances that all share similar properties. The Stein loss function is scale invariant, being based on the standardized forecast error (in matrix sense) and can be obtained as the objective function of the Wishart distribution, while the von Neumann divergence is a matrix generalization of Kullback-Leibler divergence. Both Stein and VND are asymmetric with respect to over/under predictions.

Loss Function	Formula
sfrob	$\sum_{i=1}^n \lambda_i$
Qlike	$\log S_t + \text{tr} (S_t^{-1} C_t)$
stein	$\text{tr}(S_t^{-1} C_t) - \log S_t^{-1} C_t - n$
VND	$\text{tr} (C_t \log C_t - C_t \log S_t - C_t + S_t)$

Table 2.12: Data

The table reports tickers and issue names for the stocks used in the empirical application. Dataset A contains 46 equities from the S&P 500 traded over the period 01/1999–12/2008 while Dataset B contains stocks of 50 among US largest institutions traded over 01/2007–12/2012. The assets in common (AXP, BAC, WFC) are only reported in the upper panel.

Symbol	Issue Name	Symbol	Issue Name
Dataset A: 46 selected constituents of the NYSE and NASDAQ on January 1999			
AAPL	Apple	JNJ	Johnson & Johnson
ABT	Abbott Laboratories	JPM	J P Morgan Chase & Co
AXP	American Express Company	KO	Coca Cola
BA	The Boeing Company	ETR	Entergy Corporation
BAC	Bank of America	LLY	Eli Lilly and Company
BMJ	Bristol-Myers Squibb Company	MCD	McDonald's Corporation
BP	BP p.l.c.	MMM	3M Company
C	Citigroup	MRK	Merck & Company, Inc.
CAT	Caterpillar	MS	Morgan Stanley
CL	Colgate-Palmolive Co.	MSFT	Microsoft Corporation
CSCO	Cisco Systems	SLB	Schlumberger N.V.
CVX	Chevron Corporation	T	AT&T
DELL	Dell Inc.	TWX	Time Warner
DIS	Walt Disney Company	UN	Unilever NV
EXC	Exelon Corporation	VZ	Verizon Communications
F	Ford Motor Co.	PEP	Pepsico, Inc.
FDX	FedEx Corporation	PFE	Pfizer
GE	General Electric	PG	Procter & Gamble Company
HD	The Home Depot	QCOM	QUALCOMM Incorporated
HNZ	HNZ Group	WFC	Wells Fargo & Company
HON	Honeywell International	XOM	Exxon Mobil Corporation
INTC	Intel Corporation	XRX	Xerox Corporation
Dataset B: 50 selected US financial institutions on January 2007			
ACAS	American Capital	GS	The Goldman Sachs Group
AET	Aetna	HBAN	Huntington Bancshares Incorporated
AFL	Aflac Incorporated	HIG	Hartford Financial Services Group
AIG	American International Group	HNT	Health Net Inc.
AIZ	Assurant	ICE	Intercontinental Exchange Inc.
ALL	The Allstate Corporation	JNS	Janus Capital Group, Inc
AMP	Ameriprise Financial	KEY	KeyCorp
BBT	BB&T Corporation	NTRS	Northern Trust Corporation
BEN	Franklin Resources, Inc.	NYX	NYX Gaming Group Limited
BK	Bank Of New York Mellon Corporation	PFG	Principal Financial Group Inc
BLK	BlackRock, Inc	PGR	Progressive Corporation
BRKB	Berkshire Hathaway	PNC	PNC Financial Services Group
CB	Chubb Limited	PRU	Prudential Financial
CBG	CBRE Group	RF	Regions Financial Corporation
CINF	Cincinnati Financial Corporation	SEIC	SEI Investments Company
CMA	Comerica Incorporated	SNV	Synovus Financial Corp.
COF	Capital One Financial Corporation	STI	SunTrust Banks
CVH	CV Holdings	STT	State Street Corporation
EV	Eaton Vance Corporation	TMK	Torchmark Corporation
FITB	Fifth Third Bancorp	TROW	T.Rowe Price Group
FNH	Fidelity National Financial	UNH	UnitedHealth Group
GNW	Genworth Financial Inc	UNM	Unum Group
MET	MetLife, Inc.	USB	U.S.Bancorp
MTB	M&T Bank Corporation	WU	The Western Union Company
ZION	Zions Bancorporation		

Table 2.13: Monte Carlo study results: comparison of estimators in low dimensions

The table reports Monte Carlo results for the one-step QMLE (1QMLE) and the two-step QMLE (2QMLE) across $T=500, 1000, 2000$ and $n=2, 5, 7, 10, 15$. RB and RMSE respectively denote the relative bias and the root mean square error (see Section 2.4). Results based on 500 simulations.

		T = 500		T = 1000		T = 2000	
		RB	RMSE	RB	RMSE	RB	RMSE
n=2							
$\bar{\gamma}$	1QML	0.127	0.019	0.031	0.013	0.014	0.009
	2QML	0.136	0.020	0.035	0.013	0.015	0.009
$\bar{\delta}$	1QML	-0.111	0.175	-0.037	0.093	-0.011	0.028
	2QML	-0.118	0.184	-0.041	0.101	-0.011	0.028
α	1QML	0.012	0.033	0.008	0.024	-0.001	0.018
	2QML	0.009	0.033	0.005	0.024	-0.001	0.018
β	1QML	-0.057	0.138	-0.022	0.074	-0.008	0.043
	2QML	-0.057	0.138	-0.021	0.074	-0.008	0.043
n=5							
$\bar{\gamma}$	1QML	0.142	0.015	0.053	0.009	0.014	0.006
	2QML	0.141	0.015	0.054	0.009	0.013	0.006
$\bar{\delta}$	1QML	-0.148	0.173	-0.047	0.072	-0.012	0.022
	2QML	-0.146	0.170	-0.047	0.073	-0.012	0.022
α	1QML	-0.005	0.011	-0.002	0.008	0.000	0.005
	2QML	-0.008	0.011	-0.003	0.008	-0.000	0.005
β	1QML	-0.017	0.029	-0.008	0.019	-0.004	0.013
	2QML	-0.016	0.029	-0.008	0.019	-0.004	0.012
n=7							
$\bar{\gamma}$	1QML	0.543	0.030	0.336	0.019	0.243	0.013
	2QML	0.134	0.012	0.048	0.008	0.012	0.005
$\bar{\delta}$	1QML	-0.305	0.294	-0.138	0.143	-0.067	0.065
	2QML	-0.145	0.161	-0.045	0.066	-0.012	0.020
α	1QML	0.010	0.007	0.005	0.005	0.007	0.004
	2QML	-0.003	0.007	-0.002	0.005	0.000	0.004
β	1QML	-0.020	0.025	-0.010	0.015	-0.006	0.010
	2QML	-0.016	0.022	-0.007	0.014	-0.004	0.010
n=10							
$\bar{\gamma}$	1QML	0.877	0.045	0.593	0.030	0.479	0.024
	2QML	0.128	0.011	0.045	0.007	0.003	0.004
$\bar{\delta}$	1QML	-0.314	0.286	-0.166	0.152	-0.112	0.102
	2QML	-0.154	0.162	-0.047	0.061	-0.009	0.014
α	1QML	0.013	0.005	0.010	0.004	0.005	0.003
	2QML	-0.009	0.005	-0.000	0.004	-0.000	0.002
β	1QML	-0.018	0.020	-0.010	0.012	-0.005	0.007
	2QML	-0.014	0.017	-0.007	0.011	-0.003	0.006
n=15							
$\bar{\gamma}$	1QML	0.933	0.047	0.879	0.044	0.883	0.044
	2QML	0.163	0.011	0.070	0.006	0.020	0.003
$\bar{\delta}$	1QML	-0.141	0.127	-0.118	0.106	-0.112	0.100
	2QML	-0.162	0.161	-0.055	0.064	-0.014	0.019
α	1QML	0.015	0.005	0.009	0.003	0.007	0.002
	2QML	-0.006	0.003	-0.003	0.002	-0.000	0.002
β	1QML	-0.017	0.023	-0.007	0.008	-0.005	0.005
	2QML	-0.015	0.014	-0.006	0.007	-0.004	0.005

Chapter 3

Forecasting Comparison of Long Run Component Dynamic Models For Realized Covariance Matrices¹

Abstract Novel model specifications that include a time-varying long run component in the dynamics of realized covariance matrices are proposed. The modeling framework allows the secular component to enter the model either additively or as a multiplicative factor, and to be specified parametrically, using a MIDAS filter, or non-parametrically. Estimation is performed by maximizing a Wishart quasi-likelihood function. The one-step ahead forecasting performance is assessed by means of three approaches: model confidence sets, minimum variance portfolios and Value-at-Risk. The results illustrate that the proposed models outperform benchmarks incorporating a constant long-run component both in and out-of-sample.

3.1 Introduction

Volatility models that include a time-varying long run component have been increasingly used over the last decade. Empirical evidence has indeed suggested that the level of variances and correlations is changing over time as a function of economic conditions. This can have important forecasting implications, as models reverting to a constant level in the long run will produce the same long term forecast at any point in time, thus completely ignoring the changing economic conditions. Without controlling for the different sources affecting volatility, they will end up producing spurious forecasts.

¹This chapter is inspired by the paper [Bauwens, Braione & Storti \(2016b\)](#) forthcoming in *Annals of Economics and Statistics*.

Since [Engle & Lee \(1999\)](#) introduced a GARCH model with additive long run and short run dynamic components, several others have proposed related component models for volatility. [Engle & Rangel \(2008\)](#) specify a semi-parametric model that allows for time-variation in the unconditional level of stock market volatility, while [Engle et al. \(2013\)](#) use a parametric specification involving the Mixed Data Sampling (MIDAS) scheme to extract a slowly moving secular component around which daily volatility moves. Specifically, the short run component is a GARCH specification based on daily squared returns, while the long run component is driven by realized volatilities computed over a monthly, quarterly or biannual basis. The idea has been extended to dynamic correlations in the DCC-MIDAS of [Colacito et al. \(2011\)](#) and in the multiplicative-DCC of [Bauwens et al. \(2013\)](#).

More recently, researchers have paid attention to the specification of models directly fitted to time series of realized covariances. As advocated by [Andersen et al. \(2003\)](#), high-frequency based models produce significant improvements in predictive performance relative to models that rely on daily data alone. Pioneering works in this respect can be found in [Gouriéroux et al. \(2009\)](#), [Jin & Maheu \(2013\)](#), and [Chiriac & Voev \(2011\)](#), among others. Our approach is closer to [Golosnoy et al. \(2012\)](#), who proposed a model where the realized covariance matrix of asset returns is assumed to follow a conditional Wishart distribution with a time-varying conditional expectation inspired by the full BEKK model of the multivariate GARCH literature. This conditionally autoregressive Wishart (CAW) model aims to capture the daily dynamics of the realized covariances around a constant long term level. By replacing this constant level matrix with a MIDAS filter of past realized covariance matrices, the authors specify a MIDAS-CAW model that is designed to capture long run fluctuations in the levels of daily covariances. They apply their models to daily realized covariance matrices for the returns of five stocks and show that the MIDAS-CAW specification dominates the baseline CAW model both in-sample and out-of-sample. Nevertheless, these CAW models entail estimating more than one hundred parameters (for five stocks) due to the use of a full BEKK specification and thus they prove cumbersome in higher dimensional settings.

The models and results in the papers cited above motivate the research pursued in the present work. First, they suggest that long-term dynamic models for realized covariance matrices can be applied to portfolio optimization or asset allocation strategies and financial risk measures. Such models are clearly expected to provide higher benefits than models relying on daily data alone. Second, the BEKK is only one possible specification that can be used to model the short term co-volatilities components. This suggests to use other models. To this end, we propose a wide set of models for realized covariance matrices, which account for time-variation both in co-volatilities long run levels and in their short-run dynamics around these levels. In order to maintain consistency with the operational settings typically encountered in standard financial applications, we favour parsimonious specifications applicable in a realistic dimensional framework by opting for scalar models as detailed below.

Similarly to [Golosnoy et al. \(2012\)](#), we decompose the conditional expectation of the

realized covariance matrix into a slowly evolving process and a mean-reverting short run one. We extend the existing framework discussed so far mainly into two directions. First, our model setting allows the long-run component to enter the model structure either in additive fashion, as a time-varying intercept in the co-volatility and correlation models, or as a multiplicative factor. This allows us to specify it parametrically, using a MIDAS-type filter, or even non-parametrically, by means of a matrix-variate smoother. Second, we exploit the advantages of a parsimonious modelling of the short run dynamics by employing different scalar specifications inspired by the multivariate GARCH literature, namely BEKK, DCC and DECO-type models. The estimation of the models is performed by maximizing in one step the likelihood function based on the conditional Wishart assumption. As shown in [Bauwens et al. \(2012\)](#) and [Noureldin et al. \(2012\)](#), the function has a quasi-likelihood interpretation and the estimator is consistent even if the assumed underlying distribution is not Wishart.

The remainder of the chapter is the following. Section [3.2](#) introduces the novel time-varying long run component models, where particular care is dedicated to the explanation of the different component structures employed. Specifically, depending on the combination scheme adopted for mixing the short and long run components, in the remainder indicated as *additive* or *multiplicative*, two main classes of models are presented. In addition, within each class, some constant long term models are derived as benchmarks for the forecasting comparison. The procedure for quasi-maximum likelihood (QML) estimation is covered in Section [3.3](#). Section [3.4](#) presents the results of an empirical application to a set of ten US stocks. The goal is to compare the statistical accuracy of the novel set of time-varying long-run (TVLR) component models vis-à-vis their constant long-run (CLR) equivalents, from both an in- and out-of-sample perspective. Therefore, we first compare their in-sample goodness of fit in terms of standard information criteria, then we evaluate their forecasting performance by means of three approaches: a forecasting comparison based on the use of the Model Confidence Set (MCS) as a tool for discriminating among the models considered, minimum variance portfolios and Value-at-Risk prediction. Our findings shed some light on the potential benefits deriving from using models that feature a time-varying long run component which guarantees a major flexibility and a higher responsiveness in periods of pronounced market instability. In particular, the proposed models outperform their CLR benchmarks under each of three criteria considered in our empirical application. In the forecasting comparison, the multiplicative TVLR models dominate the other classes of models in terms of frequency of occurrence in the MCS, especially at the 75% confidence level. Also, in terms of portfolio optimization, the TVLR models yield significantly lower portfolio variances than the employed counterparts and provide more accurate VaR predictions.

Some concluding remarks along with a discussion of possible lines for future research are provided in Section [3.5](#).

3.2 TVLR component models

The purpose of this section is to formally present the new multivariate time-varying long run component dynamic models for realized volatilities and correlations of asset returns. Within each subsection, models incorporating a constant long run component are derived as special cases and used as benchmarks in the empirical analysis.

The modeling framework is common to all models, and is similar to the one already recalled at the beginning of Chapter 2. For the reader's convenience, we thus keep the same notation and let C_t define the positive definite and symmetric (PDS) n -dimensional daily realized covariance matrix of order n . Again, we assume C_t to have a central conditional Wishart distribution

$$C_t | \mathfrak{S}_{t-1} \sim W_n(\nu, S_t/\nu), \quad \forall t = 1, \dots, T \quad (3.1)$$

and a conditional mean equal to

$$\mathbb{E}(C_t | \mathfrak{S}_{t-1}) = S_t, \quad (3.2)$$

so that the matrix S_t is referred to as the conditional covariance matrix of $\mathbf{r}_t$.

Equations (3.1)-(3.2) define a generic conditional autoregressive Wishart (CAW) model as proposed by Golosnoy et al. (2012). A CAW model specifies the dynamic dependence of the $\{C_t\}$ process through a dynamic equation for the scale matrix S_t of the Wishart distribution, so as to capture the serial dependences in the realized variances and covariances. In addition, S_t can be further designed to directly capture the long run fluctuations in the levels around which realized variances and covariances or correlations fluctuate from day to day. To this extent, the realized covariance dynamics are split in two components, a secular smoothly varying component and a short lived one. In the next subsections, we define several ways to specify both components.

3.2.1 Long-run additive component models for correlations

The structure of the first set of models is inspired by the DCC-MIDAS model of Colacito et al. (2011). In that model, conditional variances and correlations are modeled separately. We retain the original model structure but we adapt it to the realized covariance framework. More specifically, the scale matrix S_t in Eq.(3.2) is decomposed in terms of standard deviations and correlations:

$$S_t = D_t R_t D_t, \quad (3.3)$$

where the i -th entry of the diagonal matrix $D_t = \{\text{diag}(S_t)\}^{1/2}$ is given by the conditional standard deviation $\sqrt{S_{ii,t}}$ of asset i , and R_t is the corresponding conditional correlation matrix of $\mathbf{r}_t$. Notice that the terms 'conditional' are used for convenience and do not imply that $\sqrt{S_{ii,t}}$ is the conditional mean of $\sqrt{C_{ii,t}}$, and that R_t is the con-

ditional mean of the correlation matrix obtained from C_t . This structure allows us to model separately conditional variances and correlations, in the spirit of Engle (2002a) and Tse & Tsui (2002). Moreover, each univariate volatility equation is specified and estimated independently of the others.

Baseline models Short and long term dynamic components are accounted for, but separately treated. Namely, for every t , each conditional volatility is expressed as the product of two components:

$$\mathbb{E}(C_{ii,t}|\mathfrak{F}_{t-1}) = \bar{S}_{ii,t}S_{ii,t},$$

with $\bar{S}_{ii,t}$ and $S_{ii,t}$ denoting the secular and the mean-reverting short run component, respectively. For each asset $i = 1, \dots, n$, the short run component is specified as a mean reverting GARCH(1,1)-type process:

$$S_{ii,t} = (1 - \gamma_i - \delta_i) + \gamma_i \frac{C_{ii,t-1}}{\bar{S}_{ii,t-1}} + \delta_i S_{ii,t-1}, \quad (3.4)$$

where mean-reversion to unity is imposed for identification and stationarity is assured by imposing the usual restriction $\gamma_i + \delta_i < 1$.

The secular component is specified using a finite distributed lag specification typically encountered in the Mixed Data Sampling (MIDAS) regression framework which is assumed to be a weighted sum of a judiciously chosen number L of lagged realized variances:

$$\bar{S}_{ii,t} = \bar{m}_i + \theta_i \sum_{l=1}^L \phi_l(\omega_s) C_{ii,t-l}, \quad l = 1, \dots, L, \quad (3.5)$$

$$(3.6)$$

where $\bar{m}_i$ and θ_i are restricted to be positive scalars. Differently from standard MIDAS regressions, Eq. (3.5) does not involve mixture of different frequencies, as both $\bar{S}_{ii,t}$ and $C_{ii,t}$ are at the daily level. Nevertheless, as the mixed-frequency feature is introduced in subsequent models, we keep the notation consistent throughout the chapter and simply refer to the MIDAS filter even if not strictly proper. Notice that while Colacito et al. (2011) specified $\bar{S}_{ii,t}$ as a locally constant long run component, without loss of generality we define it based on rolling samples that change from day to day.

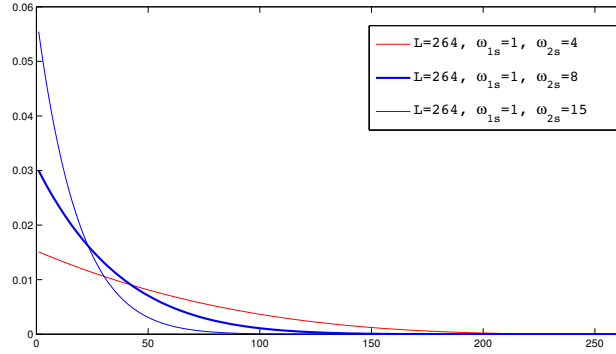
The weight function $\phi_l(\omega_s)$, normalized so that the sum of the weights is equal to

one, is specified according to the Beta² function, defined as

$$\phi_l(\omega_s) = \frac{\left(\frac{l}{L}\right)^{\omega_{1s}-1} \left(1 - \frac{l}{L}\right)^{\omega_{2s}-1}}{\sum_{j=1}^L \left(\frac{j}{L}\right)^{\omega_{1s}-1} \left(1 - \frac{j}{L}\right)^{\omega_{2s}-1}},$$

where for brevity, $\phi(\omega_s) = \phi(\omega_{1s}, \omega_{2s})$. We use the subscript s to differentiate ω_s from the similar scheme that is introduced below for correlations. In order to ensure a decaying pattern of the weights, it suffices to impose $\omega_{1s} = 1$ and estimate ω_{2s} under the restriction $\omega_{2s} > 1$; after this, the decaying speed of the weights over time is determined by the data. As an example, Figure 3.1 shows a few parameterizations of the Beta polynomial weighting function with different decaying rates. It should be obvious that

Figure 3.1: Beta polynomial weighting function
Examples of decay patterns depending on the parameterization of ω_{2s} : 4, 8 and 15. In all cases $L = 264$ and $\omega_{1s} = 1$.



the number of lags L can be chosen high enough so as to obtain a sufficiently smooth path of the long run component. In order to avoid truncating at too low a lag, which could hide important long run dependences, we set L equal to 264 so as to cover a period of one year.

The component structure for the conditional correlation matrix is inspired by the same ideas. The main element is to replace the constant intercept matrix in the DCC specification by a time-varying component expressed as a weighted average of past realized correlations. This amounts to specify R_t as

$$R_t^{DCC} = (1 - \alpha - \beta)\bar{P}_t + \alpha P_{t-1} + \beta R_{t-1}^{DCC}, \quad (3.7)$$

$$\bar{P}_t = \sum_{l=1}^L \phi_l(\omega_r) P_{t-l}, \quad (3.8)$$

where $P_t = \text{diag}\{C_t\}^{-1/2} C_t \text{diag}\{C_t\}^{-1/2}$ and $\phi_l(\omega_r)$ is again a Beta weight function that

²The function $\phi(\cdot)$ is parameterized according to the Beta polynomial in all the long-run parametric specifications we consider. Given our interest in preserving parameter parsimony, this choice appears convenient since it is flexible enough to reproduce a wide range of patterns only requiring the estimation of two additional parameters or even one, under the assumption of declining weights. A more detailed discussion of this and other weighting schemes commonly used in macroeconomic and financial applications can be found in [Ghysels et al. \(2005\)](#) and [Ghysels et al. \(2007\)](#).

applies the same weighting scheme to all the elements of the matrix P_t .

Instead of using the DCC-type specification, the dynamic equicorrelation (DECO) type of model of [Engle & Kelly \(2012\)](#) can also be used. It is written as

$$R_t^{DECO} = (1 - \rho_t)I_n + \rho_t J_n, \quad (3.9)$$

$$\rho_t = \frac{1}{n(n-1)} (\iota' R_t^{DCC} \iota), \quad (3.10)$$

where ρ_t is the equicorrelation parameter at time t , I_n is the n -dimensional identity matrix, J_n is the $n \times n$ matrix of ones and ι denotes a vector of ones.

Equations (3.4) to (3.8) denote the *Additive Midas Realized DCC (AMReDCC)* model, while replacing Eq. (3.8) with Eq. (3.9) gives the *AMReDECO* model. Note that by imposing $\theta_i = 0$ in every univariate volatility equation and assuming a constant long run level for correlations (i.e. $\bar{P}_t = \bar{P} \forall t$), one obtains the ReDCC model of [Bauwens et al. \(2012\)](#) as a particular case.³ The ReDCC can thus be considered as a benchmark for this class of models, even though it is not formally nested in the AMReDCC since the latter does not include a constant intercept matrix in Eq.(3.7).

Variations of the baseline models Two variations of the baseline models are proposed below. To start with the conditional volatility dynamics, we propose an alternative specification for the MIDAS polynomial in Eq.(3.5), in the spirit of [Golosnoy et al. \(2012\)](#). Specifically, we aggregate the univariate series of realized variances over a horizon of m trading days and consider K lagged observations, such that Eq.(3.5) is rewritten as follows:

$$\bar{S}_{ii,t} = \bar{m}_i + \theta_i \sum_{k=1}^K \phi_k(\omega_s) C_{ii,t,k}^{(m)} \quad (3.11)$$

$$C_{ii,t,k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} C_{ii,\tau}, \quad k = 1, \dots, K. \quad (3.12)$$

Clearly, for Equations (3.5) and (3.11) to be comparable, one must impose K sufficiently smaller than L . Namely, by considering a span length of the aggregation period m equal to 22 days and a number of lags K equal to 12, the series of daily realized variances are aggregated over a period of one year, thus covering the same time span as in Eq.(3.5).

A similar variation is applied to the component driving the long run correlation dynamics. The matrix $\bar{P}_t$ in Eq.(3.8) is replaced by a weighted sum of past realized

³Notice that here it is not necessary to implement the “consistent” (in the [Aielli \(2013\)](#) sense) version of the ReDCC model, as the estimation is performed with respect to all parameters in one step and hence it does not suffer from the inconsistency issue arising when targeting is used. See also Appendix 3.A for further details.

correlation matrices aggregated over the past year, i.e.

$$\bar{P}_t = \sum_{k=1}^K \phi_k(\omega_r) P_{t,k}^{(m)} \quad (3.13)$$

$$P_{t,k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} P_\tau. \quad (3.14)$$

This definition of the time-varying intercept matrix $\bar{P}_t$ does not yield a correlation matrix and consequently, R_t^{DCC} needs to be rescaled into a conditional correlation matrix. Using Equations (3.7)-(3.8), one just needs to compute

$$R_t^{DCC*} = \text{diag}\{R_t^{DCC}\}^{-1/2} R_t^{DCC} \text{diag}\{R_t^{DCC}\}^{-1/2}. \quad (3.15)$$

This new model is defined as *Additive Midas Aggregated Realized DCC (AMAReDCC)*, or *AMAReDECO* if the DECO version is used.

The second variation proposed for the specification of the conditional correlation matrix stems from the fact that the construction of R_t^{DCC} in Eq.(3.7) builds upon daily realized covariances filtered from realized standard deviations rather than conditional standard deviations. By purging the realized covariance matrices from the conditional standard deviations, one can eliminate potential heteroskedasticity left in C_t and obtain a potentially cleaner estimator. Namely, C_t can be transformed to

$$C_t^p = D_t^{-1} C_t D_t^{-1}, \quad t = 1, \dots, T, \quad (3.16)$$

and then used to compute the time-varying intercept matrix $\bar{P}_t$ in Eq. (3.7). This resembles the approach used in the two-step estimation of the DCC model of Engle (2002a), where the outer product of the first-stage (estimated) standardized residuals ξ_t is used to compute the second-stage conditional correlation matrix $R_t = \mathbb{E}_{t-1}[\xi_t \xi_t']$.

By combining the different equations proposed for the conditional variances and correlations with the two structures of the MIDAS-type filter, we obtain eight additive models that are summarized in Table 3.1.

3.2.2 Long term multiplicative component models

The component models that have been suggested in the most recent literature are mainly based on a multiplicative structure. Examples are the Spline-GARCH model of Engle & Rangel (2008), the multiplicative BEKK model of Hafner & Linton (2010), the multiplicative-DCC model of Bauwens et al. (2013) and the MIDAS-CAW model of Golosnoy et al. (2012). Our approach builds on these contributions. The proposed component models feature a multiplicative decomposition of the conditional covariance matrix S_t into a secular component $M_t = L_t L_t'$ and a short-lived component S_t^* , as

Table 3.1: **List of time-varying additive component models**

The models reported in the table feature an additive MIDAS component for correlations and are fully parametric. The rationale of each name is the following: the first letter “A” stands for additive, the second letter “M” stands for MIDAS and “ReDCC” or “ReDECO” denotes the parametric specification used to model the short term dynamics. The suffix “P” in the last four cases denotes the models built using the purged estimator introduced in Eq.(3.16). The number of parameters is equal to $5n + 3$ in all cases, where n is the number of assets. For a complete overview of the models see also Table B.1 in Appendix B.

Model	Aggregated	Purged	Equations
1 - AMReDCC			(3.4)-(3.5)-(3.7)-(3.8)
2 - AMReDECO			(3.4)-(3.5)-(3.7)-(3.9)
3 - AMAReDCC	✓		(3.4)-(3.11)-(3.7)-(3.13)-(3.15)
4 - AMAReDECO	✓		(3.4)-(3.11)-(3.7)-(3.13)-(3.9)
5 - AMReDCC-P		✓	(3.4)-(3.5)-(3.16)-(3.7)-(3.8)
6 - AMReDECO-P		✓	(3.4)-(3.5)-(3.16)-(3.7)-(3.9)
7 - AMAReDCC-P	✓	✓	(3.4)-(3.11)-(3.16)-(3.7)-(3.13)-(3.15)
8 - AMAReDECO-P	✓	✓	(3.4)-(3.11)-(3.16)-(3.7)-(3.13)-(3.15)-(3.9)

follows:

$$S_t = L_t S_t^* L_t'. \quad (3.17)$$

The matrix square root L_t can be obtained by a Cholesky factorization of the long term component matrix M_t .

The main difference with the class of additive models presented in Section 3.2.1 is that the multiplicative structure allows the matrix M_t to be modelled either via a parametric function of past realized covariances or via a nonparametric function of time. Another difference is that a unique model is specified for the long term components of the realized variances and covariances altogether, instead of different models for realized variances and correlations.

Baseline models In the baseline parametric multiplicative models, the secular component is specified as a multivariate generalization of the MIDAS filter used in Eq. (3.5). This can be written as

$$M_t = \bar{\Lambda} + \theta \sum_{l=1}^L \phi_l(\omega_r) C_{t-l}, \quad (3.18)$$

where the first term $\bar{\Lambda}$ is a $n \times n$ PDS matrix of constant parameters, θ is a positive scalar parameter and $\omega_r = (\omega_{1r}, \omega_{2r})$ is the vector of weights of the Beta function obtained under the restrictions $\omega_{1r} = 1$ and $\omega_{2r} > 1$. This long term component has $n(n+1)/2 + 2$ parameters. It is worth to emphasize that in sufficiently low dimensions this does not represent an issue, but the proliferation of parameters in large dimensions will eventually render impossible numerical maximization of the log-likelihood function. Note that by imposing $\theta = 0$ in the MIDAS specification of Eq. (3.18), the secular component becomes time invariant, being limited to the constant intercept matrix $\bar{\Lambda}$. This constant

long term structure is considered as a benchmark in the empirical application.⁴

A nonparametric formulation consists in letting the long run component matrix M_t be a smooth unknown function of the rescaled time index, i.e. $M_t = M(t/T)$, where T is the sample size of the estimation period. This assumption implies that the unconditional covariance varies over time, since using $\mathbb{E}[\mathbb{E}(C_t|\mathfrak{F}_{t-1})] = \mathbb{E}(S_t)$ and imposing $\mathbb{E}(S_t^*) = I_n$ for identification (see below), it follows that $\mathbb{E}(L_t S_t^* L_t') = L_t L_t' = M_t$. One could also assume the matrix M_t to be a smooth function of an observable variable x_t (for example a market volatility index), such that $M_t = M_t(x_{t-1})$.

The multiplicative structure in Eq. (3.17) allows for different specifications to be used for modeling the dynamics of S_t^* . In this context, BEKK, DCC and DECO parameterizations apply to the short run co-volatility components independently of the parametric or nonparametric structure assumed for the long run component. When the last two are used, S_t^* is written as

$$S_t^* = D_t^* R_t^* D_t^* \quad (3.19)$$

where $D_t^* = \{\text{diag}(S_t^*)\}^{1/2}$ is the diagonal matrix of short term conditional standard deviations and R_t^* is the corresponding short term conditional correlation matrix.

DCC and DECO allow for a separate treatment of conditional volatilities and correlations, and their scalar specifications correspond to the following equations:

$$S_{ii,t}^* = (1 - \gamma_i - \delta_i) + \gamma_i C_{ii,t-1}^* + \delta_i S_{ii,t-1}^*, \quad (3.20)$$

$$R_t^{DCC^*} = (1 - \alpha - \beta) I_n + \alpha P_{t-1}^* + \beta R_{t-1}^{DCC^*}, \quad (3.21)$$

$$R_t^{DECO^*} = (1 - \rho_t) I_n + \rho_t J_n, \quad (3.22)$$

$$\rho_t = \frac{1}{n(n-1)} \left(\iota' R_t^{DCC^*} \iota \right), \quad (3.23)$$

where

$$P_t^* = \{\text{diag}(C_t^*)\}^{-1/2} C_t^* \{\text{diag}(C_t^*)\}^{-1/2}$$

and

$$C_t^* = L_t^{-1} C_t (L_t')^{-1}. \quad (3.24)$$

The matrix C_t^* is the realized covariance matrix purged of its long term component, and the matrix P_t^* is the corresponding short term realized correlation matrix.

Finally, the scalar BEKK specification can be written as

$$S_t^* = (1 - \alpha - \beta) I_n + \alpha C_{t-1}^* + \beta S_{t-1}^*. \quad (3.25)$$

Note that, like in the additive framework, mean reversion to unity in Eq.(3.20) and $\mathbb{E}(S_t^*) = I_n$ in Equations (3.21) and (3.25) are imposed as identifying restrictions.

⁴See also Appendix C for details.

Variations of the baseline models The single variation applied to this set of models consists in the aggregation of the series of realized covariance matrices in the MIDAS filter. For completeness, this amounts to specify the long term component M_t as

$$M_t = \bar{\Lambda} + \theta \sum_{k=1}^K \phi_k(\omega_r) C_{t,k}^{(m)}, \quad (3.26)$$

$$C_{t,k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} C_{\tau} \quad (3.27)$$

where the constants m and K are set in the way discussed after Eq.(3.11).

In total, we have defined nine models in the multiplicative family. The baseline models are either labeled *Multiplicative Midas Realized DCC* (MMReDCC) or *Multiplicative Kernel Realized DCC* (MKReDCC), with variations according to the assumed short run model type (DECO or BEKK) and MIDAS filter definition. A summary of these models is given in Table 3.2.

Table 3.2: **List of time-varying multiplicative component models**

The models reported in the table feature a multiplicative component structure. The rationale of each name is the following. The first letter 'M' stands for multiplicative while the second letter, "M" or "K", denotes the specification adopted for the estimation of the long run component, respectively given by a nonparametric kernel smoother ('K') or a MIDAS polynomial ('M'). In the latter case, the third letter "A", if present, refers to the use of the 'aggregated' temporal scheme in the MIDAS scheme. The labels "ReDCC", "ReDECO" or "ReBEKK" denote the parametric specification used to model the short term dynamics. The benchmarks used for this class are obtained as special cases of the parametric models and are denominated Constant ("C") Realized DCC, DECO or BEKK models (CReDCC, CReDECO and CReBEKK for brevity). For a complete overview of the models see also Table B.1 in Appendix B.

	Model	Num. pars	Aggregated	Parametric	Semiparametric	Equations
9 -	MKReDCC	2n+3			✓	(3.29)-(3.20)-(3.21)
10 -	MKReDECO	2n+3			✓	(3.29)-(3.20)-(3.23)
11 -	MKReBEKK	3			✓	(3.29)-(3.25)
12 -	MMReDCC	$\lfloor n(n+1)/2 \rfloor + 2n + 3$		✓		(3.18)-(3.20)-(3.21)
13 -	MMReDECO	$\lfloor n(n+1)/2 \rfloor + 2n + 3$		✓		(3.18)-(3.20)-(3.23)
14 -	MMReBEKK	$\lfloor n(n+1)/2 \rfloor + 3$		✓		(3.18)-(3.25)
15 -	MMAReDCC	$\lfloor n(n+1)/2 \rfloor + 2n + 3$	✓	✓		(3.26)-(3.20)-(3.21)
16 -	MMAReDECO	$\lfloor n(n+1)/2 \rfloor + 2n + 3$	✓	✓		(3.26)-(3.20)-(3.23)
17 -	MMAReBEKK	$\lfloor n(n+1)/2 \rfloor + 3$	✓	✓		(3.26)-(3.25)

3.3 Estimation

The parametric models can be estimated by the method of maximum likelihood (ML) in one step. In general, given the Wishart assumption made on C_t , the log-likelihood function for T observations, $\ell_T(\Phi)$, where Φ is the finite-dimensional vector of model parameters to be estimated, is expressed as follows:

$$\ell_T(\Phi) = \bar{c}T - \frac{\nu}{2} \sum_{t=1}^T \left[\log(|S_t|) + \text{tr}(S_t^{-1}C_t) \right], \quad (3.28)$$

where $\bar{c}$ depends only on ν , n , and C_t . Hence, the log-likelihood function depends on Φ (through the matrix S_t) only via the last two terms. This formula is general, in the sense that it does not depend on the particular specification assumed for the long term and the short term components. In particular it also holds for the constant long run component models.

The last two terms on the right hand side of Eq.(3.28) are linear in the parameter ν . Hence, the first order conditions for the estimation of the parameter vector Φ do not depend on ν , implying that Φ can be estimated independently of the value of ν . In addition, it is known that the estimator based on the maximization of the Wishart log-likelihood function has a QML interpretation, because if the conditional expectation of C_t is correctly specified, the score of the log-likelihood function in Eq. (3.28), evaluated at the true value of the parameters, is a martingale difference sequence (MDS). Thus, under appropriate regularity conditions, the ML estimator is consistent even if the underlying distribution of C_t is not Wishart. We refer to Bauwens et al. (2012) and Noureldin et al. (2012) for technical details regarding the MDS property of the score and QML asymptotic results in this context.

The motivation for applying estimation in one step is twofold. First, the one-step maximization of the log-likelihood function simplifies inference, since one can easily compute (robust) standard errors and model selection criteria, such as the Akaike (AIC) and the Bayesian information criterion (BIC). This makes the in-sample comparison of different models easy. Second, estimation strategies based on sequential maximization of different log-likelihood components can provide inefficient estimators whose asymptotic distributions can be difficult to compute.

Nevertheless, as the one-step estimation can be computationally challenging in large-dimensional applications, multi-step estimation procedures could eventually be used to obtain estimates of model parameters. Alternatively, if possible, targeting strategies could be applied to pre-estimate some of the parameters, thus reducing computational burden. A detailed exploration of these estimation techniques goes beyond the scope of the present chapter but it will be the main object of interest of Chapter 4.

In the case of the multiplicative models that have a nonparametric long term component, a two-step approach must be used: first, a kernel smoother is used to estimate the long run component M_t , and then, conditional on the estimate of the M_t matrices, the parameters driving the short term dynamics are estimated by ML. To estimate the matrices M_t non-parametrically as a function of rescaled time, we make use of the Nadaraya-Watson kernel estimator, which is consistent under general conditions; see for example Härdle (2004). It takes the following form:

$$M_t(\tau) = \frac{\sum_{t=1}^T K_h\left(\frac{t}{T} - \tau\right) C_t}{\sum_{t=1}^T K_h\left(\frac{t}{T} - \tau\right)}, \quad (3.29)$$

where $\tau \in [0, 1]$, $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is the one-sided Gaussian kernel function and h the bandwidth parameter determining the amplitude of the movements captured by the

long term component. The bandwidth selection is performed by the least squares cross-validation method. To implement this, we use six-month rolling covariance matrices as the reference for computing the squared differences and adopt the penalizing function approach involving the bandwidth selector of [Rice \(1984\)](#). This estimator uses the same bandwidth for all the elements of the realized covariance matrices, even if this could be relaxed by smoothing separately variances and correlations with a different bandwidth for each diagonal element and a single one for the whole correlation matrix. Since we use the model for forecasting purposes, the use of a one-sided version of the kernel estimator is coherent with the idea that the information set available at time t consists only of data up to time $t - 1$.

3.4 Evidence from real data

This section presents the results of an empirical application to a time series of realized covariance matrices for U.S. stocks. First, we present the dataset being used and the forecasting framework adopted for constructing one-step ahead covariance forecasts. Then we provide full-sample estimation results and compare the estimated models in terms of information criteria. Finally, we perform an out-of-sample forecasting comparison of the proposed models vis-a-vis the corresponding benchmark models characterized by a constant long run component.

3.4.1 Data and forecasting scheme

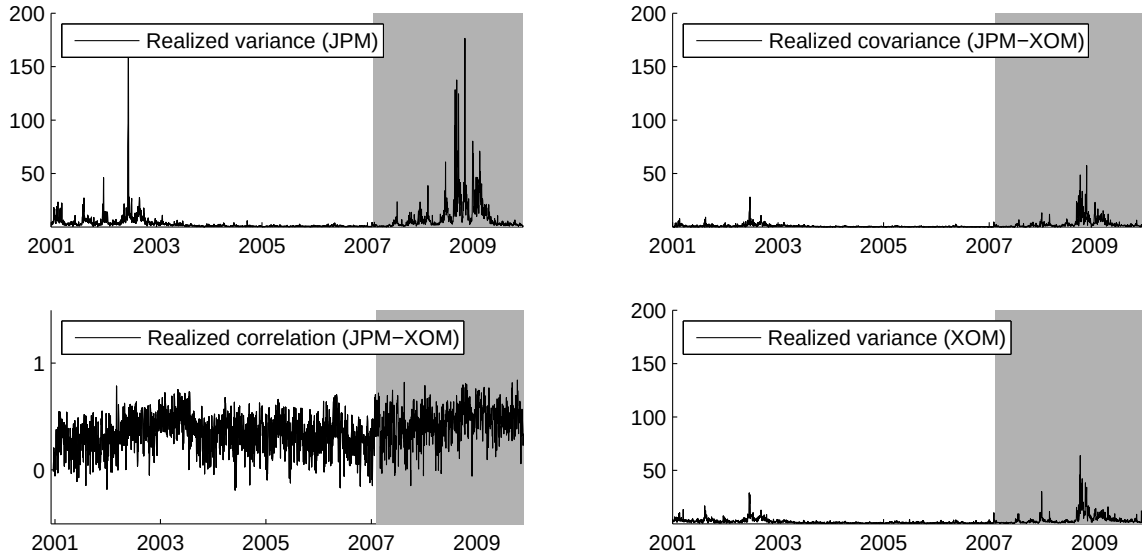
The empirical analysis is based on a series of daily realized covariance matrices of ten stocks included in the Dow Jones Industrial Average (DJIA) index. The dataset has been previously used by [Noureldin et al. \(2012\)](#) and can be downloaded from the *Oxford Man Institute Realized Library*. The data have been cleaned according to the procedure of [Barndorff-Nielsen et al. \(2009\)](#) and the open-to-close realized covariance estimator has been constructed on five-minute intraday returns aggregated with subsampling. We refer to the cited paper for a detailed explanation on the features of the dataset and the construction of the realized estimator.

The included stocks are Bank of America (BAC), JP Morgan (JPM), International Business Machines (IBM), Microsoft (MSFT), Exxon Mobil (XOM), Alcoa (AA), American Express (AXP), Du Pont (DD), General Electric (GE) and Coca Cola (KO). The dataset runs from February 1, 2001 to December 31, 2009, providing a total of 2240 observations. Descriptive statistics are provided in Table 3.7 in Appendix 3.A. Figure 3.2 shows a representative example of time series plots of the realized variances of two stocks, and the corresponding realized covariance and correlation.

Our aim is to evaluate both the in- and out-of-sample performance of the dynamic component models against benchmark models that are based on constant long run components for volatilities and correlations. For the first evaluation, we estimate the

Figure 3.2: Realized series for JPM-XOM

JP Morgan (JPM) and Exxon Mobil (XOM) realized volatility, realized covariance and realized correlation over the full-sample period 1/02/2001 – 31/12/2009. Out-of-sample period shaded in gray.



proposed models using the complete sample and then we compare the fitted models by means of information criteria. For the second, one-day ahead daily covariance matrix forecasts are constructed. Specifically, the full dataset is divided into two different periods:

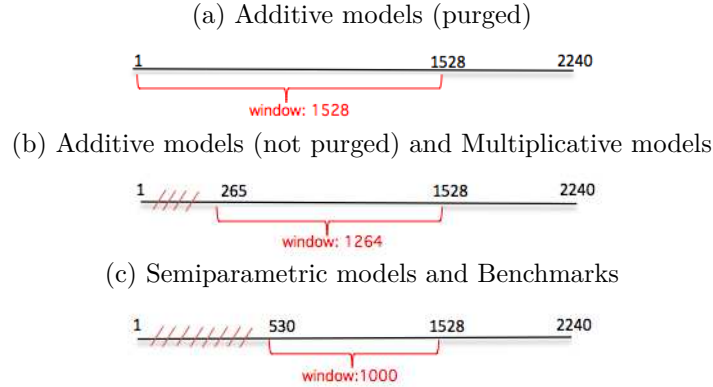
- Period I is the in-sample set for $t = 1, \dots, 1528$. It corresponds to the relatively calm period from February 2001 to December 2006 and is reserved for the model initial estimation (before forecasting).
- Period II is the out-of-sample set comprising the remaining 712 observations used for the forecast evaluations. It is characterized by a higher unconditional volatility level stemming from the 2008-2009 crisis events included in the last part of the sample.

Forecasts are constructed using a fixed rolling window scheme that satisfies the assumptions required by the MCS procedure and allows the comparison of nested models. In this respect, one clarification is necessary. The initialization of the MIDAS filter of the models discussed in Sections 3.2.1–3.2.2 requires some realized covariance matrices to be used as starting values; this amounts to reserve 529 initial observations for the additive models with purging and 264 for both the parametric multiplicative models and the remaining models in the additive group. No such loss of initial observations affects the nonparametric and the benchmark models, as their structure does not involve a MIDAS filtering scheme. Consequently, in order to obtain comparable in-sample statistics and parameter estimates, the estimation starting point and the in-sample window size need to be set differently according to the specific model structure.

The note in Figure 3.2 summarizes the implemented procedure. For each model,

Figure 3.3: Implemented rolling window setting

The length of the in-sample rolling windows is set in order to get an effective number of 1000 observations and, depending on the specific model structure, three different window sizes are used. Specifically, additive models involving a purged estimator need 528 observations for initialization (Figure(a)) and hence they require a longer in-sample window of 1528 trading days. The remaining additive models and all the parametric multiplicative models (Figure(b)) need half the number of initial observations, thus the window is shifted forward of 264 days for a total of 1264 days. The models not incorporating a MIDAS component in the long-term structure (Figure(c)) do not lose any observation, hence the initial estimation starting point is shifted forward accordingly. T denotes the residual sample size after the trimming and is equal to 1712 trading days for all the models. The out-of-sample period starts at $t = 1529$ and covers the last 712 observations of the sample.



712 forecasts are computed. We estimate the parameters of interest using the in-sample window and then we produce forecasts for the following 20 days, which approximately correspond to one month of trading. The estimation window is then shifted forward by 20 observations and the model is re-estimated. The new estimates are used to generate covariance forecasts for the subsequent 20 days and the whole procedure is repeated until all data until the end of December 2009 have been used. Irrespective of the initialization period, successive forecasts are effectively based on the most recent 1000 observations. Given the full sample of 2240 observations, the number of re-estimations of each model is equal to 36 but the last set of forecasts covers 12 days (period $t = 2228, \dots, 2240$) instead of 20.

3.4.2 Full sample results

To evaluate the relative performance of our models in fitting the data we estimate them on the whole sample. Table 3.8 in Appendix 3.A reports estimates of the parameters. The estimates of the MIDAS parameter θ of the multiplicative models are significant at the level of 5 percent at least, indicating the relevance of such a component. In most cases, when a time-varying long run component is included, the persistence of the correlation process (measured by the sum of α and β) is smaller than when the long run level is constant.

Table 3.3 reports the values of the maximized log-likelihoods for all the fitted models. For the fully parametric model specifications we also report the associated AIC and BIC values while the values of these criteria are not reported for models incorporating a nonparametric long run component. This choice is due to the fact that the penalty term appearing in the AIC and BIC formulas depends on the number of estimated parameters which cannot be explicitly determined since the long run component is estimated by

Table 3.3: **Information criteria of the models**

“Num. pars” (N_p) is the number of estimated parameters in each model. Maximized log-likelihood values are reported in the third column. Information criteria in columns 4 and 5 are computed as follows: $AIC = -2\text{LogLik} + 2N_p$ and $BIC = -2\text{LogLik} + N_p \log(T)$. The values in bold correspond to the globally best performing model while values in italics denote the best model in each category.

Models	Num. pars	LogLik	AIC	BIC
<i>Additive models</i>				
1 - AMReDCC	53	-17661	35429	35717
2 - AMReDECO	53	-18093	36292	36581
3 - AMAReDCC	53	-17675	35456	35744
4 - AMAReDECO	53	-18110	36327	36595
5 - AMReDCC-P	53	<i>-17657</i>	<i>35419</i>	<i>35709</i>
6 - AMReDECO-P	53	-18078	36261	36551
7 - AMAReDCC-P	53	-17695	35785	35800
8 - AMAReDECO-P	53	-18112	36619	36633
<i>Multiplicative models</i>				
9 - MKReDCC	–	-17611	–	–
10 - MKReDECO	–	-17726	–	–
11 - MKReBEKK	–	-17771	–	–
12 - MMReDCC	79	-17662	35481	35912
13 - MMReDECO	79	-17672	35501	35932
14 - MMReBEKK	59	-17805	35727	36049
15 - MMAReDCC	79	-17668	35494	35924
16 - MMAReDECO	79	-17696	35549	35980
17 - MMAReBEKK	59	-17790	35698	36019
<i>Benchmark models</i>				
18 - CReDCC	77	<i>-17780</i>	<i>35714</i>	36133
19 - CReDECO	77	-18184	36523	36942
20 - CReBEKK	57	-17846	35805	<i>36116</i>
21 - ReDCC	77	-17881	35915	36335

means of a kernel smoother. Nevertheless, we provide in the table the values of the maximized log-likelihood functions of the three models having a nonparametric long run component.

The first interesting comparison is between the relative performance of the time-varying long run component models and their benchmarks. Except for some cases, the maximum values of the log-likelihoods are not comparable due to different numbers of parameters. In these cases we rely on the comparison of the estimated AIC and BIC values. Whether the criterion is AIC or BIC, each benchmark model is outperformed by the corresponding time-varying counterpart(s), a clear indication that assuming a constant long term structure for volatilities and correlations can be too restrictive when the market economic conditions change over time, as it is the case in the covered sample period: e.g. AIC of 35481 for MMReDCC, 35714 for CReDCC; BIC of 36049 for MMReBEKK, 36116 for CReBEKK; AIC of 35419 for AMReDCC-P, 35915 for ReDCC.

Other comparisons are of interest. Within the additive model class, the comparison can be done by the log-likelihood values since all models have the same number of parameters. The models with a long term non aggregated structure strongly outperform their competitors, while the use of the purged estimator improves the fit only for the models with a long term non aggregated structure: the AMReDCC-P has the highest log-likelihood value but the improvement of 4 points over the AMReDCC is not spectacular.

In the multiplicative parametric class, there is no clear ranking between the models including different MIDAS filters (the difference in absolute value between one model and its corresponding aggregated version is only 15 points on average in terms of log-lik. values).

In a global ranking of the fully parametric models, three additive models (AMReDCC-P, AMReDCC, AMAReDCC) are ranked as the best fitting parametric models by AIC and BIC.

Concerning the specification of the short run component, within each class, DECO and BEKK models are always dominated by the corresponding DCC versions.

3.4.3 Forecasting results

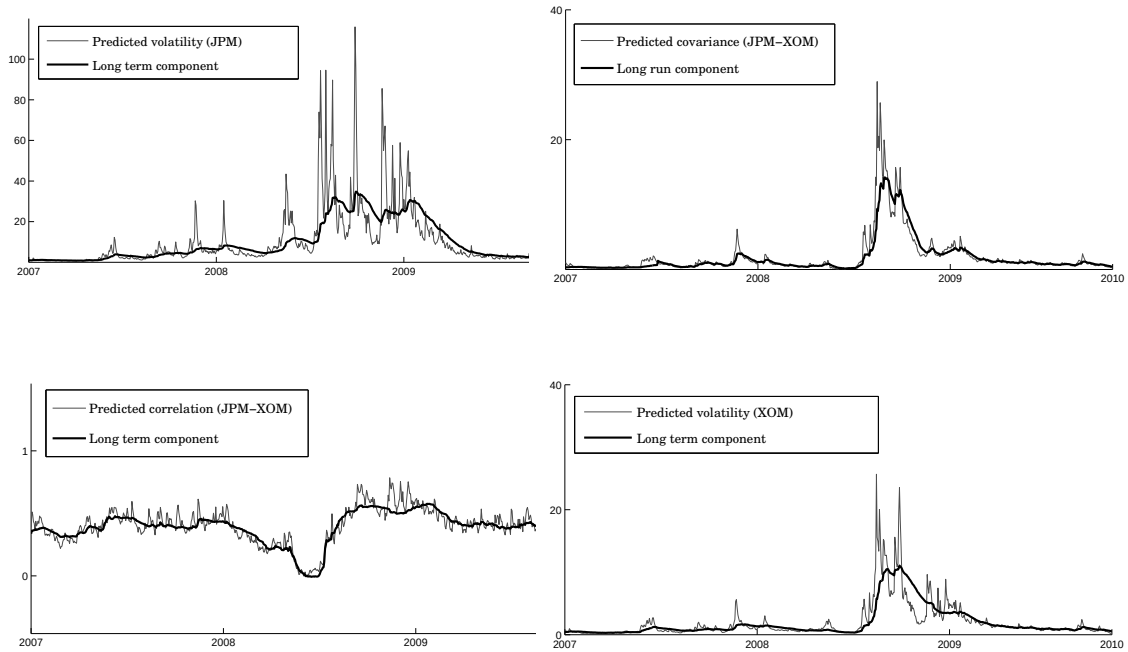
The forecasting ability of the set of proposed models is evaluated over a series of 712 out-of-sample predictions. Figure 3.4 shows representative plots of the predicted variances, covariance and correlation for the JP Morgan (JPM) and Exxon Mobil (XOM) stocks together with their respective long-run MIDAS components obtained under the MMReDCC model. The forecast exercise is performed in three different ways. Firstly, we assess the predictive performance of the models under different loss functions and evaluate the significance of loss function differences by means of the Model Confidence Set (MCS) approach. Secondly, we evaluate the out-of-sample hedging performance of the different models by constructing minimum variance portfolios. Last, we test the models ability to accurately forecast portfolio Value-at-Risk.

Model Confidence Set

The Model Confidence Set (MCS) of Hansen et al. (2011) is the first method employed to compare the forecasting performance of the proposed models. Let M^0 denote the initial set of models for which we compute the series of one-step ahead conditional covariance forecasts for period t , denoted by H_t^i , where i denotes the i -th model. The MCS is based on an iterative procedure that requires sequential testing of equal predictive accuracy (EPA); this implies that the set of candidate models is sequentially trimmed by deleting those that are found to be statistically inferior within M^0 . At a given level of confidence, the MCS contains the single model or the set of models having the best

forecasting performance. The advantage of this procedure is that it does not necessarily require to select a privileged benchmark model.

Figure 3.4: Predicted series with corresponding long term component
The figure shows predicted variance, covariance and correlation of JP Morgan (JPM) and Exxon Mobil (XOM) stocks from MMReDCC model along with the corresponding long term component (black straight line).



At the heart of the method there is a forecast loss measure. The MCS final selection is based on the ordering implied by the loss function used to evaluate the deviations of each model predictions from the true conditional covariance matrix, denoted by Σ_t . Given that the true conditional covariance is latent even ex-post, we rely on an unbiased proxy of Σ_t . Our choice falls on the 5-minutes realized covariance estimator, $\hat{\Sigma}_t$, which is a more efficient estimator than the one based on the outer product of returns under the fairly general assumptions of absence of microstructure noise and other biases; see for example [Barndorff-Nielsen & Shephard \(2001\)](#), [Aït-Sahalia et al. \(2005\)](#), and [Zhang \(2011\)](#).

As regards the choice of loss functions, recent research on the consistent ranking of volatility forecasts by [Patton \(2011\)](#), [Patton & Sheppard \(2009\)](#), and [Laurent et al. \(2013\)](#) has highlighted that care needs to be taken during the selection process in order to avoid unintended results. We therefore employ matrix loss functions that are robust to noisy proxies. In other words, on average, they are expected to provide the same ranking between two forecasts independently of whether the true conditional covariance or a conditionally unbiased proxy is used. We opt for using several robust loss functions instead of a single one, in order to assess the sensitivity of the MCS to different functions. They correspond to Euclidean and Frobenius distances, Mean Square Forecast Error (MSFE), QLIKE, Stein and von Neumann divergence (VND). Their definition is recalled in [Appendix 3.A](#). We follow [Hansen et al. \(2011\)](#) and compute the MCS at both the

75% and 90% confidence levels. The block-length bootstrap parameter and the number of bootstrap samples used to obtain the distribution under the null are set equal to 2 and 10000, respectively.

Table 3.4: **Model confidence sets at 90% and 75% levels.**

The symbol '◦' denotes models included in the 90% MCS while '•' denotes models included also in the 75% MCS. Corresponding p -values are provided in Tables 3.10-3.11 of Appendix 3.A. Loss functions defined in Section 3.4.3 and detailed in Table 3.9.

Models	Euclidean	Frobenius	MSFE	QLIKE	Stein	VND
<i>Additive models</i>						
1 - AMReDCC		◦	◦			
2 - AMReDECO		◦	◦			
3 - AMAReDCC		◦	◦			
4 - AMAReDECO		◦	◦			
5 - AMReDCC-P		◦	◦			•
6 - AMReDECO-P		◦	◦			
7 - AMAReDCC-P		◦	◦			
8 - AMAReDECO-P						
<i>Multiplicative models</i>						
9 - NPreDCC	•	•	•			•
10 - NPreDECO						
11 - NPreBEKK						
12 - MMRReDCC				•	•	•
13 - MMRReDECO				•	•	
14 - MMRReBEKK	◦	◦	◦			•
15 - MMAReDCC						◦
16 - MMAReDECO						
17 - MMAReBEKK		◦	◦			
<i>Benchmark models</i>						
18 - CReDCC		◦	◦			
19 - CReDECO						
20 - CReBEKK	◦	◦	◦			
21 - ReDCC						

The MCS results at the 90% and 75% confidence levels are given in Table 3.4. At the 90% level, two benchmark models, CReDCC and CReBEKK, are included in the MCS resulting from the Frobenius and MSFE loss functions, and CReBEKK is included also in the Euclidean MCS. At the 75%, no benchmark model is included in any MCS. These results provide reasonable support to the conjecture that accounting for a time-varying long run component may contribute to improve forecasting performance.

At the 90% level, at least two models with time-varying long run components are included in the MCS for all loss functions. The highest number of models (ten) is included for the Frobenius and the MSFE loss functions. That number is considerably reduced for the other loss functions and this is also the case for all loss functions at the 75% confidence level. The most striking result is the inclusion of the MKReDCC model in the MCS of four loss functions at both levels, and at 75% it is the single model of the MCS based on the Euclidean, Frobenius, and MSFE functions. Multiplicative parametric models slightly outperform the additive models, except at the 90% level for the Frobenius and MSFE functions. DCC and BEKK-type models are more often

selected than DECO-type models.

Out-of-sample hedging performance

In choosing among different competing models, practitioners are willing to employ the best one whose performance can be evaluated in an economically meaningful way. In portfolio management, for example, they are interested in the model providing portfolios with the lowest variance among a set of models. We accomplish this kind of comparison adopting the method proposed by [Colacito & Engle \(2006\)](#) pertaining to minimum variance portfolio management. Namely, by using out-of-sample covariance forecasts from the set of competing models, we consider two minimum variance portfolios: the global minimum variance portfolio, and the portfolio with minimum variance under the constraint of achieving a target expected return. The unconstrained portfolio weights are obtained as solution to the optimization problem

$$\min_{w_t} w_t' H_t w_t \quad s.t. \sum_{j=1}^n w_{t,j} = 1,$$

where w_t is the vector of portfolio weights for time t chosen at time $t - 1$, H_t denotes the conditional covariance forecast from a generic model, and the only requirement is on the vector of weights to sum up to unity. The second portfolio is achieved by adding an additional constraint on the vector of expected returns μ , i.e. $w_t' \mu \geq q$. In both cases, short sales are not allowed. As pointed out by [Colacito & Engle \(2006\)](#), the portfolio volatility is smallest for the correctly specified covariance matrix for any vector of expected returns. Hence, for computational ease, we set the conditional mean return vector equal to the historical mean and, like [Engle & Kelly \(2012\)](#), we impose a target expected annual return of 10%.

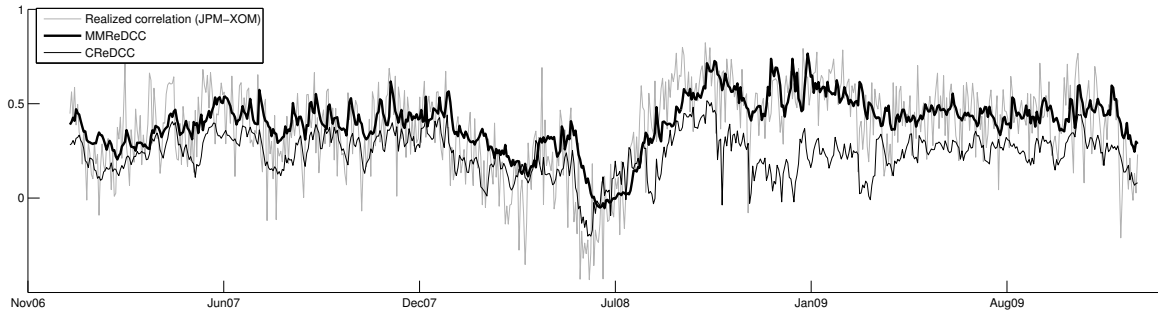
As a result, we get series of 712 out-of-sample daily portfolio returns and their corresponding daily variances for each model. According to the initial statement, a superior model is required to produce optimal portfolios with lower variance realizations. In order to test for the significance of the differences between portfolio variances, we employ a [Diebold & Mariano \(1995\)](#) test between the overall best performing model and the first best model in the benchmark group. Table 3.5 reports standard deviations of the forecasted portfolio returns time series along with each model position in the ranking. Results for the constrained portfolios are shown in the second column of the table. Noticeably, there are no benchmark models up to the eighth position.

The five best performing models belong to the multiplicative class, and the MM-ReDCC achieves the lowest global minimum variance portfolio with a standard deviation of 1.0711. This improves at the 1% significance level over the first best benchmark, the CReDCC, which achieves a standard deviation of 1.0815. A plausible reason for this difference stems from the comparison of the correlations extracted from both models, as shown in Figure 4 together with the realized correlation. The MMReDCC series tracks more closely the realized correlation than CReDCC. In particular, the rise in the correlations at the end of 2008 persists for the MMReDCC like the realized series,

whereas the CReDCC series reverts to a lower level. Hence, the CReDCC model tends to underestimate correlations during the more volatile period, thus leading to a less appreciable gain from portfolio diversification.

Figure 3.5: Predicted conditional correlations JPM-XOM

The figure shows predicted correlations of JP Morgan (JPM) and Exxon Mobil (XOM) stocks from MMRéDCC and CReDCC models over the forecasting period (December 2006 to December 2009)



The other constant long term models have broadly the same inferior performance, except that the CReBEKK model slightly improves over its competitors involving a BEKK-type structure. Not really surprisingly, as it reminds the finding obtained in the MCS evaluation, the ReDCC closes the ranking with the highest standard deviation of 1.1218. Similar results are found for the constrained portfolios in the fourth column. The MMRéDCC model confirms its predominance (1.0958) and significantly improves over the first best constant long run model (CReBEKK, 1.1102) at the 1% significance level. The other benchmarks are all ranked in the last positions.

Briefly, we can interpret minimum variance portfolios results from this forecasting exercise as providing clear evidence that there can be hedging benefits from employing models that account for a time-varying long run component. In some cases, these benefits can be remarkable, depending on the chosen long run component-type structure or the short run multivariate specification.

Portfolio VaR forecasting

In this last application, we consider the forecasting of portfolio Value-at-Risk (VaR). Our analysis is concerned with the forecasting of the *long* side of the daily VaR. This corresponds to the VaR level for traders having long positions in their asset holdings, and thus the predictive power of a specific model is related to its ability to model large negative returns. For simplicity, we abstract from the Markovitz optimization setting employed in the previous application, considering only equally-weighted portfolios.

For each model, the portfolio VaR at level α on day t , conditional on the information

Table 3.5: **Out-of-sample hedging performance**

The table reports out-of-sample standard deviations of the portfolios return time series along with the corresponding rank in parenthesis. “Unconstrained” refers to the global minimum variance portfolio, while “constrained” refers to the minimum variance portfolio with 10% expected return. The best performing model in each column is in bold. A Diebold-Mariano test is performed to compare the variances achieved by the best model against the first best performing model within the benchmarks in the same column. The best model is accompanied by * or ** if the difference is significant at the 5% or 1% level, respectively.

Models	Unconstrained		Constrained	
	Std. Deviation	Rank	Std. Deviation	Rank
<i>Additive models</i>				
1 - AMReDCC	1.07963	(7)	1.10497	(8)
2 - AMReDECO	1.08335	(9)	1.11160	(14)
3 - AMAReDCC	1.08545	(13)	1.11072	(13)
4 - AMAReDECO	1.08672	(15)	1.11683	(15)
5 - AMReDCC-P	1.07792	(6)	1.10341	(6)
6 - AMReDECO-P	1.08590	(14)	1.11440	(16)
7 - AMAReDCC-P	1.08487	(12)	1.10968	(10)
8 - AMAReDECO-P	1.08782	(16)	1.11441	(17)
<i>Multiplicative models</i>				
9 - MKReDCC	1.07190	(2)	1.09910	(4)
10 - MKReDECO	1.08404	(11)	1.10465	(7)
11 - MKReBEKK	1.11439	(20)	1.13175	(20)
12 - MMRReDCC	1.0711 **	(1)	1.0958 **	(1)
13 - MMRReDECO	1.07307	(3)	1.09874	(3)
14 - MMRReBEKK	1.08886	(18)	1.10897	(9)
15 - MMAReDCC	1.07308	(4)	1.09687	(2)
16 - MMAReDECO	1.07790	(5)	1.10264	(5)
17 - MMAReBEKK	1.08958	(19)	1.11053	(12)
<i>Benchmark models</i>				
18 - CReDCC	1.08149	(8)	1.11168	(18)
19 - CReDECO	1.08357	(10)	1.13073	(19)
20 - CReBEKK	1.08858	(17)	1.11026	(11)
21 - ReDCC	1.12175	(21)	1.14693	(21)

available at time $t - 1$, is computed as:

$$VaR_t(\alpha) = z_\alpha \sqrt{w' H_t w},$$

where w is the given n -dimensional vector of equal weights, H_t is the forecasted conditional covariance matrix for a generic model, and z_α is the $\alpha\%$ left-quantile of the standard normal distribution. The same analysis was done assuming the more flexible Student distribution, but as this did not lead to significant improvements, results are not reported. VaR at levels α equal to 5%, 2.5% and 1% are forecasted, and their performance is then assessed using two statistical backtesting methods. First, the Likelihood Ratio Conditional Coverage (LR_{cc}) test of Christoffersen (1998) is used; its construction

relies on the so-called hit function, or indicator function, obtained as follows:

$$I_t(\alpha) = \begin{cases} 1 & \text{if } w'r_t \leq VaR_t(\alpha) \\ 0 & \text{if } w'r_t > VaR_t(\alpha). \end{cases}$$

According to [Finger \(2005\)](#), good VaR models are capable of reacting to changing volatility and correlations in a way that exceptions occur independently of each other, whereas bad models tend to produce a sequence of consecutive exceptions. Christoffersen's test accounts for both properties of a good VaR model, namely the correct failure rate and the independence of exceptions. The LR_{cc} test statistic is χ_1^2 distributed. The second method is the regression-based test of [Engle & Manganelli \(2004\)](#), also known as the Dynamic Quantile (DQ) test. Instead of directly considering the hit sequence, the test is based on its associated quantile process $H_t(\alpha) = I_t(\alpha) - \alpha$, formally expressed as

$$H_t(\alpha) = \begin{cases} 1 - \alpha & \text{if } I_t = 1 \\ -\alpha & \text{if } I_t = 0. \end{cases}$$

Its purpose is to link the current margin exceedances to past violations or past information and subsequently testing for different restrictions on the parameters of the regression. We run the regression $H_t(\alpha) = \delta + \sum_{k=1}^K \beta_k H_{t-k}(\alpha) + \epsilon_t$ for $K = 3$ and we test the joint hypothesis $H_0(DQ_{cc}) : \delta = \beta_1 = \dots = \beta_K = 0$. This assumption coincides with the null of Christoffersen's LR_{cc} test. It is also possible to split the test and separately test the independence hypothesis and the unconditional coverage hypothesis, respectively as $H_0(DQ_{ind}) : \beta_1 = \dots = \beta_K = 0$ and $H_0(DQ_{uc}) : \delta = 0$. Empirical results from the tests are given in Table 3.6. For each VaR level, we report the p -value corresponding to each test. The first column of each panel depicts the results for the LR_{cc} test while the last three columns show the results for the DQ_{uc} , DQ_{ind} and DQ_{cc} tests. Across the different VaR panels, the LR_{cc} and DQ_{cc} tests basically tell the same story. Rejections of the first test at the 5% level correspond to rejections of the second, with very few exceptions.

Of main interest, the last four rows of the table report p -values from the benchmark models. Results are quite homogeneous among the three VaR panels and, at least for the multiplicative models, they are considerably inferior to those obtained by the corresponding time-varying counterparts.

Irrespective of their model structure, the tests for the benchmark models lead to rejections in a vast majority of cases. On the contrary, the tests for the multiplicative component models show a remarkably better performance. Both the nonparametric and the parametric versions pass all the tests for $\alpha = 5\%$ showing occasional rejections of the DQ_{uc} test for the most extreme quantiles. Apparently, the flexibility of the multiplicative structure allows to adequately model the left tail of the distribution and to deliver superior VaR forecasts. This holds for any type of short term (DCC, BEKK

Table 3.6: **Likelihood Ratio Test and Dynamic Quantile Test Results**

The table reports resulting p-values of the Likelihood Ratio (LR) Conditional Coverage Test and of the Dynamic Quantile (Unconditional Coverage, Independence and Conditional Coverage) test for the equal weighted portfolios Value-at-Risk at confidence levels $\alpha=1\%$, 2.5% and 5%. Values in bold denote significance at the 5% level.

Models	VaR 1%				VaR 2.5%				VaR 5%			
	LR		DQ		LR		DQ		LR		DQ	
	cc	uc	ind	cc	cc	uc	ind	cc	cc	uc	ind	cc
<i>Additive models</i>												
1 - AMReDCC	0.14	0.02	0.50	0.07	0.27	0.07	0.25	0.11	0.02	0.00	0.23	0.01
2 - AMReDECO	0.02	0.00	0.36	0.00	0.26	0.07	0.25	0.11	0.01	0.00	0.06	0.00
3 - AMAReDCC	0.24	0.06	0.55	0.15	0.36	0.12	0.28	0.18	0.07	0.02	0.35	0.05
4 - AMAReDECO	0.04	0.00	0.41	0.01	0.12	0.02	0.20	0.04	0.02	0.00	0.23	0.01
5 - AMReDCC-P	0.40	0.14	0.60	0.30	0.26	0.07	0.25	0.11	0.03	0.01	0.27	0.02
6 - AMReDECO-P	0.04	0.00	0.41	0.01	0.27	0.07	0.25	0.11	0.00	0.00	0.71	0.00
7 - AMAReDCC-P	0.14	0.02	0.51	0.07	0.37	0.12	0.28	0.18	0.06	0.02	0.34	0.04
8 - AMAReDECO-P	0.04	0.00	0.41	0.01	0.18	0.04	0.22	0.07	0.01	0.00	0.47	0.01
<i>Multiplicative models</i>												
9 - NPreDCC	0.40	0.13	0.60	0.29	0.12	0.02	0.19	0.04	0.17	0.05	0.94	0.14
10 - NPreDECO	0.13	0.02	0.50	0.07	0.49	0.19	0.31	0.27	0.25	0.09	0.54	0.22
11 - NPreBEKK	0.14	0.17	0.01	0.01	0.12	0.03	0.97	0.09	0.12	0.09	0.19	0.11
12 - MMRReDCC	0.40	0.14	0.60	0.29	0.76	0.42	0.37	0.50	0.06	0.05	0.14	0.05
13 - MMRReDECO	0.95	0.73	0.74	0.89	0.62	0.29	0.34	0.38	0.12	0.09	0.19	0.11
14 - MMRReBEKK	0.59	0.27	0.65	0.50	0.27	0.07	0.25	0.11	0.25	0.09	0.55	0.22
15 - MMRReDCC	0.24	0.06	0.55	0.15	0.76	0.43	0.38	0.50	0.12	0.09	0.19	0.11
16 - MMRReDECO	0.59	0.27	0.65	0.50	0.62	0.29	0.35	0.38	0.12	0.09	0.19	0.11
17 - MMRReBEKK	0.40	0.13	0.60	0.29	0.27	0.07	0.25	0.11	0.46	0.43	0.35	0.48
<i>Benchmark models</i>												
18 - CReDCC	0.01	0.00	0.31	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.23	0.00
19 - CReDECO	0.00	0.00	0.09	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.12	0.00
20 - CReBEKK	0.40	0.13	0.60	0.03	0.12	0.02	0.19	0.03	0.06	0.01	0.34	0.04
21 - ReDCC	0.07	0.01	0.45	0.02	0.01	0.00	0.09	0.00	0.00	0.00	0.14	0.00

DECO) specification.

It clearly appears that the additive class fails in predicting the VaR adequately. This is particularly visible for the 5% VaR, where the p -values for the null hypothesis of the various tests are often smaller than 0.05 for most models.

However, by looking at the results of the DQ_{ind} test across the board, violations do not appear to be dependent. For VaR at levels $\alpha = 2.5\%$ and $\alpha = 1\%$, models including a DCC specification show a slight improvement in capturing the conditional coverage of both tests, but the additive class still appears to be almost uniformly rejected. The only exception is the AMAReDCC model, which passes all the tests with a single rejection at the 5% level for the 1% VaR.

The conclusion that can be drawn is that, for equally weighted portfolios, reliable VaR forecasts can be obtained under the simple assumption of conditionally normally standardized portfolio returns, by using time-varying long run component models that include an appropriately chosen long term structure.

3.5 Conclusions

We propose a new set of component models allowing for time variation in the long run levels of realized variances and correlations. Our modelling framework allows for the secular component to enter the model structure either in additive fashion, as a time-varying intercept in the conditional correlation model, or as a multiplicative factor. In the latter case, it is specified parametrically, using a MIDAS-type specification, or non-parametrically, by means of a matrix-variate smoother. Different specifications for modelling the short-term dynamics are also investigated and compared. As a general finding, additive-type models have good in-sample fits while multiplicative models tend to be preferred out-of-sample. While the parametric long term component can be easily interchanged with a nonparametric smoother without drastically altering the performance of the model, the choice of the multivariate GARCH specification plays a crucial role for accurately capturing the short-term dynamics and, among the three alternative proposed, the DCC appears to be prevailing. In all cases, the greatest strength of these models is to overcome the usual assumption of time-invariant long term co-volatility dynamics by introducing a higher level of flexibility yet preserving parsimony and computational tractability. Indeed, when the dimension of the application is moderately small, estimation can be performed by maximizing a Wishart quasi likelihood function in one step which yields consistent and asymptotically efficient estimates of the parameters.

The main achievement of the present work is to provide compelling evidence that these models are useful in practical applications and outperform models not accounting for such time-varying long run dynamics. Our out-of-sample forecasting analysis, designed to guarantee robustness to the choice i) of the components of the model universe and ii) of the loss function used for assessing predictive accuracy, showed that accounting for time variation in the levels of volatilities and correlations clearly contributes to provide more accurate out-of-sample conditional covariance forecasts.

The research conducted in this chapter lends itself to extensions along two main directions. First, the set of proposed models featuring a MIDAS component is potentially applicable to larger dimensional frameworks. However, as the cross-sectional size increases, substantial complications arise due to the proliferation of model parameters, with the main source of burden represented by the intercept matrix in the MIDAS equation, whose parameters increase at rate $O(n^2)$. In order to alleviate the computational burden and to make estimation feasible even in large dimension, it may be necessary to abstract from the standard 1-step QML approach and to investigate alternative estimation approaches based on the implementation of shrinkage or targeting-type strategies.

Second, it should be mentioned that component models featuring short and long-run dynamics are conceivably expected to be beneficial for longer-term forecasts. In particular, it is reasonable to expect that increasing the forecast horizon is likely to increase the gap between constant and time-varying long run component models in favor of the latter. Unfortunately, the generation of long term forecasts is complicated by

the complexity of the proposed specifications which rules out the possibility of deriving exact analytical predictors. Alternatively, the use of simulation or bootstrap based procedures could be considered as a rigorous and flexible approach to the generation of point long term forecasts from the parametric specifications considered in the chapter. Under this respect it is worth noting that the use of bootstrap procedures would also have the advantage of easily allowing the computation of interval forecasts.

In the next chapter, we investigate the aforementioned research topics for the special case of the MMReDCC model. Extensions to other parametric specifications are still under investigation.

Appendix

3.A Tables

Table 3.7: Descriptive statistics of realized variances

The table reports descriptive statistics of the realized variances of the stocks used in the empirical application. The three panels report the statistics for the in-sample period, the out-of-sample period and the full sample period, respectively.

Symbol	Mean	Max.	Min.	Std.dev.	Skewness	Kurtosis
ESTIMATION SAMPLE: FEBRUARY 1, 2001 TO DECEMBER 31, 2006 (1528 OBSERVATIONS)						
BAC	1.31	19.35	0.07	1.88	4.66	32.54
JPM	2.88	168.72	0.11	6.45	14.86	329.89
IBM	1.60	40.69	0.08	2.27	6.23	73.62
MSFT	2.19	29.65	0.10	2.86	3.98	26.93
XOM	1.59	40.89	0.13	1.98	8.29	126.45
AA	2.85	30.57	0.29	2.71	3.71	24.24
AXP	2.33	62.94	0.08	4.01	6.23	65.13
DD	1.83	28.93	0.16	2.07	4.98	46.68
GE	1.98	55.74	0.10	3.06	6.49	80.14
KO	1.29	22.21	0.04	1.60	5.57	52.96
FORECASTING SAMPLE: JANUARY 1, 2007 TO DECEMBER 31, 2009 (712 OBSERVATIONS)						
BAC	14.36	277.31	0.10	27.69	4.07	25.22
JPM	9.74	176.48	0.16	16.32	4.74	34.81
IBM	2.65	57.54	0.10	4.88	5.86	49.89
MSFT	3.02	43.11	0.08	4.32	4.65	31.49
XOM	3.12	115.38	0.17	6.66	9.28	129.77
AA	9.44	160.24	0.49	14.36	4.82	36.89
AXP	8.91	201.88	0.24	14.14	6.09	64.94
DD	4.02	63.87	0.17	5.60	4.75	35.36
GE	5.81	114.26	0.18	11.37	4.66	31.05
KO	1.69	56.51	0.12	3.23	8.79	125.68
FULL SAMPLE: FEBRUARY 1, 2001 TO DECEMBER 31, 2009 (2240 OBSERVATIONS)						
BAC	5.46	277.31	0.07	16.82	7.17	72.51
JPM	5.06	176.48	0.11	11.10	7.53	84.62
IBM	1.93	57.54	0.08	3.36	7.32	84.96
MSFT	2.45	43.11	0.08	3.41	4.73	36.29
XOM	2.07	115.38	0.13	4.16	13.29	287.87
AA	4.95	160.24	0.29	8.94	7.63	92.05
AXP	4.42	201.88	0.08	9.16	8.53	133.60
DD	2.53	63.87	0.16	3.73	6.44	68.47
GE	3.20	114.26	0.10	7.12	7.23	75.42
KO	1.41	56.51	0.04	2.26	9.71	180.08

Table 3.8: QML parameter estimates

The table reports QML parameter estimates and corresponding robust standard errors in brackets. For the univariate MIDAS and GARCH estimates we report average values among series and corresponding mean asymptotic square errors (MASE). Asymptotic standard errors are calculated using the “sandwich” estimator of Bollerslev et al.(1988). The results are based on the full-sample dataset, i.e. from February 1, 2001 to December 31, 2009.

Models	Kernel	MIDAS			GARCH		Correlation		
	h	θ	ω_{2s}	$\bar{m}$	γ	δ	α	β	ω_{2r}
<i>Additive models</i>									
1 - AMReDCC		0.68 (0.080)	7.51 (2.506)	0.34 (0.09)	0.33 (0.03)	0.50 (0.06)	0.12 (0.00)	0.56 (0.15)	12.05 (2.84)
2 - AMReDECO		0.64 (0.30)	7.19 (3.20)	0.40 (0.46)	0.35 (0.05)	0.50 (0.09)	0.24 (0.04)	0.63 (0.18)	10.73 (14.47)
3 - AMArDCC		0.02 (0.00)	8.34 (8.55)	0.53 (0.35)	0.33 (0.04)	0.56 (0.08)	0.27 (0.06)	0.70 (0.08)	12.10 (3.83)
4 - AMArDECO		0.02 (0.01)	4.91 (5.11)	0.67 (0.50)	0.37 (0.04)	0.55 (0.05)	0.47 (0.09)	0.51 (0.10)	20.6 (9.93)
5 - AMReDCC-P		0.61 (0.11)	6.68 (2.50)	0.42 (0.13)	0.33 (0.03)	0.52 (0.06)	0.09 (0.00)	0.89 (0.00)	2.91 (1.44)
6 - AMReDECO-P		0.60 (0.27)	7.84 (16.40)	0.52 (0.28)	0.35 (0.053)	0.44 (0.12)	0.25 (0.05)	0.70 (0.07)	2.46 (11.2)
7 - AMArDCC-P		0.03 (0.01)	7.84 (23.78)	0.72 (1.89)	0.35 (0.12)	0.58 (0.09)	0.25 (0.11)	0.73 (0.13)	11.75 (24.29)
8 - AMArDECO-P		0.02 (0.00)	5.42 (7.30)	0.74 (0.19)	0.37 (0.03)	0.55 (0.06)	0.40 (0.19)	0.60 (0.21)	27.30 (9.29)
<i>Multiplicative models</i>									
9 - NPreDCC	0.078				0.37 (0.03)	0.60 (0.04)	0.08 (0.00)	0.88 (0.00)	
10 - NPreDECO	0.078				0.36 (0.03)	0.60 (0.04)	0.20 (0.02)	0.74 (0.03)	
11 - NPreBEKK	0.078						0.23 (0.01)	0.74 (0.01)	
12 - MMReDCC		0.97 (0.42)			0.23 (0.04)	0.50 (0.17)	0.07 (0.01)	0.82 (0.14)	7.21 (2.80)
13 - MMReDECO		0.97 (0.07)			0.39 (0.09)	0.47 (0.13)	0.19 (0.08)	0.72 (0.13)	13.87 (1.72)
14 - MMReBEKK		0.87 (0.04)					0.20 (0.02)	0.55 (0.07)	19.89 (7.63)
15 - MMArDCC		0.04 (0.00)			0.38 (0.04)	0.52 (0.05)	0.06 (0.00)	0.87 (0.01)	5.79 (0.62)
16 - MMArDECO		0.04 (0.00)			0.37 (0.05)	0.48 (0.13)	0.20 (0.08)	0.72 (0.19)	12.17 (5.48)
17 - MMArBEKK		0.04 (0.00)					0.24 (0.01)	0.62 (0.02)	8.62 (1.01)
<i>Benchmark models</i>									
18 - CRDCC					0.22 (0.02)	0.70 (0.02)	0.27 (0.00)	0.65 (0.01)	
19 - CRDECO					0.14 (0.05)	0.79 (0.10)	0.32 (0.09)	0.70 (0.19)	
20 - CRBEKK							0.28 (0.03)	0.74 (0.03)	
21 - ReDCC					0.32 (0.12)	0.56 (0.21)	0.19 (0.08)	0.70 (0.09)	

Table 3.9: **Robust loss functions definition**

The set of robust loss functions used in the empirical application involves Euclidean and Frobenius distances, Mean Square Forecast Error (MSFE), QLIKE, Stein and von Neumann divergence (VND). In the table, C_t is the proxy used for the true conditional covariance matrix while S_t denotes each model predicted covariance matrix for the day t (the subscript i on S_t has been removed for simplicity). The first three loss functions are symmetric quadratic loss functions based on the forecast error. The Euclidean distance only accounts for the unique elements of the covariance matrix while the Frobenius distance double counts the loss associated to the conditional covariances. For the description of the other losses also used in Chapter 2, see the note in Table 2.11.

Loss Function	Formula	Type
Euclidean	$\text{vech}(C_t - S_t)' \text{vech}(C_t - S_t)$	Symmetric
Frobenius	$\text{tr}[(C_t - S_t)'(C_t - S_t)]$	Symmetric
MSFE	$\frac{1}{T} \text{vec}(C_t - S_t)' \text{vec}(C_t - S_t)$	Symmetric
QLIKE	$\log S_t + \text{tr}(S_t^{-1} C_t)$	Symmetric
Stein	$\text{tr}(S_t^{-1} C_t) - \log S_t^{-1} C_t - n$	Asymmetric
VND	$\text{tr}(C_t \log C_t - C_t \log S_t - C_t + S_t)$	Asymmetric

Table 3.10: **90% MCS p -values**

Entries are the p -values resulting from the 90% MCS. Models are included if the p -value is larger than 0.10.

Models	Euclidean	Frobenius	MSFE	QLIKE	Stein	VND
<i>Additive models</i>						
1 - AMReDCC	0.0973	0.1799	0.1637	0.0067	0.0053	0.0987
2 - AMReDECO	0.0973	0.1799	0.1637	0.0000	0.0000	0.0000
3 - AMArDCC	0.0973	0.1799	0.1637	0.0067	0.0001	0.0929
4 - AMArDECO	0.0973	0.1799	0.1637	0.0000	0.0000	0.0000
5 - AMReDCC-P	0.0996	0.1927	0.1936	0.0067	0.0053	0.203
6 - AMReDECO-	0.0608	0.1257	0.1214	0.0000	0.0000	0.0000
7 - AMArDCC-P	0.0973	0.1799	0.1637	0.0003	0.0000	0.0929
8 - AMArDECO-P	0.0608	0.0943	0.0854	0.0000	0.0000	0.0000
<i>Multiplicative models</i>						
9 - NPrDCC	1.000	1.000	1.000	0.0000	0.0000	1.000
10 - NPrDECO	0.0608	0.0877	0.0854	0.0000	0.0000	0.0013
11 - NPrBEKK	0.0608	0.0943	0.0854	0.0000	0.0000	0.0013
12 - MMrDCC	0.0608	0.0943	0.0854	1.000	1.000	0.203
13 - MMrDECO	0.0608	0.0943	0.0854	0.2993	0.3005	0.0987
14 - MMrBEKK	0.1885	0.1927	0.1936	0.0000	0.0000	0.203
15 - MMarDCC	0.0608	0.0943	0.0854	0.0067	0.0053	0.1612
16 - MMarDECO	0.0608	0.0877	0.0854	0.0003	0.0001	0.0168
17 - MMarBEKK	0.0996	0.1927	0.1936	0.0000	0.0000	0.0388
<i>Benchmark models</i>						
18 - CRDCC	0.0608	0.1257	0.1214	0.0000	0.0000	0.0168
19 - CRDECO	0.0608	0.0943	0.0854	0.0000	0.0000	0.0000
20 - CRBEKK	0.1885	0.206	0.1993	0.0000	0.0000	0.0987
21 - RDCC	0.0533	0.076	0.0697	0.0000	0.0000	0.0000

Table 3.11: **75% MCS p -values**

Entries are the p -values resulting from the 75% MCS. Models are included if the p -value is larger than 0.25.

Models	Euclidean	Frobenius	MSFE	QLIKE	Stein	VND
<i>Additive models</i>						
1 - AMReDCC	0.103	0.192	0.167	0.006	0.006	0.104
2 - AMReDECO	0.103	0.192	0.167	0.000	0.000	0.000
3 - AMAReDCC	0.103	0.192	0.167	0.006	0.006	0.094
4 - AMAReDECO	0.103	0.192	0.167	0.000	0.000	0.000
5 - AMReDCC-P	0.103	0.203	0.188	0.006	0.006	0.273
6 - AMReDECO-P	0.066	0.128	0.130	0.000	0.000	0.000
7 - AMAReDCC-P	0.103	0.192	0.167	0.001	0.001	0.094
8 - AMAReDECO-P	0.066	0.103	0.087	0.000	0.000	0.000
<i>Multiplicative models</i>						
9 - NP-ReDCC	1.000	1.000	1.000	0.000	0.000	1.000
10 - NP-ReDECO	0.066	0.086	0.087	0.000	0.000	0.001
11 - NP-ReBEKK	0.066	0.103	0.087	0.000	0.000	0.001
12 - MMReDCC	0.066	0.103	0.087	1.000	1.000	0.273
13 - MMReDECO	0.066	0.103	0.087	0.309	0.289	0.104
14 - MMReBEKK	0.185	0.203	0.188	0.000	0.000	0.273
15 - MMAReDCC	0.066	0.103	0.087	0.006	0.006	0.181
16 - MMAReDECO	0.066	0.086	0.087	0.001	0.001	0.018
17 - MMAReBEKK	0.103	0.203	0.188	0.000	0.000	0.043
<i>Benchmark models</i>						
18 - CReDCC	0.066	0.128	0.130	0.000	0.000	0.018
19 - CReDECO	0.066	0.103	0.087	0.000	0.000	0.000
20 - CReBEKK	0.185	0.203	0.210	0.000	0.000	0.104
21 - RDCC	0.046	0.073	0.069	0.000	0.000	0.000

Chapter 4

Feasible inference procedures for forecasting high-dimensional realized covariance matrices¹

Abstract The Multiplicative MIDAS Realized DCC (MMReDCC) model introduced in Chapter 3 showed remarkably good results compared to the selected competitors. However, due to the high number of parameters involved, its estimation becomes unfeasible for large cross-sectional dimensions. This chapter aims at improving the applicability of the model in two directions. First, by mitigating the over-parameterization problem to make estimation feasible in large dimensions and second, by allowing for multi-step ahead predictions. The proposed inference procedures are illustrated in the content of the chapter and their properties are analyzed through detailed Monte Carlo simulations. In an empirical application on a dataset of forty-six equities, the MMReDCC model is found to statistically outperform the selected benchmarks in terms of in-sample fit as well as in terms of out-of-sample covariance predictions. The latter are mostly significant in periods of high market volatility.

4.1 Introduction

Building models for predicting the volatility of high dimensional portfolios is important in risk management and asset allocation. Previous developments on time-varying covariances in large dimensions include the constant conditional correlation (CCC) model of [Bollerslev \(1990\)](#), where the volatilities of each asset are allowed to vary through time but the correlations are time invariant, the RiskMetrics model by [Morgan \(1994\)](#) and the DECO model by [Engle & Kelly \(2012\)](#), where pairwise correlations are assumed equal at a point in time, whilst being time varying. Recently, [Andersen, Bollerslev,](#)

¹This chapter is inspired by the paper [Bauwens, Braione & Storti \(2016a\)](#) forthcoming in *Econometrics and Statistics*.

Diebold & Ebens (2001), Barndorff-Nielsen & Shephard (2001) and Barndorff-Nielsen et al. (2011), among others, opened up a new channel for increasing the precision of covariance matrix estimates and forecasts by exploiting the information of high frequency asset returns. This development has motivated several researchers to investigate models directly fitted to series of realized covariance matrices (see Gouriéroux et al. (2009), Jin & Maheu (2013) and Chiriac & Voev (2011), to name a few).

Despite the superiority of these models, illustrated for example by Hautsch et al. (2015), there still remain technical and practical challenges one needs to deal with when constructing covariance matrix forecasts for high-dimensional systems. Primarily, the well-known “curse of dimensionality” problem, implying that the number of parameters grows as a power function of the cross-sectional model dimension. In order to save parameters, a simple solution is represented by the so called covariance (or correlation) targeting approach of Engle (2009), which consists in pre-estimating the constant intercept matrix in the model specification by linking it to the unconditional covariance matrix of returns. This method can be applied under the stationarity assumption of the model and is one of the most widely employed techniques to simplifying parameter estimation and reducing the computational burden when the numerical maximization of the likelihood function becomes difficult. However, in order to implement it, the user needs to have knowledge of the analytical relationship between the mean of the targeted process and the other model parameters. In many cases, due to the complex dependence structure encountered in the dynamic behavior of co-volatility matrices, this is far from being obvious (see Aielli (2013)). Moreover, even in cases in which targeting is feasible, severe homogeneity constraints still have to be imposed on the model parameters in order to make the model tractable for inference purposes (see e.g. Bauwens, Grigoryeva & Ortega (2016)). For their ability to separately analyze short and long run movements, component models have been effectively used in the Multivariate GARCH literature as a parsimonious way of describing this rich dynamic structure (see e.g. Colacito et al. (2011), Hafner & Linton (2010)). More recently, their applicability has been extended to dynamic models for realized covariance matrices mainly by Golosnoy et al. (2012).

In Chapter 3 we investigated a wide class of multivariate models that simultaneously account for short and long-term dynamics in the conditional co-volatilities and correlations of asset returns, in line with the empirical evidence suggesting that their level is changing over time as a function of the economic conditions (see, among others, Engle et al. (2013)). Herein we focus on the Multiplicative MIDAS Realized DCC (MMReDCC) model, whose main ingredients are a multiplicative component structure, a Mixed Data Sampling (MIDAS)-type filter for modeling the secular dynamics and a short term parameterization directly inspired by the Dynamic Conditional Correlation (DCC) model of Engle (2002a).² The following considerations support the choice for this model. First, the multiplicative component structure entails specifying a unique long term factor for the whole covariance matrix, thus preserving tractability in the

²We refer to Engle (2002a), Ghysels et al. (2007) and references therein for detailed discussions of the DCC model and standard MIDAS regressions.

estimation process. Second, the MIDAS-type filter used in the specification of the secular component, compared to non-parametric smoothers, offers a much more convenient framework for the generation of multi-step ahead forecasts and potentially allows for the inclusion of external explanatory variables (such as low-frequency macroeconomic indicators). Also, if compared to other parametric alternatives, it represents an appealing compromise between parsimony and versatility. Finally, the short run DCC structure improves in flexibility over more basic linear recursions (e.g. the BEKK model by Engle & Kroner (1995)).

On an empirical ground, our choice is further supported by the results of the extensive out-of-sample forecasting comparison performed in Chapter 3. Although not identifying a unique winner, these findings show that the MMReDCC model gives remarkably good performances in important financial applications such as Value-at-Risk forecasting and portfolio allocation. Notwithstanding, these results are limited to a relatively low dimensional setting (ten assets) and to a short term forecasting horizon (one day).

This chapter extends the work done so far along these directions: estimation for high-dimensional systems and multi-step forecasting. We contribute to the first line of research by developing a computationally feasible procedure for the estimation of vast dimensional MMReDCC models. In this respect, it is important to remark that, although the introduction of a dynamic secular component in the structure of the model adds a major element of flexibility and enables to obtain more accurate forecasts than standard models reverting to constant mean levels, it also dramatically increases the number of parameters to be estimated. Specifically, the long term component incorporates a scale intercept matrix with number of parameters equal to $n(n+1)/2$, where n denotes the number of assets. In a vast dimensional framework, this quickly translates into the impossibility of estimating the model since the intercept matrix cannot be directly targeted.

Therefore, we propose to overcome this estimation issue by proposing an iterative procedure inspired by the covariance targeting of Engle (2009). More precisely, based on a Method of Moments estimator, we profile out the parameters of the intercept matrix and iteratively maximize the likelihood in terms of the other parameters of interest. We refer to this as the Iterative Moment-Based Profiling (IMP) estimator, as opposed to the one-step Quasi Maximum Likelihood (1QML) estimator which directly maximizes the likelihood with respect to the full parameter vector.

It is worth noticing that the proposed estimation procedure can be considered a *switching algorithm* in the sense discussed by Boswijk (1995) and Cubadda et al. (2015) since the maximization of the overall likelihood is obtained by switching between optimizations over different blocks of parameters. This idea has a long standing tradition in the econometric analysis of time series. A simple, well known example of switching algorithm is given by the Cochrane-Orcutt iterative estimation procedure. Compared to conventional switching algorithms, the procedure that is here implemented incorporates an additional targeting step. In particular, it reduces the dimension of the optimization problem to be solved by concentrating out some of the parameters, the elements of the intercept matrix, by means of an iteratively re-computed moment-based

estimator. A comprehensive simulation study is performed to assess the finite-sample properties of the proposed estimator which is found to deliver unbiased estimates and to quickly converge, as no more than three iterations are required in general.

The second contribution of the paper is the development of a resampling based approach for the generation of multi-step ahead forecasts of the realized covariance matrices, which is motivated by the impossibility to derive a closed-form expression for the h -step predictor from the MMReDCC model. The bootstrap has been a standard tool for generating multi-step forecasts from non-linear and non-Gaussian time series models for more than two decades (see e.g. [Clements & Smith \(1997\)](#)). Its use has been later extended to univariate volatility modeling (see e.g. [Pascual et al. \(2006\)](#); [Shephard & Sheppard \(2010\)](#)) and more recently to the multivariate setting by [Fresoli et al. \(2015\)](#) and [Fresoli & Ruiz \(2016\)](#).

In the same spirit of [Fresoli & Ruiz \(2016\)](#), we propose a resampling algorithm making use of residual bootstrap with the appealing feature of being distribution-free, namely not relying on any particular assumption about the distribution of the realized covariance matrices. Differently from them, who generate multi-step forecasts from DCC models, our algorithm is adapted to the dynamic modeling of realized covariance matrices.

Two different applications are performed to illustrate the usefulness of the model on real data. In the first one, we focus on a low dimensional setting (ten assets), in which both the IMP and 1QML estimation procedures are feasible, and compare the estimates obtained by means of both algorithms. We find that the IMP-based estimates are sufficiently close to the 1QML ones, so that using the IMP method in large dimensions is a sensible approach. We further consider the case in which the IMP estimated parameters are used as starting values for the one-step QMLE: despite an increase in the maximized likelihood value, the improvement can be considered rather marginal, thus suggesting that the implementation of the IMP algorithm alone may be sufficient in practical applications.

In the second application, the MMReDCC model estimated for forty-six assets by the IMP method is used to generate forecasts of the realized covariance matrix up to twenty days ahead and compared to existing benchmarks not accounting for short and long term co-volatility dynamics. It clearly emerges that over periods of relatively lower market volatility, where short and long run components are hardly distinguishable, simpler model specifications tend to be preferred especially at the shortest horizons. The opposite holds during more volatile periods, such as the one of the 2007-2008 financial crisis, where accounting for time-varying long term dynamics in the conditional covariance process appears crucial to generating superior forecasts. The latter are particularly sizeable at the longest horizons.

The remainder of the chapter is organized as follows. Section [4.2](#) briefly recalls the structure of the MMReDCC model and explains the curse of dimensionality issue. Section [4.3](#) introduces the IMP algorithm and Section [4.4](#) presents the results of a Monte Carlo experiment aimed at assessing the finite sample statistical properties of

the proposed estimation algorithm. The bootstrap procedure for computing multi-step ahead forecasts is explained in Section 4.5, along with a simulation study to assess its final sample behavior. Section 4.6 contains the empirical results for the in-sample estimation comparison and the out-of-sample forecasting exercise. Section 4.7 concludes with some final remarks.

4.2 Curse of dimensionality issue

Consider again the general framework illustrated in Section 3.2, where it is assumed that, for all t , it holds

$$C_t | \mathfrak{S}_{t-1} \sim W_n(\nu, S_t/\nu), \quad (4.1)$$

so that by the properties of the Wishart distribution, S_t represents the conditional covariance matrix of returns.

In the MMReDCC model, S_t is designed to take into account the long run movements in the levels around which realized (co)variances (and by extension, correlations) fluctuate from day to day. To this extent, the model features a multiplicative decomposition of the conditional covariance matrix S_t into a smoothly varying or *secular* component $M_t = L_t L_t'$ and a short-lived component S_t^* , such that S_t can be rewritten as $S_t = L_t S_t^* L_t'$, where the matrix square root L_t can be obtained by a Cholesky factorization of M_t . These components can then be modeled separately.

First, as already discussed in Section 3.2.2, the secular component is specified parametrically and extracted by means of the following MIDAS filter,

$$M_t = \Lambda + \theta \sum_{k=1}^K \phi_k(\omega) C_{t-k}. \quad (4.2)$$

where the first right-hand-side term Λ is a $n \times n$ symmetric and semi-positive definite matrix of constant parameters, θ is a positive scalar and $\phi_k(\cdot)$ is the scalar Beta weight function.

Second, the dynamics of the short term component S_t^* is specified according to a scalar DCC parametrization that enables a separate treatment of conditional volatilities and correlations, in the following way:

$$S_{ii,t}^* = (1 - \gamma_i - \delta_i) + \gamma_i C_{ii,t-1}^* + \delta_i S_{ii,t-1}^*, \quad \forall i = 1, \dots, n \quad (4.3)$$

$$R_t^* = (1 - \alpha - \beta) I_n + \alpha P_{t-1}^* + \beta R_{t-1}^*, \quad (4.4)$$

where $\gamma_i > 0$, $\delta_i \geq 0$, $\gamma_i + \delta_i < 1$, $\alpha > 0$, $\beta \geq 0$, $\alpha + \beta < 1$, $C_t^* = L_t^{-1} C_t (L_t')^{-1}$ and $P_t^* = (\text{diag}\{C_t^*\})^{-1/2} C_t^* (\text{diag}\{C_t^*\})^{-1/2}$. The matrix C_t^* is the realized covariance matrix purged of its long term component and the matrix P_t^* is the corresponding short term realized correlation matrix. Let $\boldsymbol{\gamma} = \{\gamma_1, \dots, \gamma_n\}$, $\boldsymbol{\delta} = \{\delta_1, \dots, \delta_n\}$ for further use.

Table 4.1: Number of parameters of MMReDCC models

Note: Entries report the number of parameters as a function of the dimension n ; n_Λ denotes the number of unique parameters contained in the Λ matrix, ψ denotes the full vector of model parameters and $\tilde{\psi}$ the vector of parameters excluding n_Λ .

	$n = 5$	$n = 10$	$n = 20$	$n = 50$	$n = 100$
n_Λ	15	55	210	1275	5050
ψ	29	79	254	1379	5254
$\tilde{\psi}$	14	24	44	104	204

The parameters can be estimated by maximizing the following Wishart (quasi) log-likelihood function in one step:

$$\ell_T(\psi) = -\frac{1}{2} \sum_{t=1}^T \left\{ \log |S_t(\psi)| + \text{tr}[S_t(\psi)^{-1} C_t] \right\}. \quad (4.5)$$

The finite-dimensional parameter vector³ $\psi = \{\text{vech}(\Lambda), \theta, \omega, \gamma, \delta, \alpha, \beta\}$, has length $\{n_\Lambda + 2n + 4\}$ where $n_\Lambda = n(n+1)/2 = O(n^2)$ denotes the number of unique parameters included in the intercept matrix Λ of Eq.(4.2). It is obvious that, as n increases, the curse of dimensionality problem quickly arises, leading to the number of parameters listed in the first two rows of Table 4.1. Observe that estimation becomes already cumbersome after $n = 20$ and almost impossible for $n \geq 50$.

On the other hand, the last row of Table 4.1 shows that an intuitive way to keep the model tractable is to avoid estimating the parameters of the matrix Λ . This would be sufficient to reduce the order to $2n + 4 = O(n)$, thus making the model estimable also for large n .

In the following section we put forward a feasible estimation procedure that aims at overcoming the direct estimation of the long term component intercept matrix, thus crucially mitigating the computational complexity of the model.

4.3 An Iterative Moment based Profiling (IMP) algorithm

In this section we discuss an iterative procedure for fitting the MMReDCC model to large dimensional datasets. The basic idea underlying the proposed algorithm is to eliminate from the likelihood maximization the parameters of the intercept matrix Λ using a technique that builds upon the covariance targeting discussed in Pedersen & Rahbek (2014) for BEKK and Engle et al. (2008) for DCC models. First of all, notice that from Eq.(4.2) and the following relation

$$\Lambda = \mathbb{E}(M_t) - \theta \sum_{k=1}^K \phi_k(\omega) \mathbb{E}(C_{t-k}),$$

³Note that ψ does not include the degrees of freedom parameter ν , as the first order conditions for the estimation of the parameter vector ψ do not depend on ν by linearity in ν (see Section 3.3).

a moment based estimator of the Λ intercept matrix is

$$\hat{\Lambda} = \frac{1}{T} \sum_{t=1}^T \left[M_t - \theta \sum_{k=1}^K \phi_k(\omega) C_{t-k} \right]. \quad (4.6)$$

Obviously, given the latent nature of M_t , the estimator in Eq.(4.6) cannot be computed in practice and hence the covariance targeting approach cannot be applied in the usual way. It is worth noting that, if L_t and S_t^* were assumed to be independent, given $\mathbb{E}(S_t^*) = I_n$, it would hold that $\mathbb{E}(C_t) = \mathbb{E}(M_t)$, implying that an asymptotically equivalent version of Eq.(4.6) could be explicitly computed replacing M_t by C_t . However, this is not the approach we pursue, since the assumption of independence of the short and long term sources is difficult to justify and would result in a rather counterintuitive and arbitrary constraint. Hence, we adopt a different method.

It can be seen from Eq.(4.6) that no estimate of Λ makes sense regardless of the value of (θ, ω) , so that by making this dependence explicit, it is possible to obtain an estimate of Λ as a function of (θ, ω) , i.e. $\hat{\Lambda}(\theta, \omega)$. In this way, a different estimate of Λ is required for each different value of the other two parameters. Therefore, by substituting $\hat{\Lambda}(\theta, \omega)$ for Λ in the Wishart QML function stated in Eq.(4.5), the following moment based QML approximation is obtained:

$$\tilde{\ell}_T(\tilde{\psi}) = -\frac{1}{2} \sum_{t=1}^T \left\{ \log |\tilde{L}_t(\theta, \omega) S_t^*(\tilde{\psi}) \tilde{L}_t'(\theta, \omega)| + \text{tr} \left\{ [\tilde{L}_t(\theta, \omega) S_t^*(\tilde{\psi}) \tilde{L}_t'(\theta, \omega)]^{-1} C_t \right\} \right\} \quad (4.7)$$

with $\tilde{\psi} = (\omega, \theta, \psi_{S^*})'$, $\psi_{S^*}' = (\gamma, \delta, \alpha, \beta)$ and

$$\tilde{M}_t(\theta, \omega) = \tilde{L}_t(\theta, \omega) \tilde{L}_t'(\theta, \omega) = \hat{\Lambda}(\theta, \omega) + \theta \sum_{k=1}^K \phi_k(\omega) C_{t-k}. \quad (4.8)$$

The method we propose consists in estimating the parameters in $\tilde{\psi}$ by a block-wise maximization of the moment-based QML function given in Eq.(4.7). First, conditional on some reasonable initial guess of (θ, ω) , $\tilde{\ell}_T(\tilde{\psi})$ is maximized with respect to the short term parameters ψ_{S^*} and then, conditional on $\hat{\psi}_{S^*}$, the same function is maximized with respect to (θ, ω) . The procedure is iterated for $j = 0, \dots, J$ until some pre-specified convergence criterion on the likelihood is met.

To initialize the algorithm at $j = 0$, one can reasonably use as starting values the parameter estimates obtained by fitting the model to low dimensional subsets of data; also, an initial guess for the long term component $M_{t,0}$ could be either provided in a naive way, i.e. using the series of observed realized covariance matrices directly, or in a more sophisticated manner, by fitting to the data a nonparametric kernel smoother with an optimized bandwidth parameter. Note that in order to guarantee the positive definiteness of $\tilde{M}_t(\theta, \omega)$ in Eq.(4.8), it suffices to initialize $M_{t,0}$ from a PDS matrix and to impose $\theta > 0$. Given that the observed series of C_t , for every t , is PDS by definition, $\hat{\Lambda}(\theta, \omega)$ is assured to be at least semi-positive definite at each iteration $j > 0$.

Once $\Lambda_j(\theta_j, \omega_j)$ has been computed at the initial iteration $j = 0$, for every $j > 0$ the steps conducted in the algorithm are as follows:

Step 1 Plug $\Lambda_{j-1}(\theta_{j-1}, \omega_{j-1})$ into Eq.(4.2), then get $\tilde{M}_{t,j}$ and $\tilde{L}_{t,j} = \text{chol}(\tilde{M}_{t,j})$ for all t ;

Step 2 For each asset $i = 1, \dots, n$, obtain the short term GARCH(1,1) parameters of Eq.(4.4) as follows

$$\{\gamma_{i,j}, \delta_{i,j}\} = \underset{\{\gamma_i, \delta_i\}}{\operatorname{argmax}} \tilde{\ell}_T(\theta_{j-1}, \omega_{j-1}, \alpha_{j-1}, \beta_{j-1});$$

Step 3 Conditional on the estimated vectors $\boldsymbol{\gamma}_j = (\gamma_{1,j}, \dots, \gamma_{n,j})'$ and $\boldsymbol{\delta}_j = (\delta_{1,j}, \dots, \delta_{n,j})'$, maximize the same log-likelihood function with respect to the short term DCC correlation parameters:

$$\{\alpha_j, \beta_j\} = \underset{\{\alpha, \beta\}}{\operatorname{argmax}} \tilde{\ell}_T(\theta_{j-1}, \omega_{j-1}, \boldsymbol{\gamma}_j, \boldsymbol{\delta}_j);$$

Step 4 Finally, conditional on the vector of short term parameter estimates $\boldsymbol{\phi}_{s*} = \{\boldsymbol{\gamma}_j, \boldsymbol{\delta}_j, \alpha_j, \beta_j\}$, maximize $\tilde{\ell}_T$ with respect to $\{\theta_j, \omega_j\}$; these estimates are used to compute an updated version of $\Lambda_j(\theta_j, \omega_j)$;

Step 5 Check for convergence, *i.e.* if

$$\left| \frac{\tilde{\ell}_T(\tilde{\boldsymbol{\psi}}_j) - \tilde{\ell}_T(\tilde{\boldsymbol{\psi}}_{j-1})}{\tilde{\ell}_T(\tilde{\boldsymbol{\psi}}_{j-1})} \right| < \epsilon, \quad \epsilon = 0.000001;$$

convergence is achieved and the algorithm stops; otherwise update all parameter estimates and go back to Step 1.

It is worth to stress that although $\tilde{\ell}_T(\tilde{\boldsymbol{\psi}})$ looks like a profile likelihood, it is not since $\hat{\Lambda}(\theta, \omega)$ is not a QML estimator but a feasible moment estimator. This motivates our choice to refer to Steps 1 – 5 as the Iterative Moment based Profiling algorithm, or IMP for short. This implies that $\tilde{\boldsymbol{\psi}}$ is typically less efficient than the standard 1QML estimator that maximizes Eq.(4.5) in one step. We come back to this issue in Section 4.6.1.

4.4 Simulation study

A Monte Carlo study is conducted to analyze the finite sample properties of the IMP estimator. We assume the MMReDCC to be the DGP and we generate 500 processes of length $T = 1000$ and 2000 for $n = 10, 20, 40$ and 50 , with true parameter values inspired by the estimates obtained in Chapter 3, as summarized in Table 4.2.

It is important to stress that, in order to initialize the algorithm, parameter values have to be carefully chosen. This is a standard requirement in every optimization

Table 4.2: **Simulation setting**

In Panel A, for every $i = 1, \dots, n$ it holds $\{\gamma_i + \delta_i\} < 1$. Entries of Panel B are scalar parameters chosen to initialize the algorithm in both sets of simulation exercises.

Panel A: Parameters					
θ	0.5				
K	264				
ω	15				
Λ	$\Lambda_{i,i} = 0.02, \Lambda_{i,j} = 0.002$ for $i \neq j$				
γ_i	$\sim U(\gamma_0 - 0.02, \gamma_0 + 0.02), \gamma_0 = 0.2$				
δ_i	$\sim U(2\delta_0 + \gamma_i - 1 + 0.01, 1 - \gamma_i - 0.01), \delta_0 = 0.7$				
α	0.2				
β	0.7				
ν	$2n$				
T	1000, 2000 (+ 1000 burn-in)				
Panel B: Initial values					
θ_0	ω_0	$\gamma_{i,0}$	$\delta_{i,0}$	α_0	β_0
0.8	10	0.05	0.90	0.05	0.90

Table 4.3: **Simulation exercise I: summary statistics**

The table reports summary statistics of the first set of simulations where $M_{t,0}$ is initialized using a nonparametric kernel estimator, see Section 4.3. To save on space, $\bar{\gamma}$ and $\bar{\delta}$ are reported as averaged values across series and replications. RB denotes the Relative Bias computed over 500 replications. True parameter values used to simulate the process at the top of the table.

T= 1000							T= 2000						
	$\bar{\gamma}$	$\bar{\delta}$	α	β	θ	ω		$\bar{\gamma}$	$\bar{\delta}$	α	β	θ	ω_2
	0.197	0.705	0.2	0.7	0.5	15		0.197	0.705	0.2	0.7	0.5	15
n=10							n=10						
RB	0.098	-0.036	0.020	0.003	-0.058	-0.120	RB	-0.039	-0.002	0.033	0.003	0.053	-0.095
IQR	0.048	0.093	0.006	0.010	0.044	1.641	IQR	0.032	0.068	0.006	0.008	0.037	1.819
Mean	0.202	0.699	0.204	0.702	0.475	14.820	Mean	0.2	0.713	0.207	0.702	0.526	13.58
Min	0.176	0.660	0.191	0.679	0.393	7.460	Min	0.153	0.669	0.183	0.523	0.072	1.949
Max	0.214	0.735	0.220	0.728	0.709	18.943	Max	0.22	0.803	0.37	0.817	1.000	17.74
n=20							n=20						
RB	0.049	-0.009	0.019	0.001	-0.056	-0.110	RB	0.036	0.011	0.024	0.002	-0.014	-0.080
IQR	0.046	0.083	0.003	0.006	0.019	0.713	IQR	0.031	0.079	0.002	0.004	0.015	0.622
Mean	0.202	0.701	0.204	0.701	0.472	14.782	Mean	0.209	0.708	0.205	0.702	0.496	13.801
Min	0.171	0.633	0.197	0.687	0.430	12.802	Min	0.197	0.678	0.200	0.694	0.001	2.440
Max	0.219	0.744	0.211	0.711	0.532	16.759	Max	0.221	0.739	0.220	0.748	0.598	15.270
n=40							n=40						
RB	0.028	0.023	0.015	0.002	-0.049	-0.080	RB	0.033	0.029	0.022	0.002	-0.014	-0.072
IQR	0.042	0.077	0.002	0.002	0.011	0.372	IQR	0.030	0.064	0.001	0.002	0.007	0.263
Mean	0.208	0.715	0.203	0.701	0.476	14.810	Mean	0.209	0.719	0.204	0.702	0.493	13.925
Min	0.190	0.656	0.060	0.695	0.446	3.735	Min	0.181	0.671	0.182	0.674	0.172	1.000
Max	0.217	0.762	0.222	0.799	0.705	16.500	Max	0.221	0.761	0.223	0.744	0.837	14.760
n=50							n=50						
RB	0.027	0.011	0.016	0.001	-0.045	0.012	RB	0.029	0.027	0.017	0.001	-0.011	-0.037
IQR	0.042	0.076	0.001	0.002	0.008	0.293	IQR	0.030	0.056	0.001	0.002	0.007	0.220
Mean	0.208	0.716	0.203	0.701	0.473	15.182	Mean	0.208	0.720	0.203	0.701	0.494	14.442
Min	0.162	0.644	0.200	0.697	0.455	13.150	Min	0.191	0.657	0.199	0.695	0.474	12.089
Max	0.220	0.814	0.207	0.705	0.525	15.830	Max	0.222	0.763	0.207	0.707	0.529	16.238

based procedure where the initial amount of information on the model parameters can be limited or even null. In our situation we are mainly concerned with the impact that different choices of $M_{t,0}$, more than the remaining set of parameters, may have on the convergence of the IMP algorithm. We evaluate this by performing a robustness check based on the two possible initializations of $M_{t,0}$ mentioned in Section 4.3.

Table 4.4: Simulation exercise II: summary statistics

The table reports summary statistics of the second set of simulations where $M_{t,0}$ is initialized from the series of realized covariance matrices, see Section 4.3. To save on space, $\bar{\gamma}$ and $\bar{\delta}$ are reported as averaged values across series and replications. RB denotes the Relative Bias computed over 500 replications. True parameter values used to simulate the process at the top of the table.

T= 1000							T= 2000						
	$\bar{\gamma}$	$\bar{\delta}$	α	β	θ	ω		$\bar{\gamma}$	$\bar{\delta}$	α	β	θ	ω_2
	0.197	0.705	0.2	0.7	0.5	15		0.197	0.705	0.2	0.7	0.5	15
n=10							n=10						
RB	-0.077	0.013	0.019	0.002	-0.032	-0.007	RB	0.043	0.006	0.025	0.002	-0.012	-0.069
IQR	0.070	0.113	0.007	0.010	0.043	1.664	IQR	0.028	0.046	0.004	0.007	0.024	0.870
Mean	0.191	0.708	0.204	0.701	0.484	14.892	Mean	0.211	0.707	0.205	0.701	0.494	13.959
Min	0.107	0.588	0.191	0.678	0.393	6.678	Min	0.189	0.654	0.196	0.686	0.435	9.614
Max	0.231	0.833	0.218	0.723	0.927	19.237	Max	0.232	0.748	0.213	0.715	0.623	15.535
n=20							n=20						
RB	0.048	-0.002	0.018	0.002	-0.045	-0.008	RB	0.006	0.039	0.023	0.002	-0.011	-0.064
IQR	0.042	0.085	0.003	0.005	0.021	0.719	IQR	0.030	0.060	0.002	0.004	0.013	0.510
Mean	0.207	0.702	0.204	0.701	0.477	14.887	Mean	0.206	0.721	0.205	0.701	0.494	14.047
Min	0.196	0.649	0.197	0.689	0.429	6.751	Min	0.164	0.680	0.199	0.692	0.467	11.108
Max	0.220	0.740	0.214	0.713	0.887	16.573	Max	0.225	0.772	0.209	0.710	0.555	15.241
n=40							n=40						
RB	0.046	0.017	0.017	0.003	-0.045	0.005	RB	0.050	0.016	0.021	0.002	-0.012	-0.057
IQR	0.042	0.080	0.002	0.003	0.010	0.421	IQR	0.029	0.053	0.001	0.002	0.007	0.245
Mean	0.209	0.713	0.203	0.702	0.477	15.078	Mean	0.210	0.718	0.204	0.701	0.494	14.148
Min	0.197	0.682	0.193	0.696	0.116	6.895	Min	0.194	0.658	0.202	0.697	0.479	8.433
Max	0.219	0.756	0.216	0.807	0.955	49.985	Max	0.223	0.768	0.212	0.708	0.726	14.638
n=50							n=50						
RB	0.029	0.028	0.016	0.002	-0.053	0.015	RB	0.028	0.018	0.020	0.002	-0.017	-0.048
IQR	0.041	0.071	0.001	0.002	0.009	0.342	IQR	0.030	0.061	0.001	0.001	0.005	0.195
Mean	0.208	0.715	0.203	0.701	0.473	15.225	Mean	0.208	0.721	0.204	0.701	0.492	14.278
Min	0.190	0.632	0.200	0.698	0.457	14.190	Min	0.184	0.674	0.202	0.698	0.437	11.178
Max	0.220	0.766	0.206	0.705	0.494	15.844	Max	0.216	0.766	0.209	0.713	0.508	14.742

In the first set of repetitions $M_{t,0}$ is computed by fitting to the series of simulated realized covariance matrices a Nadaraya-Watson kernel estimator with a single bandwidth parameter for the whole covariance matrix. As already done in Chapter 3 (see also e.g. Bauwens et al. (2013)), the optimal bandwidth is selected by a least squares cross-validation criterion, where the six-month rolling covariance is used as the reference for the computation of least squares. In the second (equivalent) simulation study, $M_{t,0}$ is obtained by substituting in Eq.(4.6) the observed C_t for the latent matrix M_t at each t . In both cases, the initial scalar model parameters are set equal to the values listed in Panel B of Table 4.2. The estimation bias is evaluated by the relative bias (RB), computed as $\frac{1}{500} \sum_{i=1}^{500} \frac{\hat{\psi}_i - \psi}{\psi}$, along with the interquartile range (IQR), mean, minimum and maximum of the obtained parameter estimates. To save space, we report averaged bias results for the parameters of the MIDAS intercept matrix in a separate table.

Table 4.3 reports results from the first simulation exercise. The emerging picture looks encouraging. As expected, the relative biases decrease as n or T increases. The biases for the parameters of the short term volatility and correlation components are very small, being smaller than five per cent in most of the cases, with one exception recorded for $\bar{\gamma}$ at $T = 1000$ for $n = 10$. As for the scalar parameters in the MIDAS specification, the bias for θ is negative in seven out of eight cases (the exception occurs

for $n = 10$ at $T = 2000$) and ranging from the maximum of 5.8% (in absolute value) for $n = 10$ and $T = 1000$ to the lowest value of 1.1% for $n = 50$ and $T = 2000$. The bias on the ω parameter, also generally negative, tends to decrease with n but is usually of higher order (from 1.1 to 12% in absolute value). A similar behavior is observed for the IQR measure, which decreases across n and T but remains on higher values for the parameter ω . However, this does not represent a major concern as the Beta weight function is not very sensitive to small variations of this parameter and therefore we do not expect the likelihood function to be either.

Table 4.4 gives an idea of the robustness of the results to the other initialization of the long term component. Entries can be directly compared to those in Table 4.3. As hoped for, the initial choice has a minor impact on the overall accuracy of the estimator, as the parameter biases are in the same range of magnitude and the comments made earlier are still valid under this alternative scenario.

Figure 4.1 contains plots of the Monte Carlo standard deviations of the estimated θ, ω, α and β parameters against the cross-section size. In all cases, standard deviations tend to decline as the cross-section dimension grows, with a faster decline when $T = 2000$. The two approaches produce similar parameter standard deviations, with slightly bigger values recorded for θ and ω under the second simulation experiment in correspondence with the higher cross-section sizes.

Figure 4.1: IMP Monte Carlo standard deviations

The figure shows standard deviations of the IMP Monte Carlo estimated scalar parameters θ, ω, α and β against the cross-section dimension ranging from 10 to 50. Results from the first (Sim.I) and second (Sim.II) simulation study are jointly reported in Panel A for $T=1000$ and Panel B for $T=2000$.

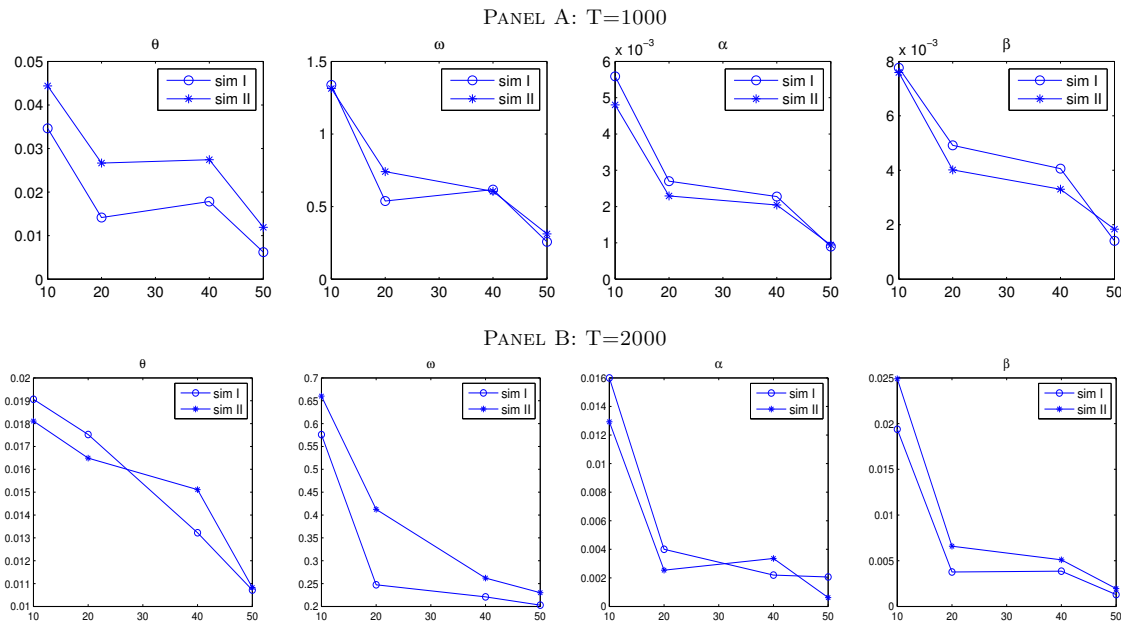


Table 4.5: Bias results for the scale MIDAS intercept matrix

$RB_{\{i,i\}}$ denotes averaged values over diagonal terms, while $RB_{\{i,j\}}$ denotes averages over off diagonal terms. Panel A reports summary statistics of the first simulation exercise where $M_{t,0}$ is initialized from a nonparametric smoother while Panel B reports results from the second simulation exercise where the series of observed realized covariance matrices are used.

Panel A: Simulation exercise I				Panel B: Simulation exercise II			
T=1000		T=2000		T=1000		T=2000	
n=10		n=10		n=10		n=10	
$RB_{\{i,i\}}$	0.080	$RB_{\{i,i\}}$	0.033	$RB_{\{i,i\}}$	0.075	$RB_{\{i,i\}}$	0.060
$RB_{\{i,j\}}$	0.079	$RB_{\{i,j\}}$	0.000	$RB_{\{i,j\}}$	0.066	$RB_{\{i,j\}}$	0.040
n=20		n=20		n=20		n=20	
$RB_{\{i,i\}}$	0.073	$RB_{\{i,i\}}$	0.068	$RB_{\{i,i\}}$	0.072	$RB_{\{i,i\}}$	0.058
$RB_{\{i,j\}}$	0.062	$RB_{\{i,j\}}$	0.048	$RB_{\{i,j\}}$	0.060	$RB_{\{i,j\}}$	0.044
n=40		n=40		n=40		n=40	
$RB_{\{i,i\}}$	0.073	$RB_{\{i,i\}}$	0.062	$RB_{\{i,i\}}$	0.073	$RB_{\{i,i\}}$	0.058
$RB_{\{i,j\}}$	0.061	$RB_{\{i,j\}}$	0.046	$RB_{\{i,j\}}$	0.159	$RB_{\{i,j\}}$	0.043
n=50		n=50		n=50		n=50	
$RB_{\{i,i\}}$	0.072	$RB_{\{i,i\}}$	0.004	$RB_{\{i,i\}}$	0.072	$RB_{\{i,i\}}$	0.058
$RB_{\{i,j\}}$	0.057	$RB_{\{i,j\}}$	0.037	$RB_{\{i,j\}}$	0.058	$RB_{\{i,j\}}$	0.043

If we move to analyzing the bias results for the scale MIDAS intercept matrix, Table 4.5 shows that under both sets of simulation exercises the estimator $\hat{\Lambda}(\theta, \omega)$ well approximates the true Λ matrix at all cross-section dimensions, with the parameter bias (averaged across diagonal and off-diagonal elements) clearly improving with increasing n and T . Again, the direct comparison of Panels A and B confirms that the algorithm initialized from the series of realized covariance matrices overall performs no worse than the one initialized from a nonparametric smoother.

Finally, as the final interest is in the overall accuracy of the model in fitting conditional (co)variances and correlations, we complement this section with an additional table that extends the Monte Carlo study to analyze the properties of the estimated in-sample series. Specifically, for each of the two exercises performed, we compare simulated and estimated variances, covariances and correlations in terms of mean, standard deviation, lower and upper quantiles. Moreover, we compare the performance, in terms of variability, of the equally weighted portfolios constructed employing the true and the estimated conditional covariance matrix.

Table 4.6 illustrates the strong similarity between the true series and the series obtained using the estimated model parameters, which supports the set of results discussed previously. Once again, no substantial difference can be detected between the two panels indexing the chosen initialization approach.

To summarize, the simulation study carried out in this section suggests that the proposed algorithm works quite accurately in finite samples and converges irrespective of the initialization choice made. Overall, the moment-based estimator used for iteratively targeting the constant intercept matrix in the secular component does not create a severe bias problem in the estimation of the other parameters, thus representing a feasible solution to alleviate the curse of dimensionality issue that would otherwise prevent the use of the MMReDCC model in high dimensional applications.

Both initialization methods for $M_{t,0}$ can be used in practice. In the empirical section, we have opted for the nonparametric smoother.

Table 4.6: **Properties of the estimator**

The table shows summary statistics of simulated (True) and estimated variance, covariance and correlation series, with values reported as averages across assets (n) and simulations (500). *Sim.I* denotes the first simulation exercise where $M_{t,0}$ is initialized from a nonparametric smoother while *Sim.II* uses the series of realized covariance matrices. In the last two columns we compare the properties of the volatilities of the n -dimensional equally weighted portfolios constructed using the true and the estimated conditional covariance matrix at each t . Q(1) and Q(3) respectively denote the lower (0.25) and the upper (0.75) quantile.

n	T	Stat.	Volatility (e-03)			Covariance (e-03)			Correlation			Portf. volatility		
			True	Sim.I	Sim.II	True	Sim.I	Sim.II	True	Sim.I	Sim.II	True	Sim.I	Sim.II
10	1000	Mean	0.924	0.932	0.938	0.109	0.109	0.110	0.116	0.116	0.118	0.014	0.014	0.015
		Std	0.217	0.220	0.230	0.148	0.149	0.147	0.149	0.148	0.145	0.002	0.002	0.001
		Q(1)	0.765	0.771	0.757	0.010	0.010	0.011	0.011	0.012	0.012	0.013	0.013	0.014
		Q(2)	1.050	1.060	1.059	0.203	0.204	0.205	0.223	0.222	0.243	0.015	0.015	0.016
	2000	Mean	0.995	1.006	0.962	0.156	0.157	0.157	0.151	0.150	0.148	0.014	0.014	0.014
		Std	0.172	0.181	0.202	0.118	0.122	0.121	0.114	0.115	0.117	0.001	0.001	0.001
		Q(1)	0.836	0.840	0.827	0.060	0.059	0.058	0.062	0.060	0.061	0.014	0.014	0.013
		Q(2)	1.128	1.145	1.150	0.244	0.247	0.220	0.240	0.240	0.238	0.015	0.015	0.015
	4000	Mean	0.963	0.971	0.972	0.136	0.136	0.136	0.139	0.138	0.138	0.013	0.013	0.013
		Std	0.157	0.161	0.162	0.110	0.111	0.111	0.108	0.108	0.108	0.001	0.001	0.001
		Q(1)	0.828	0.833	0.832	0.055	0.055	0.055	0.060	0.060	0.060	0.013	0.013	0.013
		Q(2)	1.075	1.086	1.087	0.210	0.211	0.212	0.218	0.218	0.218	0.014	0.014	0.014
20	1000	Mean	0.963	0.973	1.002	0.136	0.137	0.157	0.139	0.138	0.152	0.013	0.013	0.014
		Std	0.161	0.168	0.158	0.113	0.115	0.107	0.111	0.111	0.101	0.001	0.001	0.001
		Q(1)	0.828	0.832	0.854	0.055	0.055	0.072	0.060	0.059	0.075	0.013	0.013	0.013
		Q(2)	1.075	1.089	1.128	0.210	0.212	0.235	0.218	0.218	0.230	0.014	0.014	0.014
	2000	Mean	0.963	0.973	1.002	0.136	0.137	0.157	0.139	0.138	0.152	0.013	0.013	0.014
		Std	0.161	0.168	0.158	0.113	0.115	0.107	0.111	0.111	0.101	0.001	0.001	0.001
		Q(1)	0.828	0.832	0.854	0.055	0.055	0.072	0.060	0.059	0.075	0.013	0.013	0.013
		Q(2)	1.075	1.089	1.128	0.210	0.212	0.235	0.218	0.218	0.230	0.014	0.014	0.014
	4000	Mean	0.963	0.973	1.002	0.136	0.137	0.157	0.139	0.138	0.152	0.013	0.013	0.014
		Std	0.161	0.168	0.158	0.113	0.115	0.107	0.111	0.111	0.101	0.001	0.001	0.001
		Q(1)	0.828	0.832	0.854	0.055	0.055	0.072	0.060	0.059	0.075	0.013	0.013	0.013
		Q(2)	1.075	1.089	1.128	0.210	0.212	0.235	0.218	0.218	0.230	0.014	0.014	0.014
40	1000	Mean	1.040	1.048	1.294	0.188	0.189	0.235	0.176	0.175	0.178	0.014	0.015	0.015
		Std	0.123	0.131	0.805	0.087	0.088	0.215	0.079	0.079	0.076	0.001	0.001	0.001
		Q(1)	0.894	0.896	0.902	0.102	0.103	0.106	0.106	0.106	0.109	0.014	0.014	0.014
		Q(2)	1.168	1.181	1.285	0.262	0.264	0.288	0.245	0.244	0.248	0.015	0.015	0.015
	2000	Mean	1.046	1.042	1.059	0.191	0.198	0.195	0.178	0.179	0.179	0.015	0.015	0.015
		Std	0.123	0.125	0.129	0.088	0.089	0.088	0.078	0.078	0.078	0.001	0.001	0.001
		Q(1)	0.899	0.876	0.905	0.106	0.109	0.107	0.109	0.109	0.110	0.014	0.014	0.014
		Q(2)	1.176	1.057	1.194	0.266	0.274	0.271	0.247	0.251	0.248	0.015	0.015	0.015
	4000	Mean	1.078	1.086	1.098	0.212	0.213	0.221	0.191	0.190	0.195	0.015	0.015	0.015
		Std	0.115	0.123	0.120	0.082	0.082	0.080	0.071	0.071	0.068	0.001	0.001	0.001
		Q(1)	0.918	0.920	0.928	0.118	0.118	0.124	0.119	0.119	0.124	0.015	0.015	0.015
		Q(2)	1.222	1.234	1.249	0.294	0.296	0.304	0.262	0.261	0.266	0.016	0.016	0.016
50	1000	Mean	1.088	1.098	1.088	0.219	0.220	0.214	0.195	0.195	0.191	0.015	0.015	0.015
		Std	0.114	0.121	0.123	0.082	0.083	0.084	0.070	0.070	0.072	0.001	0.001	0.001
		Q(1)	0.925	0.928	0.922	0.123	0.123	0.119	0.124	0.123	0.120	0.015	0.015	0.015
		Q(2)	1.235	1.249	1.235	0.302	0.304	0.295	0.266	0.265	0.261	0.016	0.016	0.016
	2000	Mean	1.088	1.098	1.088	0.219	0.220	0.214	0.195	0.195	0.191	0.015	0.015	0.015
		Std	0.114	0.121	0.123	0.082	0.083	0.084	0.070	0.070	0.072	0.001	0.001	0.001
		Q(1)	0.925	0.928	0.922	0.123	0.123	0.119	0.124	0.123	0.120	0.015	0.015	0.015
		Q(2)	1.235	1.249	1.235	0.302	0.304	0.295	0.266	0.265	0.261	0.016	0.016	0.016
	4000	Mean	1.088	1.098	1.088	0.219	0.220	0.214	0.195	0.195	0.191	0.015	0.015	0.015
		Std	0.114	0.121	0.123	0.082	0.083	0.084	0.070	0.070	0.072	0.001	0.001	0.001
		Q(1)	0.925	0.928	0.922	0.123	0.123	0.119	0.124	0.123	0.120	0.015	0.015	0.015
		Q(2)	1.235	1.249	1.235	0.302	0.304	0.295	0.266	0.265	0.261	0.016	0.016	0.016

4.5 Multi-step Forecasting

Models featuring short and long-run dynamics are particularly attractive for computing multi-step-ahead predictions, as their dynamic component structure is conceivably expected to be beneficial for longer-term forecasts. Unfortunately, the complex nonlinear structure of the MMReDCC model makes the analytical derivation of closed-form solutions not possible. In order to overcome this limit, we propose to compute multi-step predictions by means of a procedure based on bootstrap resampling.

4.5.1 A conditional bootstrap (CB) procedure

At the outset, notice that Eq.(4.1) implies that $\mathbb{E}(C_t|\mathfrak{F}_{t-1}) = S_t$, so that C_t can be represented as

$$C_t = S_t^{1/2} U_t (S_t^{1/2})', \quad (4.9)$$

where U_t is an element of a sequence of iid random matrices with $\mathbb{E}(U_t) = I_n$, and $S_t^{1/2}$ denotes the lower-triangular Cholesky factor of S_t such that $S_t^{1/2}(S_t^{1/2})' = S_t$. If $U_t \sim W_n(\nu, I_n/\nu)$, the Wishart assumption of Eq.(4.1) is recovered, but this is not needed to justify the bootstrap procedure used for generating multi-step-ahead forecasts of the realized covariance matrix C_t . The procedure is described in the following six steps.

Step 1 Estimate the model on $\{C_t, t = 1, \dots, T\}$ and obtain the parameter vector $\hat{\psi} = \{vech(\hat{\Lambda}), \hat{\theta}, \hat{\omega}, \hat{\gamma}, \hat{\delta}, \hat{\alpha}, \hat{\beta}\}$ to compute the estimated conditional covariance $\hat{S}_t$.

Step 2 Compute the estimated residuals

$$\hat{U}_t = \hat{S}_t^{-1/2} C_t (\hat{S}_t^{-1/2})', \quad t = 1, \dots, T$$

and rescale them to enforce their sample mean to be equal to I_n :

$$\tilde{U}_t = (\hat{E}_u^{-1/2}) \hat{U}_t (\hat{E}_u^{-1/2})',$$

where $\hat{E}_u = (1/T) \sum_{t=1}^T \hat{U}_t$. The rescaled $\tilde{U}_t$ can then be used to generate bootstrap replicates of C_{T+j} , for $j = 1, \dots, h$, where h denotes the chosen forecast horizon.

Step 3 Draw with replacement a bootstrap sample $\{\tilde{U}_{T+1|T}, \dots, \tilde{U}_{T+h|T}\}$ of length h from the empirical CDF of $\{\tilde{U}_t, t = 1, \dots, T\}$.

Step 4 For $j = 1 \dots, h$, recursively generate a sequence of bootstrap replicates of C_{T+j}

as follows:

$$\begin{aligned}
M_{T+j|T} &= \hat{\Lambda}(\theta, \omega) + \hat{\theta} \sum_{k=1}^K \phi_k(\hat{\omega}) C_{T-k+j|T}, \\
L_{T+j|T} &= M_{T+j|T}^{1/2}, \\
C_{T+j|T}^* &= L_{T+j|T}^{-1} C_{T+j|T} (L_{T+j|T}')^{-1}, \\
P_{T+j|T}^* &= (\text{diag}\{C_{T+j|T}^*\})^{-1/2} C_{T+j|T}^* (\text{diag}\{C_{T+j|T}^*\})^{-1/2}, \\
S_{ii,T+j|T}^* &= (1 - \hat{\gamma}_i - \hat{\delta}_i) + \hat{\gamma}_i C_{ii,T+j-1|T}^* + \hat{\delta}_i S_{ii,T+j-1|T}^*, \\
R_{T+j|T}^* &= (1 - \hat{\alpha} - \hat{\beta}) I_n + \hat{\alpha} P_{T+j-1|T}^* + \hat{\beta} R_{T+j-1|T}^*, \\
S_{T+j|T}^* &= (\text{diag}\{S_{T+j|T}^*\})^{1/2} R_{T+j|T}^* (\text{diag}\{S_{T+j|T}^*\})^{1/2}, \\
S_{T+j|T} &= L_{T+j|T} S_{T+j|T}^* L_{T+j|T}', \\
C_{T+j|T} &= S_{T+j|T}^{1/2} \tilde{U}_{T+j} (S_{T+j|T}^{1/2})'.
\end{aligned}$$

The procedure is initialized at C_T , $\hat{R}_T$ and $\hat{S}_T$.

Step 5 Repeat steps 3–4 B times, where B is set sufficiently large (e.g. $B=5000$). As a result, the procedure generates an array of $h \times B$ bootstrap replicates $(C_{T+j|T}^{(1)}, \dots, C_{T+j|T}^{(B)})$ for $C_{T+j|T}$.

Step 6 Finally, the h -steps-ahead forecast is obtained as

$$\hat{S}_{T,h} = \frac{1}{B} \sum_{b=1}^B C_{T+h|T}^{(b)}.$$

It is worth to mention that the proposed procedure is applied by conditioning on the estimated model parameters, namely by keeping fixed the parameter estimates in all bootstrap forecasts of C_{T+j} , for $j = 1, \dots, h$, such that the achieved bootstrap h -steps ahead prediction depends only on the resampled residuals. Clearly, this could be relaxed in order to account for the variability associated to parameter estimation by re-estimating ψ on each bootstrap replicate, but it would come at the not negligible cost of increase computational complexity and time. Overall, the results presented in Section 4.5.2 provide sufficient evidence that our approach works fine in finite samples.

Finally, even if our primary interest is in forecasting from MMReDCC models, the proposed forecasting procedure is very general and can be readily adapted to any model that admits the representation in Eq.(4.9), where S_t is modeled as a function of past information $\mathfrak{F}_{t-1}$. For example, in the empirical application which is being presented in Section 4.6.2, we also use it to generate multi-step ahead forecasts of C_t from the cReDCC model of Bauwens et al. (2012). To this purpose, the dynamic equations in step 4 must be replaced by those pertaining to the specific model of interest (see Appendix 4.A for details on the adapted algorithm).

4.5.2 Finite sample properties

In order to analyze the finite sample behavior of the proposed bootstrap procedure, we devise a simple Monte Carlo experiment. Namely, we generate 1000 series with the MMReDCC model using the same simulation design used in Section 4.4, with the nonparametric smoother chosen as the initialization method for $M_{t,0}$ ⁴. The sample sizes considered are $T = 1000$ and 2000 and the cross-sectional dimensions are $n = 5, 10$ and 20 . In each case, we generate $j = 1, \dots, h$ future values of the simulated series, where $h = 2, 10$ and 20 , that are considered as the reference forecasting sample.

For each obtained set of B bootstrap replicates $(C_{T+j|T}^{(1)}, \dots, C_{T+j|T}^{(B)})$, we compute the corresponding prediction limits, defined as the quantiles of the bootstrap distribution function of $C_{T+j|T}^{(b)}$ ($b = 1, \dots, B$). More specifically, if $D_{c,B}(z)$ is an estimate of the distribution function $D_c(z) = Pr(C_{T+j|T} \leq z)$, then, a $100(1 - \vartheta)\%$ prediction interval for C_{T+j} is achieved as

$$[L_{C,B}(z), U_{C,B}(z)] = \left[Q_{C,B} \left(\frac{\vartheta}{2} \right), Q_{C,B} \left(1 - \frac{\vartheta}{2} \right) \right],$$

with $Q_{C,B} = D_{c,B}^{-1}$. Bootstrap intervals are constructed based on $B=999$ replicates with nominal coverages $1 - \vartheta$ equal to 0.90, 0.95 and 0.99. After this, we compute the empirical coverage by counting the number of future values inside the corresponding intervals as $(1 - \vartheta)^* = \{L_{C,B} \leq C_{T+j} \leq U_{C,B}\}$. In addition we compute left and right coverage as the proportion of predictions falling below $L_{C,B}$ and above $U_{C,B}$, respectively.

Results are reported in Table 4.7 as averages across univariate variance (left panel) and covariance (right panel) series. First of all, we can notice that results are qualitatively similar across the two panels. It emerges that the intervals for future volatilities and covolatilities at multiple-steps ahead have average coverages close to the nominal values, and that their performance improves as the sample size increases from 1000 to 2000 observations. For illustration, Figure 4.2 plots one realization of the predicted variance, for $j = 1, \dots, h$, together with corresponding point forecasts and bootstrap intervals at 90, 95 and 99% coverage. Note that the h -steps bootstrap forecast is bounded from below by construction, so that prediction intervals are generally larger to the right.

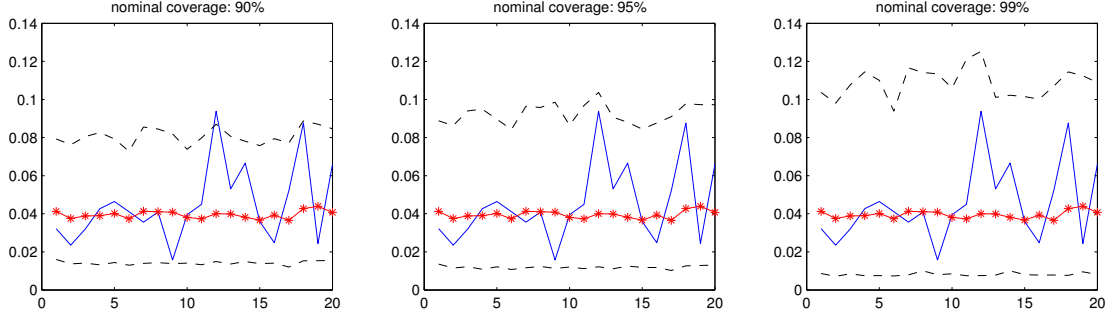
The table also shows that the coverage rates depend on the forecast horizon, and that they have a tendency to decrease as horizons increase. This comes as a consequence of the parameter estimation variability that is not accounted for, as well as from the addition of error uncertainty. However, even if for $h > 2$ there is a slight undercoverage, coverage values are never below their nominal levels by more than 2.7%, and this happens irrespective of the forecast horizon or the considered sample size.

Overall, Monte Carlo results show that the proposed bootstrap procedure is capable

⁴Results obtained employing the alternative initialization approach are qualitatively similar and thus not reported for brevity.

of generating accurate point and interval forecasts from the MMReDCC model.

Figure 4.2: Bootstrapped prediction intervals
90, 95 and 99% h -steps ahead prediction intervals ($h=1,\dots,20$) of future volatility together with real observations (continuous line) and point linear predictions ($-\star$).



4.6 Empirical Applications

This section contains two empirical applications. The first application provides the estimation results for the IMP estimator in comparison with the 1QML estimator in the ideal case where both can be computed. The second one is performed in a large dimensional system and aims at evaluating both the full-sample fit of the model and its forecasting performance. Specifically, we evaluate the ability of the MMReDCC model to provide accurate multi-steps-ahead covariance predictions against existing competitors not accounting for time-varying long term dynamics.

4.6.1 Small sample accuracy comparison

In small dimensional frameworks, according to Table 4.1, the 1QMLE is applicable and represents the most efficient estimator. Hence, a simple way to evaluate the in-sample performance of the proposed approach, is to compare the estimates provided by the IMP method to those obtained by maximizing the quasi-likelihood (QL) with respect to the full parameter vector. To this purpose, we fix the cross-sectional dimension equal to ten assets and fit the MMReDCC model to three different datasets. An overview of the data being used is given in Table 4.12 in the Appendix 4.A. The first dataset comprises the assets used in Chapter 3 and includes series of daily realized covariance matrices estimated using five minute intraday returns over the period February 2001 until December 2009 (2242 observations); the second set is made up of some of the most liquid equities of the S&P 500 traded from October 1997 to July 2008 (2524 observations) while the last one consists of an arbitrarily selected subsample from the dataset used in the work of Boudt et al. (2014). The latter contains series of daily realized covariance matrices obtained with the *CholCov* estimator over the period January 2007 until December 2012 (1499 trading days). As already mentioned before, the choice of

CHAPTER 4. FEASIBLE INFERENCE PROCEDURES FOR FORECASTING HIGH-DIMENSIONAL REALIZED COVARIANCE MATRICES

100

Table 4.7: Prediction intervals

The table reports prediction intervals for variances (left panel) and covariances (right panel) of MMRdCC model across 1000 simulations.

VARIANCES							COVARIANCES						
n	Lead time	T	Nominal coverage	Average coverage	Av.cov. below	Av.cov. above	n	Lead time	T	Nominal coverage	Average coverage	Av.cov. below	Av.cov. above
5	2	1000	0.90	0.891	0.054	0.054	5	2	1000	0.90	0.897	0.052	0.051
			0.95	0.945	0.026	0.029				0.95	0.946	0.028	0.026
			0.99	0.978	0.008	0.014				0.99	0.981	0.007	0.011
		2000	0.90	0.894	0.053	0.053			2000	0.90	0.906	0.046	0.048
			0.95	0.951	0.024	0.025				0.95	0.954	0.025	0.021
			0.99	0.982	0.005	0.012				0.99	0.982	0.007	0.011
	10	1000	0.90	0.878	0.064	0.059		10	1000	0.90	0.877	0.061	0.062
			0.95	0.932	0.036	0.032				0.95	0.930	0.035	0.035
			0.99	0.971	0.012	0.017				0.99	0.974	0.010	0.016
		2000	0.90	0.882	0.057	0.061			2000	0.90	0.882	0.060	0.058
			0.95	0.935	0.030	0.034				0.95	0.935	0.035	0.031
			0.99	0.973	0.009	0.018				0.99	0.978	0.010	0.013
	20	1000	0.90	0.873	0.062	0.065		20	1000	0.90	0.874	0.065	0.061
			0.95	0.932	0.033	0.035				0.95	0.930	0.036	0.036
			0.99	0.977	0.006	0.017				0.99	0.973	0.010	0.017
		2000	0.90	0.874	0.067	0.065			2000	0.90	0.879	0.064	0.057
			0.95	0.923	0.035	0.041				0.95	0.933	0.036	0.034
			0.99	0.970	0.008	0.022				0.99	0.971	0.012	0.018
10	2	1000	0.90	0.891	0.053	0.057	10	2	1000	0.90	0.889	0.056	0.055
			0.95	0.940	0.028	0.031				0.95	0.943	0.028	0.029
			0.99	0.979	0.007	0.015				0.99	0.980	0.007	0.013
		2000	0.90	0.895	0.054	0.052			2000	0.90	0.895	0.053	0.053
			0.95	0.948	0.027	0.025				0.95	0.946	0.028	0.027
			0.99	0.983	0.005	0.013				0.99	0.981	0.007	0.012
	10	1000	0.90	0.876	0.060	0.064		10	1000	0.90	0.883	0.060	0.058
			0.95	0.933	0.030	0.037				0.95	0.935	0.032	0.033
			0.99	0.975	0.007	0.018				0.99	0.975	0.010	0.016
		2000	0.90	0.883	0.060	0.057			2000	0.90	0.880	0.058	0.060
			0.95	0.938	0.030	0.030				0.95	0.935	0.032	0.032
			0.99	0.978	0.006	0.014				0.99	0.977	0.009	0.014
	20	1000	0.90	0.873	0.065	0.066		20	1000	0.90	0.873	0.064	0.063
			0.95	0.930	0.033	0.037				0.95	0.932	0.037	0.035
			0.99	0.975	0.007	0.018				0.99	0.973	0.011	0.016
		2000	0.90	0.876	0.060	0.063			2000	0.90	0.870	0.063	0.064
			0.95	0.930	0.033	0.037				0.95	0.934	0.035	0.038
			0.99	0.974	0.008	0.017				0.99	0.972	0.010	0.017
20	2	1000	0.90	0.892	0.049	0.059	20	2	1000	0.90	0.886	0.057	0.057
			0.95	0.943	0.025	0.032				0.95	0.939	0.030	0.031
			0.99	0.981	0.006	0.014				0.99	0.979	0.007	0.013
		2000	0.90	0.892	0.051	0.057			2000	0.90	0.887	0.057	0.056
			0.95	0.946	0.026	0.029				0.95	0.940	0.030	0.030
			0.99	0.983	0.005	0.012				0.99	0.979	0.007	0.013
	10	1000	0.90	0.880	0.055	0.065		10	1000	0.90	0.876	0.062	0.062
			0.95	0.935	0.029	0.036				0.95	0.931	0.034	0.034
			0.99	0.976	0.007	0.017				0.99	0.975	0.009	0.016
		2000	0.90	0.883	0.054	0.063			2000	0.90	0.876	0.061	0.062
			0.95	0.936	0.028	0.036				0.95	0.932	0.034	0.034
			0.99	0.975	0.008	0.017				0.99	0.975	0.009	0.016
	20	1000	0.90	0.877	0.060	0.063		20	1000	0.90	0.873	0.066	0.068
			0.95	0.933	0.032	0.036				0.95	0.930	0.037	0.038
			0.99	0.972	0.009	0.019				0.99	0.970	0.011	0.018
		2000	0.90	0.877	0.059	0.064			2000	0.90	0.875	0.065	0.066
			0.95	0.933	0.033	0.034				0.95	0.931	0.037	0.037
			0.99	0.977	0.007	0.016				0.99	0.972	0.011	0.017
50	2	1000	0.90	0.899	0.045	0.056	50	2	1000	0.90	0.886	0.057	0.057
			0.95	0.950	0.023	0.027				0.95	0.939	0.030	0.030
			0.99	0.985	0.005	0.010				0.99	0.979	0.007	0.013
		2000	0.90	0.897	0.050	0.053			2000	0.90	0.889	0.055	0.056
			0.95	0.948	0.025	0.026				0.95	0.942	0.029	0.029
			0.99	0.985	0.005	0.010				0.99	0.980	0.007	0.013
	10	1000	0.90	0.888	0.051	0.061		10	1000	0.90	0.874	0.063	0.063
			0.95	0.940	0.027	0.033				0.95	0.931	0.034	0.035
			0.99	0.979	0.007	0.015				0.99	0.975	0.009	0.016
		2000	0.90	0.888	0.056	0.057			2000	0.90	0.878	0.061	0.061
			0.95	0.941	0.028	0.030				0.95	0.934	0.033	0.033
			0.99	0.980	0.007	0.013				0.99	0.977	0.008	0.015
	20	1000	0.90	0.879	0.054	0.067		20	1000	0.90	0.864	0.068	0.068
			0.95	0.934	0.029	0.037				0.95	0.923	0.038	0.038
			0.99	0.977	0.007	0.017				0.99	0.971	0.011	0.018
		2000	0.90	0.880	0.057	0.063			2000	0.90	0.868	0.066	0.066
			0.95	0.935	0.031	0.034				0.95	0.926	0.037	0.036
			0.99	0.977	0.007	0.016				0.99	0.973	0.010	0.017

the realized estimator is not an issue here as the model can be fitted to any series of realized variance-covariance matrices as long as they are guaranteed to be PDS.

As suggested by a Referee, the IMP estimator could also be used to provide accurate initial values for direct QML estimation leading to a reduction of the number of iterations needed to reach convergence and, hence, to substantial computational savings. In this spirit we also consider the additional estimator, denoted as IMP(+), obtained performing a one-step QL optimization taking the IMP estimates as starting points.

Estimation results for the MMReDCC model parameters by the 1QML, IMP and IMP(+) estimators are collected in Panel A of Table 4.8. In the three datasets considered, all methods appear to deliver similar estimates. Short term GARCH coefficients tend to be quite homogeneous across assets and generally significant; the same applies to the short term correlation estimates. As for the parameters driving the long term component, it can be noticed that the estimated θ and ω coefficients are regularly lower for the IMP than for the 1QML method. This is in line with the prevailing negative bias reported from the simulation study. Coming to the analysis of the maximized quasi log-likelihood values, it can be seen that, as expected, the 1QMLE returns the highest QL value for all the datasets, but those obtained by the IMP estimator are very close, and the gap never exceeds 0.04% in relative terms. The IMP(+) further reduces the discrepancy but its contribution is as small as 0.014% on average, thus far from impressive.

The bottom line of Panel A reports test statistics and corresponding p-values of a simple score test (ST) performed to assess convergence of the IMP and IMP(+) estimators. In practice we test the null hypothesis that $\psi^* = \mathbf{0}$ in the unrestricted model parameterized by $\psi = \hat{\psi}_M + \psi^*$ where $\hat{\psi}_M$ denotes the estimate of ψ obtained by estimation method M . In practice ψ^* can be interpreted as the bias potentially affecting the estimated ψ in case of lack of convergence of the IMP and IMP+ algorithms.

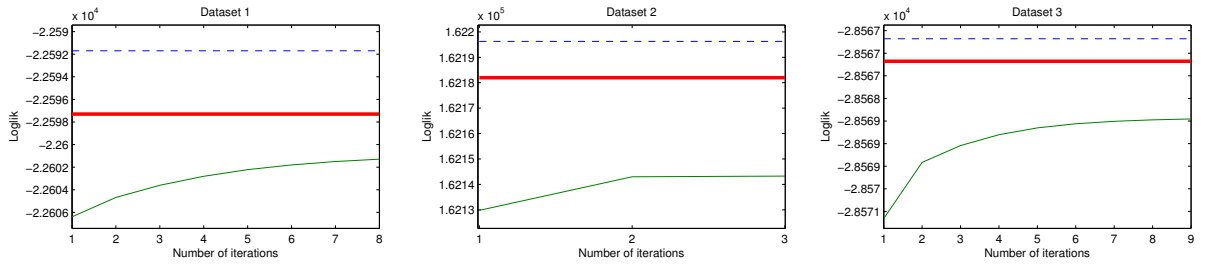
In order to double check our results the test is repeated for the 1QML estimator. The null is accepted in all cases. This result is confirmed by Figure 4.3 that compares the values of the QL function recorded for the IMP estimator in each iteration (continuous green line) with the maximum obtained by direct maximization of the QL function (continuous red line) and by the IMP(+) estimator (dotted blue line). The plot shows that the IMP algorithm increases the value of the QL function at each step monotonically converging to a value which is very close to the maximum yielded by the direct QML estimator.

Finally, Panel B of Table 4.8 provides further information on the performance of the estimators measured in terms of mean, standard deviation, first and third quantile of the estimated conditional variance, covariance and correlation series. Furthermore, to gain deeper insight on the practical impact that the choice of the estimation algorithm can have on risk management applications, the comparison is extended to the estimated conditional variances of the equally weighted portfolio returns. In general, the distribution of the estimated series do not appear to be very sensitive to the estima-

tion method adopted. The most sizeable differences are observable between 1QML and IMP(+) estimated conditional variances and covariances of Dataset 3. Nevertheless, the discrepancy becomes negligible if we focus on conditional correlations and, in particular, on portfolio volatility which is the main quantity of interest for risk management applications.

Figure 4.3: Maximized likelihood values

The figure shows the log-likelihood value at the maximum (y-axis) achieved by the 1-step QMLE (dotted blue line), the IMPE (continuous green line) and the IMPE after an additional Newton-Raphson iteration (continuous red line), for each of the considered datasets. The number of iterations needed for the IMPE to converge is given on the x-axis (8 for Dataset 1, 3 for Dataset 2 and 9 for Dataset 3).



4.6.2 Forecasting performance

In this subsection we push the analysis to higher dimensions, with the aim of assessing the usefulness of the MMReDCC model in a forecasting framework. As benchmarks we consider the consistent ReDCC (cReDCC) model of [Bauwens et al. \(2012\)](#) as the closest competitor⁵ and a simple Exponentially Weighted Moving Average (EWMA) model. The EWMA predictor appears a natural candidate due to its widespread diffusion among practitioners and in risk management systems like RiskMetrics. If applied to series of daily realized covariance matrices, it is defined by

$$S_t = (1 - \lambda)C_{t-1} + \lambda S_{t-1},$$

where the λ parameter is set equal to the value 0.94 (see also [Golosnoy et al. \(2012\)](#)).

On the other hand, the choice of the cReDCC as a benchmark is supported by two main reasons. First, it assumes that conditional volatilities and correlations mean revert to constant quantities, thus it can be considered as a simplified version of the MMReDCC model despite not being formally nested in it. Second, the findings of [Boudt et al. \(2014\)](#) show that the cReDCC model favorably compares with some widely used competitors, such as the HEAVY ([Noureldin et al. \(2012\)](#)) and the cDCC ([Aielli \(2013\)](#)) model, in forecasting Value-at-Risk. In order to estimate the cReDCC in high dimension, we apply a three stage QML estimation procedure as suggested by [Bauwens et al. \(2012\)](#), where the constant long term covariance matrix is consistently targeted

⁵See Appendix C for details.

Table 4.8: **Application I. In-sample comparison**

Panel A shows parameter estimates with corresponding standard errors in brackets for each of the estimators employed and the three datasets considered (see table A.12 in the Appendix). IMP(+) denotes results achieved after one iteration of the one-step QL optimization, taking the IMP estimate as starting point. The last two rows respectively report the log-likelihood (Loglik) values at the maximum and Score test (ST) statistics with corresponding p-values in brackets. Panel B reports summary statistics for the estimated series of volatilities (Vol), covariances (Cov) and correlations (Corr) across estimators and datasets. The properties of the volatilities (Portf. vol) of the n -dimensional equally weighted portfolios constructed using the estimated conditional covariance matrices are also compared. The number of in-sample observations is 2242 for Dataset 1, 2524 for Dataset 2 and 1499 for Dataset 3. Realized series of the assets comprised in Dataset 2 multiplied by 10000.

DATASET 1						DATASET 2						DATASET 3							
PANEL A: PARAMETER ESTIMATES																			
1QML		IMP		IMP(+)		1QML		IMP		IMP(+)		1QML		IMP		IMP(+)			
γ_i	δ_i	γ_i	δ_i	γ_i	δ_i	γ_i	δ_i	γ_i	δ_i	γ_i	δ_i	γ_i	δ_i	γ_i	δ_i	γ_i	δ_i		
0.35 (0.10)	0.60 (0.03)	0.34 (0.06)	0.59 (0.03)	0.37 (0.15)	0.55 (0.08)	0.12 (0.04)	0.88 (0.19)	0.10 (0.06)	0.85 (0.10)	0.09 (0.03)	0.82 (0.15)	0.18 (0.07)	0.75 (0.06)	0.18 (0.16)	0.75 (0.07)	0.17 (0.09)	0.75 (0.05)		
0.34 (0.38)	0.58 (0.08)	0.31 (0.76)	0.60 (0.05)	0.33 (0.86)	0.58 (0.07)	0.32 (0.15)	0.80 (0.18)	0.33 (0.07)	0.83 (0.12)	0.29 (0.05)	0.82 (0.18)	0.23 (0.20)	0.69 (0.04)	0.23 (0.90)	0.69 (0.05)	0.24 (0.92)	0.68 (0.04)		
0.41 (0.04)	0.57 (0.06)	0.40 (0.03)	0.52 (0.05)	0.41 (0.05)	0.51 (0.09)	0.33 (0.08)	0.61 (1.01)	0.28 (0.07)	0.55 (0.10)	0.33 (0.30)	0.52 (0.43)	0.24 (0.04)	0.62 (0.06)	0.25 (0.16)	0.62 (0.21)	0.25 (0.06)	0.61 (0.09)		
0.43 (0.06)	0.53 (0.06)	0.38 (0.04)	0.49 (0.06)	0.41 (0.06)	0.48 (0.07)	0.26 (0.03)	0.43 (0.18)	0.24 (0.06)	0.41 (0.01)	0.28 (0.12)	0.38 (0.06)	0.20 (0.05)	0.67 (0.07)	0.19 (0.06)	0.69 (0.07)	0.20 (0.07)	0.68 (0.06)		
0.37 (0.05)	0.49 (0.05)	0.37 (0.03)	0.49 (0.04)	0.40 (0.05)	0.48 (0.06)	0.46 (0.08)	0.18 (0.09)	0.49 (0.14)	0.23 (0.05)	0.50 (0.09)	0.22 (0.06)	0.20 (0.06)	0.60 (0.09)	0.21 (0.11)	0.65 (0.12)	0.20 (0.06)	0.59 (0.10)		
0.34 (0.05)	0.65 (0.12)	0.32 (0.04)	0.61 (0.06)	0.33 (0.06)	0.60 (0.06)	0.35 (0.14)	0.76 (0.25)	0.34 (0.20)	0.79 (0.28)	0.35 (0.15)	0.78 (0.20)	0.18 (0.11)	0.69 (0.20)	0.18 (0.09)	0.73 (0.13)	0.19 (0.10)	0.68 (0.18)		
0.40 (0.05)	0.52 (0.11)	0.42 (0.04)	0.51 (0.05)	0.44 (0.06)	0.49 (0.10)	0.43 (0.09)	0.52 (0.09)	0.23 (0.10)	0.44 (0.62)	0.31 (0.10)	0.36 (0.09)	0.15 (0.09)	0.68 (0.23)	0.14 (0.08)	0.74 (0.19)	0.16 (0.10)	0.66 (0.23)		
0.47 (0.04)	0.43 (0.04)	0.45 (0.03)	0.40 (0.05)	0.45 (0.04)	0.40 (0.05)	0.45 (0.21)	0.72 (0.17)	0.43 (0.16)	0.70 (0.01)	0.45 (0.19)	0.68 (0.03)	0.18 (0.07)	0.75 (0.16)	0.18 (0.30)	0.75 (0.40)	0.18 (0.08)	0.75 (0.16)		
0.31 (0.06)	0.61 (0.08)	0.31 (0.04)	0.59 (0.05)	0.32 (0.04)	0.57 (0.08)	0.18 (0.06)	0.59 (0.19)	0.16 (0.15)	0.63 (0.04)	0.20 (0.17)	0.60 (0.13)	0.18 (0.10)	0.74 (0.27)	0.17 (0.14)	0.75 (0.23)	0.17 (0.20)	0.74 (0.50)		
0.34 (0.06)	0.55 (0.09)	0.31 (0.04)	0.58 (0.06)	0.35 (0.05)	0.56 (0.08)	0.43 (0.11)	0.36 (0.09)	0.47 (0.17)	0.27 (0.01)	0.48 (0.12)	0.28 (0.10)	0.14 (0.04)	0.78 (0.05)	0.14 (0.06)	0.80 (0.10)	0.14 (0.04)	0.78 (0.05)		
α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β	α	β		
0.07 (0.02)	0.91 (0.08)	0.08 (0.01)	0.88 (0.06)	0.07 (0.02)	0.88 (0.08)	0.05 (0.01)	0.86 (0.16)	0.04 (0.01)	0.89 (0.10)	0.05 (0.00)	0.87 (0.11)	0.01 (0.00)	0.80 (0.10)	0.01 (0.00)	0.88 (0.05)	0.01 (0.00)	0.77 (0.05)		
θ	ω	θ	ω	θ	ω	θ	ω	θ	ω	θ	ω	θ	ω	θ	ω	θ	ω		
0.88 (0.01)	4.57 (0.01)	0.83 (0.00)	4.22 (0.01)	0.87 (0.00)	5.69 (0.03)	0.91 (0.01)	3.67 (0.02)	0.88 (0.02)	2.91 (0.07)	0.90 (0.06)	3.46 (0.19)	0.84 (0.02)	5.52 (1.58)	0.81 (0.18)	5.16 (0.57)	0.82 (0.01)	5.56 (0.44)		
Loglik		Loglik		Loglik		Loglik		Loglik		Loglik		Loglik		Loglik		Loglik			
-22592		-22601		-22597		162196		162143		162180		-28567		-28568		-28567			
ST		ST		ST		ST		ST		ST		ST		ST		ST			
-0.001 (1.00)		-0.026 (1.00)		-0.001 (1.00)		0.103 (1.00)		0.345 (1.00)		0.290 (1.00)		0.000 (1.00)		0.030 (1.00)		0.001 (1.00)			
PANEL B: PROPERTIES OF ESTIMATED SERIES																			
		1QML	IMP	IMP(+)			1QML	IMP	IMP(+)			1QML	IMP	IMP(+)			1QML	IMP	IMP(+)
Vol	mean	2.98	3.06	3.14	Vol	mean	3.60	3.59	3.91	Vol	mean	17.33	12.83	11.75	Vol	mean	17.33	12.83	11.75
	std	5.70	5.84	5.89		std	8.66	9.03	9.50		std	23.93	19.86	18.35		std	23.93	19.86	18.35
	Q1	0.70	0.70	0.74		Q(1)	0.95	1.07	1.19		Q1	4.38	3.30	3.07		Q1	4.38	3.30	3.07
	Q3	2.60	2.69	2.78		Q(2)	4.12	4.28	4.70		Q3	12.19	11.15	10.23		Q3	12.19	11.15	10.23
Cov	mean	1.20	1.24	1.26	Cov	mean	0.67	0.71	0.73	Cov	mean	7.16	5.20	4.81	Cov	mean	7.16	5.20	4.81
	std	2.45	2.47	2.49		std	1.53	1.85	1.96		std	8.94	7.03	6.53		std	8.94	7.03	6.53
	Q1	0.20	0.21	0.22		Q(1)	0.24	0.26	0.28		Q1	1.85	1.23	1.14		Q1	1.85	1.23	1.14
	Q3	1.03	1.09	1.10		Q(2)	0.81	0.87	0.92		Q3	7.17	5.21	4.76		Q3	7.17	5.21	4.76
Corr	mean	0.37	0.39	0.38	Corr	mean	0.26	0.27	0.25	Corr	mean	0.42	0.42	0.42	Corr	mean	0.42	0.42	0.42
	std	0.10	0.10	0.10		std	0.12	0.13	0.12		std	0.08	0.09	0.09		std	0.08	0.09	0.09
	Q1	0.30	0.31	0.31		Q(1)	0.17	0.17	0.17		Q1	0.36	0.35	0.35		Q1	0.36	0.35	0.35
	Q3	0.44	0.45	0.44		Q(2)	0.35	0.36	0.35		Q3	0.48	0.48	0.48		Q3	0.48	0.48	0.48
Portf.	mean	0.95	0.97	0.99	Portf.	mean	0.91	0.92	0.95	Portf.	mean	2.29	2.10	2.02	Portf.	mean	2.29	2.10	2.02
	std	0.69	0.69	0.69		std	0.38	0.37	0.37		std	1.40	1.25	1.20		std	1.40	1.25	1.20
	Q1	0.52	0.53	0.54		Q(1)	0.60	0.62	0.63		Q1	1.47	1.20	1.16		Q1	1.47	1.20	1.16
	Q3	1.10	1.13	1.14		Q(2)	1.10	1.11	1.13		Q3	2.63	2.41	2.30		Q3	2.63	2.41	2.30

by the unconditional covariance. This drastically reduces the number of parameters to be estimated to $2n + 2$.

The dataset, also used by [Laurent et al. \(2012\)](#), contains realized covariance matrices based on intraday returns computed from 6-minute intervals last mid-quotes of 46 assets traded in the NYSE and NASDAQ over the period January 5, 1999 to November 14, 2008, for a total of 2483 observations.

Table 4.9 reports parameter estimates obtained by fitting the MMReDCC and cReDCC models over the full sample period. As emerges from Panel A, the MMReDCC outperforms the cReDCC in terms of the AIC and BIC criteria, which are both minimized for the MMReDCC. The univariate GARCH(1,1) parameters $\bar{\gamma}$ and $\bar{\delta}$, reported in averaged values across series, largely agree with each other, while the correlation estimates are slightly different across the two models, with the cReDCC showing a higher level of persistence.

To closely examine the difference in the fit of the models, consider the estimated conditional correlations between two representative pairs of stocks. The first, presented in the upper panel of Figure 4.4, includes two financial assets: American Express (AXP) and Bank of America (BAC), while the second pair, in the lower panel, includes stocks from different sectors, i.e. McDonald's (MCD) and Wells & Fargo (WFC). Analogously to what shown by Figure 3.5 in Chapter 3, in general the correlations returned by the MMReDCC model appear to be characterized by more pronounced fluctuations. At the beginning of the sample the correlation paths from the two models evolve around the same mean level while at the burst of the dot-com bubble in 2002 the MMReDCC correlations appear to be characterized by a positive level shift which is not present in the cReDCC series. For the reminder of the sample the cReDCC correlations are on average lower than those obtained from the MMReDCC. Given the close similarity between the models, this can be reasonably explained by the fact that the parameters θ and ω driving the long term co-volatilities dynamics of the MMReDCC allow for a higher flexibility of the model and consequently for a better responsiveness of correlations in periods of pronounced market volatility.

To determine whether the MMReDCC model can lead to gains in forecasting accuracy we compute predictions of the conditional covariance matrix of daily returns at 1, 5, 10 and 20 steps-ahead making use of the bootstrap procedure explained in Section 4.5.⁶ A similar approach is also applied to the cReDCC model, while predictions from the EWMA are obtained analytically, since this model implies that $\mathbb{E}(C_{t+h}|\mathfrak{F}_t) = \mathbb{E}(C_{t+h-1}|\mathfrak{F}_t)$.

To shorten computational time, estimation is performed using a fixed-rolling window scheme with window length equal to 1483 observations that shifts forward every twenty days, over which parameter estimates are kept fixed. The number of re-estimations of each model is equal to fifty, and the out-of-sample evaluation is performed on 1000 trading days.

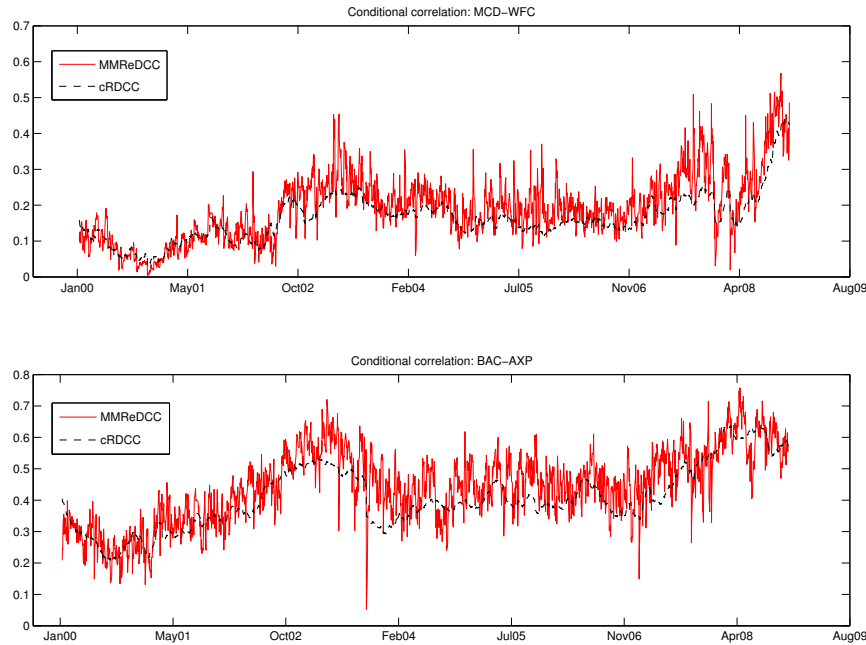
⁶We performed the same exercise using a number of bootstrap replicates equal to 1000, 2000 and 5000, without achieving qualitatively different results. Those we report are for $B = 5000$.

Table 4.9: **Application II. Full sample estimates and implemented loss functions**

Panel A reports full sample estimates from the MMRcDCC and cRDCC model, where $\bar{\gamma}$ and $\bar{\delta}$ are averaged across the n series. AIC and BIC criteria have been rescaled by the number of observations. Panel B contains the loss functions chosen to evaluate the models forecasting ability. $\hat{S}_t$ denotes the predicted conditional covariance matrix while C_t is the 6-minute realized measure chosen as proxy for the latent covariance matrix.

PANEL A: FULL SAMPLE ESTIMATES		
	MMReDCC	cRDCC
$\bar{\gamma}$	0.381 (0.087)	0.373 (0.180)
$\bar{\delta}$	0.543 (0.110)	0.551 (0.201)
α	0.016 (0.001)	0.020 (0.005)
β	0.950 (0.002)	0.974 (0.007)
θ	0.761 (0.027)	
ω	3.278 (0.785)	
Loglik	787304	683817
AIC	-633	-617
BIC	-630	-616
PANEL B: LOSS FUNCTIONS		
ST	Stein	$\text{tr}(S_t^{-1}C_t) - \log S_t^{-1}C_t - n$
vND	von Neumann Divergence	$\text{tr}(C_t \log C_t - C_t \log S_t - C_t + S_t)$
QLIK	Qlike	$\log S_t + \text{tr}(S_t^{-1}C_t)$

Figure 4.4: Estimated conditional correlation of MCD-WFC (upper panel) and of AXP-BAC (lower panel).



The forecasting period is characterized by drastic changes in volatility dynamics, as emerges from the summary statistics given in Table 4.13 in Appendix 4.A. To better analyze to what extent this impacts on the performance of the models, we break the evaluation sample into two sub-samples. Their differences can be visualized by looking at Figure 4.5, which shows the realized variance of the equally weighted portfolio made of the 46 assets used in the application. The upper panel spans the period from November,

2004 until end of June, 2007, where the market experiences a situation of stability after the turmoil of the 2000-2003 dot-com bubble. On the other hand, the period from July, 2007 to November, 2008, highlighting a widespread turbulence on the market, coincides with the burst of the recent financial crisis which reaches its peak with the collapse of the Lehman Brothers in September 2008. During the last four months of the latter sub-sample, the unconditional volatility of the portfolio is roughly ten times higher than over the first sub-period.

The comparison of the models forecasting ability is performed using the consistent⁷ loss functions defined in Panel B of Table 4.9, for which we report averaged values over the two out-of-sample periods considered. The test of [Giacomini & White \(2006\)](#) (GW) is used to examine the relative performance of the MMRéDCC model with respect to the cReDCC and the EWMA. Namely, for each loss function L , the loss differential at time t is denoted as $\delta_t = L(C_{t+h}, S_{t+h|t}^{MMReDCC}) - L(C_{t+h}, S_{t+h|t}^{benchmark})$, where C_t is the 6-minute realized covariance chosen as proxy for the true matrix. This difference is expected to be zero if neither of the models is superior, otherwise, negative (positive) values correspond to a superior forecasting performance of the MMRéDCC (benchmark). In addition, consistently with the in-sample analysis performed in Section 4.6.1, we also assess, using the univariate version of the QLIK loss function⁸, the ability of the different models to accurately predict the volatility of an equally weighted portfolio including all the 46 assets.

The null of equal predictive accuracy is implemented as the following t-statistic

$$GW = \frac{\sum_{t=\tau+h}^T \delta_t}{\sqrt{T - \tau + h} \sqrt{\text{avar}(\delta_t)}},$$

and the Newey-West estimator is used to consistently estimate the long-run variance of δ_t .

Table 4.10 holds the resulting average loss differences summarized by horizon. The value achieved by either the EWMA or the cReDCC is in italic if the model is favored by the GW test, in bold if the test favors the MMRéDCC or underlined if the test is indecisive.

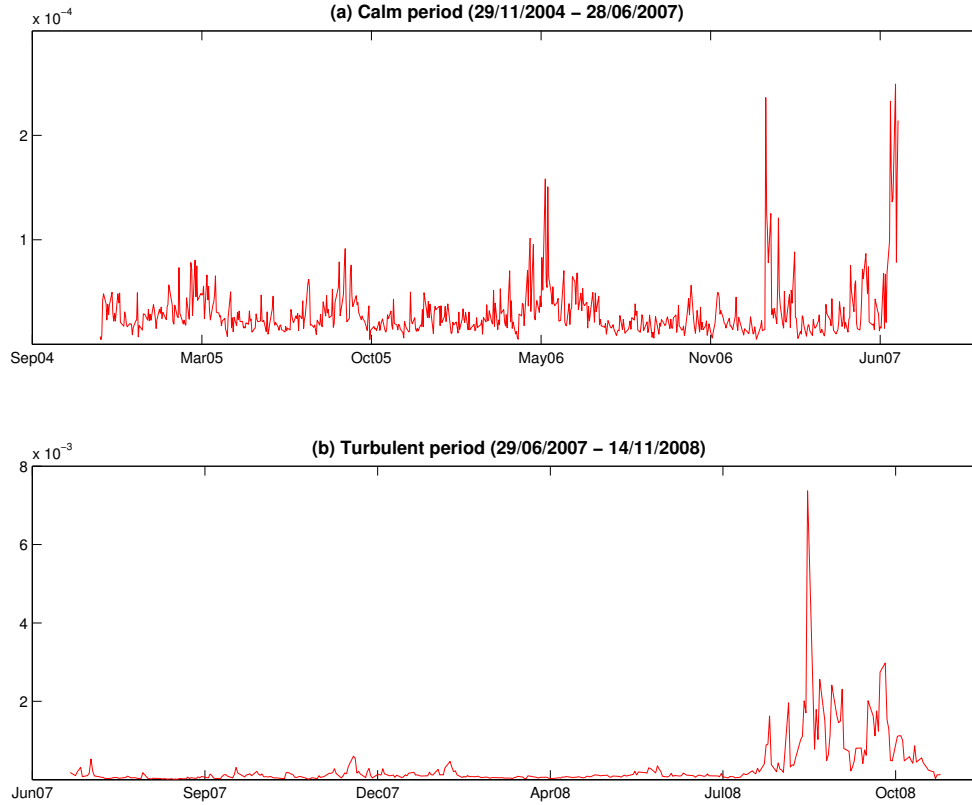
According to Panel A, the MMRéDCC model is not leading to particularly impressive gains in forecasting accuracy compared to the two benchmarks. When focusing on direct evaluation of forecasts of the whole RC matrix, the cReDCC is prevailing at the shortest horizon according to all loss functions, while for $h > 1$ they fail to point towards a unique winner as the overall performance of the models is pretty similar and the test is often inconclusive. Despite being the simplest model, the EWMA is found to perform no worse than the other two and to be preferred twice by the vND

⁷The term *consistent* is used according to [Laurent et al. \(2013\)](#).

⁸As given by [Patton \(2011\)](#), the formula of the QLIK loss function for the volatility of the i -th asset is $QLIK_t = \log(S_{ii,t}) + \frac{C_{ii,t}}{S_{ii,t}}$, whereas its portfolio version is obtained by replacing $C_{ii,t}$ and $S_{ii,t}$ by their portfolio equivalents.

Figure 4.5: Equally weighted portfolio daily realized variance

The figure shows the daily realized variance over the forecasting period of the equally wighted portfolio composed of the 46 assets used in the empirical application and listed in Table 4.13. The sample is divided into a more calm period (upper panel, 650 observations) and a more volatile period (lower panel, 350 observations).



over the MMRdCC. Considering that the period covered by Panel A is characterized by a relatively small and slow-moving market volatility, these results are probably not surprising for two main reasons. First, it is known that in such circumstances highly parameterized sophisticated models suffer from additional parameter uncertainty, thus being more heavily penalized than model featuring simple parameterizations. Second, when the underlying process exhibits smooth dynamics, it is more complicated for the MMRdCC model to disentangle the different volatility components, thus suggesting that accurate predictions can be obtained by employing models that do not necessarily account for time-varying long run levels.

As we move to analyze the results in Panel B, the situation is quickly reversed. The cReDCC is still minimizing two out of three loss functions at the one-step horizon, but in all other cases it is evident that, whenever the models predictive abilities can be distinguished, the GW test decides in favor of the MMRdCC. The out-performance of the MMRdCC is particularly strong at the 20-day horizon, when it delivers the optimal covariance forecasts according to the whole set of losses.

A slightly different situation arises when we move to the analysis of portfolio volatility forecasts. In this case the forecasting performance of the MMRdCC model appears to be more stable over time. In period A, at the 1-day horizon the MMRdCC is

outperforming the EWMA and doing not significantly worse than the cReDCC. At longer horizons, it is always prevailing over its competitors, the only exception being represented by the cReDCC, for which at horizon 20 we cannot reject the null of equal predictive ability. In period B, again the worst performance is recorded for the 1-day horizon but, for longer horizons, the MMReDCC is performing significantly better than its competitors in all cases except for the 10-day horizon when the comparison with the EWMA results indecisive.

Table 4.11 offers a closer inspection of the models relative performance by reporting GW test results across univariate co-volatilities and correlations. For sake of space, only the QLIK case is considered. Each panel of the table records the number of series (out of 1081 co-volatilities and 1035 correlations) for which the test favors the MMReDCC, one of the benchmarks or gives no decision. Results are mostly in accordance with those achieved for the whole covariance matrix and stress the evidence that in periods of calm (Panel A) there is almost no benefit from employing the MMReDCC model for co-volatility prediction. The gain in terms of correlations is marginal and only achieved with respect to the EWMA. On the other hand, sensibly better results are obtained over the final period (Panel B), both in terms of co-volatility and correlation dynamics. Noticeably, the number of cases favoring the MMReDCC increases with h and becomes striking already at $h = 10$.

Overall, the main message we can get from these empirical results is that while constant long-term models may be preferred in moderately volatile time periods, the benefits from the MMReDCC model can be fully appreciated in periods of market instability. In those cases, the flexibility of the model assures a higher responsiveness and more reliable out-of-sample forecasts.

Table 4.10: Multi-step-ahead forecast evaluation

The table reports averaged values of the loss functions listed in Table 4.9. Results are reported across the out-of-sample period divided into a more calm period (Panel A, November 2004 to July 2007) and a more turbulent period (Panel B, July 2007 to November 2008). In the last row of each panel we also report the averaged QLIK of the out-of-sample variance of the equally weighted portfolio obtained using the predicted covariance matrices of the models. For each horizon, we perform pairwise Giacomini-White (GW) tests for the significance at 5% level of the loss difference between the MMReDCC and each benchmark: the competitor is in italic if it is favored by the test, in bold if the MMReDCC is favored and underlined if the test is indecisive.

	Horizon 1			Horizon 5			Horizon 10			Horizon 20		
	MMReDCC	EWMA	cReDCC	MMReDCC	EWMA	cReDCC	MMReDCC	EWMA	cReDCC	MMReDCC	EWMA	cReDCC
PANEL A: 29-11-2004/01-07-2007												
ST	37.37	37.52	<i>36.56</i>	38.08	38.62	38.39	38.80	<u>39.24</u>	39.45	39.05	<u>40.01</u>	41.43
vND	0.003	<u>0.003</u>	<i>0.003</i>	0.003	<i>0.003</i>	<u>0.003</u>	0.003	0.003	<u>0.003</u>	0.003	<i>0.003</i>	0.004
QLIK	-386.62	<u>-386.46</u>	<i>-386.43</i>	-385.74	-385.20	-385.43	-385.02	<u>-384.58</u>	-384.38	-383.96	<u>-383.79</u>	-382.38
Portf. vol	-9.69	-9.55	<u>-9.69</u>	-9.75	-9.49	-9.70	-9.789	-9.48	-9.75	-9.83	-9.46	<u>-9.84</u>
PANEL B: 02-07-2007/14-11-2008												
ST	40.12	<u>43.88</u>	<i>39.29</i>	48.03	50.04	<u>49.75</u>	52.76	54.38	58.05	63.08	69.28	72.25
vND	0.023	0.026	0.025	0.030	<u>0.032</u>	0.033	0.043	0.047	0.044	0.045	0.054	0.054
QLIK	-338.75	<u>-338.00</u>	<i>-342.58</i>	-337.35	-332.04	-335.33	-328.48	<u>-327.86</u>	-324.19	-318.29	-312.09	-309.12
Portf. vol	-7.57	-7.43	<i>-7.69</i>	-7.70	-7.41	-7.34	-7.61	<u>-7.37</u>	-6.88	-7.16	-6.72	-5.88

Table 4.11: **GW test results across series**

Unconditional Giacomini-White (GW) test results at 5% level of significance using the QLIK function. Each panel of the table records the number of series for which the MMReDCC is favored, for which there is no decision and for which one of the benchmarks is favored. Results are differentiated between conditional co-volatilities (1081 series) and correlations (1035 series).

Co-volatilities					Correlations				
PANEL A: 29-11-2004/01-07-2007									
Horizon	Favors MMReDCC	Indecisive	Favors B	Model B	Favors MMReDCC	Indecisive	Favors B	Model B	
1	155	732	194	EWMA	62	669	304	EWMA	
	24	447	610	cReDCC	87	522	426	cReDCC	
5	121	833	127	EWMA	89	882	64	EWMA	
	20	369	692	cReDCC	114	427	494	cReDCC	
10	33	948	100	EWMA	72	906	57	EWMA	
	37	356	688	cReDCC	213	503	319	cReDCC	
20	51	850	180	EWMA	118	877	40	EWMA	
	77	512	492	cReDCC	108	605	322	cReDCC	
PANEL B: 02-07-2007/14-11-2008									
Horizon	Favors MMReDCC	Indecisive	Favors B	Model B	Favors MMReDCC	Indecisive	Favors B	Model B	
1	41	853	187	EWMA	130	689	133	EWMA	
	5	187	889	cReDCC	266	627	142	cReDCC	
5	106	910	65	EWMA	78	911	46	EWMA	
	429	620	32	cReDCC	502	505	28	cReDCC	
10	198	883	0	EWMA	217	791	27	EWMA	
	739	340	2	cReDCC	567	454	14	cReDCC	
20	193	888	0	EWMA	339	687	9	EWMA	
	835	246	0	cReDCC	626	405	4	cReDCC	

4.7 Conclusions

The estimation procedure proposed in the paper allows to extend the range of applicability of the MMReDCC model to large dimensional portfolios such as those encountered in risk management practice. In order to reach this objective, we face two well-known challenges in multivariate time series modeling, namely high-dimensional estimation and multi-step ahead forecasting.

To face the former challenge, we implement a feasible estimation procedure, the Iterative Moment based Profiling (IMP) algorithm. It profiles out the parameters of the scale MIDAS intercept matrix and iteratively maximizes the likelihood in terms of the other parameters of interest. Whilst not providing an asymptotic inference theory for this method, we investigate the finite sample properties of the estimator via a simulation study, which demonstrates that the IMP estimator delivers accurate estimates irrespective of the initialization method employed. We also compare the one-step QML and IMP estimators on real data sets of small dimension (ten) and find that not only the two estimators deliver very similar in-sample estimates, but also that the loss of the IMP in terms of likelihood values can be considered as negligible.

Another application illustrates the usefulness of the IMP algorithm when the model has to be fitted to high dimensional realized covariance matrices. In this respect, the IMP algorithm is reliable from the computational point of view and easy to implement despite the large number of parameters involved in the MMReDCC model. Given its flexibility, we fairly believe that it could be applied to datasets of even larger dimensions.

As regards the second challenge, we develop a bootstrap approach to the generation of multi-step-ahead predictions. In an application to a portfolio of forty-six stocks, we provide compelling evidence that the MMReDCC model is useful for out-of-sample forecasting purposes in periods of pronounced market volatility. If compared with existing multivariate competitors not accounting for time-varying long term dynamics, the MMReDCC is found to deliver the most accurate predictions especially at long term horizons, thus indicating the importance of allowing for a more flexible long run component.

Appendix

4.A Tables

Table 4.12: Descriptive statistics of daily realized variances used in Application I
The table reports descriptive statistics of daily realized variances of the assets included in Dataset 1, 2 and 3 used in the first empirical application. Realized variances of the assets comprised in Dataset 2 multiplied by 10000.

Symbol	Issue name	Mean	Max.	Min.	Std.dev.	Skewness	Kurtosis
Dataset 1: February, 2001 – December, 2009							
AA	Alcoa	5.458	277.308	0.074	16.811	7.178	72.570
AXP	American Express	5.055	176.478	0.112	11.094	7.529	84.686
BAC	Bank of America	1.934	57.543	0.075	3.362	7.319	85.006
KO	Coca Cola	2.455	43.106	0.084	3.412	4.724	36.234
DD	Du Pont	2.073	115.378	0.126	4.155	13.296	288.066
GE	General Electric	4.944	160.241	0.294	8.935	7.635	92.124
IBM	International Business Machines	4.420	201.879	0.077	9.154	8.536	133.699
JPM	JP Morgan	2.529	63.874	0.163	3.728	6.442	68.505
MSFT	Microsoft	3.196	114.256	0.097	7.114	7.232	75.484
XOM	Exxon Mobil	1.414	56.505	0.039	2.254	9.715	180.206
Dataset 2: May, 1997 – July, 2008							
ABT	Abbott Laboratories	3.831	249.408	0.197	8.588	17.264	397.922
T	AT& T	5.490	458.044	0.166	14.953	17.145	412.001
FISV	Fiserv	7.828	628.583	0.270	15.588	26.114	999.133
ALL	Allstate Corporation	4.423	980.762	0.106	20.731	41.512	1934.757
GPC	Genuine Parts Company	3.765	887.997	0.171	18.221	45.217	2180.553
AFL	Aflac Incorporated	4.443	267.045	0.153	10.955	14.110	268.010
AA	Alcoa	5.417	123.771	0.354	7.422	7.331	86.153
GE	General Electric	3.973	366.830	0.131	11.481	20.633	559.319
CTL	Century Link	4.472	687.135	0.217	18.814	24.783	775.201
C	Cytigroup Inc.	6.244	861.435	0.157	21.306	27.458	1036.670
Dataset 3: January, 2007 – December, 2012							
ACAS	American Capital	8.576	331.786	0.060	20.844	7.226	78.667
AET	Aetna	8.163	771.525	0.109	26.593	17.882	467.969
AFL	Aflac Incorporated	9.113	675.348	0.133	27.345	13.811	284.791
AIG	American International Group	8.799	555.098	0.103	26.382	11.459	185.778
AIZ	Assurant	8.613	325.167	0.101	23.230	7.712	79.082
ALL	The Allstate Corporation	8.213	543.714	0.186	24.593	11.277	189.052
AMP	Ameriprise Financial	7.679	264.761	0.129	17.790	6.098	54.262
AXP	American Express Company	8.076	945.750	0.095	30.571	21.795	618.891
BAC	Bank of America	8.450	332.586	0.130	22.830	8.458	96.824
BBT	BB&T Corporation	9.093	613.826	0.087	28.801	11.267	184.837

CHAPTER 4. FEASIBLE INFERENCE PROCEDURES FOR FORECASTING HIGH-DIMENSIONAL REALIZED COVARIANCE MATRICES

112

Table 4.13: Descriptive statistics of daily realized variances used in Application II

The table reports descriptive statistics of daily realized variances of the assets used in the second empirical application, where the forecasting sample, spanning from November 29, 2004 to November 14, 2008, is divided into a calm period (650 observations) and a more turbulent period (350 observations) including the 2008 financial crisis. Realized volatilities multiplied by 10000.

CALM PERIOD: NOVEMBER 29, 2004 TO JUNE 28, 2007 (650 OBSERVATIONS)								TURBULENT PERIOD: JUNE 29, 2007 TO NOVEMBER 14, 2008 (350 OBSERVATIONS)					
Stock	Issue name	Mean	Max.	Min.	Std.dev.	Skewness	Kurtosis	Mean	Max.	Min.	Std.dev.	Skewness	Kurtosis
AAPL	Apple	3.114	44.664	0.257	2.997	6.109	68.647	8.579	109.719	0.609	12.570	4.249	26.370
ABT	Abbott	1.040	11.301	0.151	0.818	4.773	45.820	3.225	81.740	0.258	5.702	8.409	106.373
AXP	American Express	0.876	10.932	0.072	1.038	5.715	44.530	12.572	312.208	0.838	22.702	7.818	92.250
BA	The Boeing Comp.	1.147	5.444	0.254	0.767	2.388	10.830	5.151	57.291	0.532	7.796	3.231	15.171
BAC	Bank of America	0.701	13.063	0.117	0.764	8.737	120.508	14.876	276.824	0.279	26.233	4.942	38.722
BMJ	Bristol-Myers Squibb Comp.	1.270	20.256	0.179	1.286	7.526	90.830	4.259	87.928	0.242	6.620	6.861	76.824
BP	BP b.l.c.	0.751	3.873	0.139	0.448	2.482	12.914	3.906	65.784	0.219	6.626	4.212	28.869
C	Citigroup Inc.	0.846	12.850	0.123	0.963	7.033	72.020	21.287	1015.764	0.519	64.196	11.716	171.168
CAT	Caterpillar	1.605	11.470	0.303	1.233	3.277	19.679	6.805	118.706	0.566	11.322	4.502	33.594
CL	Colgate-Palmolive	0.800	11.800	0.121	0.688	8.066	112.163	2.628	89.214	0.243	5.673	10.719	157.531
CSCO	Cisco Systems	1.552	11.784	0.189	1.014	3.324	24.630	6.178	114.330	0.478	9.424	5.839	55.640
CVX	Chevron Corp.	1.679	14.566	0.289	1.349	3.580	24.108	6.200	166.296	0.149	12.609	7.428	82.139
DELL	Dell	1.756	14.185	0.242	1.298	2.927	19.247	7.323	143.320	0.655	12.045	6.050	55.752
DIS	Walt Disney	1.014	7.920	0.201	0.712	3.932	30.039	5.089	118.243	0.453	9.420	6.197	63.696
EK	Eastman Kodak	1.416	22.304	0.179	1.642	6.306	60.391	6.751	134.678	0.566	11.362	5.373	50.263
EXC	Exelon	2.841	24.425	0.401	2.296	3.978	28.797	27.900	1981.687	0.937	116.942	13.976	226.720
F	Ford Motor	1.307	9.113	0.083	0.947	2.928	16.728	5.261	71.372	0.235	7.107	4.048	27.759
FDX	FedEX Corp.	0.663	3.940	0.126	0.447	3.009	16.528	6.789	97.782	0.292	13.465	4.100	23.033
GE	General Electric	1.305	6.982	0.181	0.858	2.407	11.664	7.993	160.418	0.495	11.651	7.441	87.734
GM	General Motors	0.666	6.577	0.105	0.589	4.342	32.564	2.303	39.958	0.208	3.666	4.777	38.049
HD	The Home Depot	1.306	10.663	0.173	1.003	4.422	34.275	5.686	139.350	0.541	10.774	6.729	72.384
HNZ	HNZ Group	0.796	4.691	0.128	0.549	2.948	15.566	4.494	85.792	0.490	7.507	5.156	44.724
HON	Honeywell	1.541	7.739	0.332	0.840	1.937	9.764	5.841	76.334	0.478	7.396	4.103	29.515
IBM	Intern. Business Machines	0.495	2.679	0.063	0.330	2.193	10.384	2.034	59.372	0.086	4.539	7.056	77.545
INTC	Intel Corp.	0.955	13.492	0.140	0.939	5.838	59.503	14.577	234.008	0.332	25.147	4.980	34.276
JNJ	Johnson & Johnson	0.577	4.866	0.076	0.366	3.806	34.495	2.874	101.443	0.189	6.550	10.331	148.976
JPM	JP Morgan	0.962	9.037	0.142	0.654	4.891	49.285	3.471	89.880	0.265	6.457	7.790	94.853
KO	Coca Cola	1.143	8.119	0.197	0.816	2.961	17.698	3.730	148.108	0.273	8.824	12.878	207.094
LLY	Eli Lilly and Co.	0.884	8.232	0.099	0.702	5.275	47.421	3.849	60.731	0.286	6.388	4.156	26.768
MCD	McDonald	1.456	39.785	0.139	2.373	10.527	141.696	5.405	153.458	0.532	10.592	8.776	113.275
MMM	3M Company	1.769	22.744	0.346	2.065	6.197	53.214	42.394	1836.085	1.213	156.907	8.427	82.761
MOT	Motorola	0.892	4.629	0.179	0.556	2.350	11.471	4.559	46.751	0.385	6.309	3.449	17.267
MRK	Merck & Co.	1.816	14.715	0.366	1.130	3.854	35.605	5.723	61.006	0.665	8.031	3.857	20.906
MS	Morgan Stanley	0.608	4.486	0.118	0.408	3.538	24.998	2.905	138.670	0.168	8.233	13.189	213.668
MSFT	Microsoft	1.094	19.804	0.146	1.058	9.820	157.073	3.556	63.957	0.305	5.699	4.909	41.144
SLB	Schlumberger Limited	0.679	6.771	0.079	0.530	4.601	38.636	2.770	123.527	0.196	7.582	12.047	186.371
T	AT&T	2.710	22.356	0.423	1.834	3.716	30.143	7.569	196.766	0.648	13.960	8.415	102.982
TWX	Time Warner	3.020	13.708	0.206	1.905	1.995	8.992	10.437	166.001	0.494	15.246	4.652	36.751
UN	Unilever	1.024	9.537	0.117	0.824	4.180	31.694	5.943	156.416	0.276	11.229	7.837	95.552
VZ	Verizon Communications	1.184	48.463	0.241	2.020	20.043	463.495	6.341	114.383	0.316	10.664	4.681	36.406
ORCL	Oracle Corp.	0.509	4.318	0.092	0.361	3.483	25.912	2.333	41.487	0.215	3.697	4.803	40.510
PEP	Pepsico	0.954	8.887	0.172	0.768	4.086	30.276	5.142	133.199	0.147	9.425	8.057	100.502
PFE	Pfizer Inc.	0.693	8.324	0.075	0.695	5.557	47.417	13.576	209.392	0.572	20.115	4.414	32.475
PG	Procter & Gamble	0.914	7.693	0.131	0.632	4.083	34.957	3.700	92.274	0.234	6.315	8.804	114.371
QCOM	Qualcomm	1.414	11.781	0.246	1.121	4.085	29.777	5.506	188.347	0.331	12.581	9.788	132.406
WFC	Well Fargo & Co.	1.790	44.518	0.243	2.202	12.661	226.808	6.288	182.259	0.653	12.009	9.676	134.541

4.B Bootstrap procedure for the cReDCC model

The conditional bootstrap procedure presented in Section 4.5.1 is adapted in order to generate multi-step forecasts from the consistent ReDCC (cReDCC) model of Bauwens et al. (2012). Here we summarize the main steps of the algorithm, similarly to the one described in the content of the thesis.

Let C_t be distributed as $C_t | \mathfrak{S}_{t-1} \sim W_n(\nu, S_t/\nu)$ and let S_t be defined as $S_t = D_t R_t D_t$.

The diagonal elements of S_t , for $i = 1, \dots, n$, are assumed to follow the dynamic equation

$$S_{t,ii} = (1 - \gamma_i - \delta_i) \bar{S}_i + \gamma_i C_{t-1,ii} + \delta_i S_{t-1,ii}$$

while the conditional correlation matrix R_t is assumed to be generated by the following equations

$$\begin{aligned} Q_t &= (1 - \alpha - \beta) \bar{Q} + \alpha P_{t-1} + \beta Q_{t-1} \\ P_t &= \{diag(Q_t)\}^{1/2} D_t^{-1} C_t D_t^{-1} \{diag(Q_t)\}^{1/2} \\ R_t &= \{diag(Q_t)\}^{-1/2} Q_t \{diag(Q_t)\}^{-1/2} \end{aligned}$$

The algorithm

Consider to re-write C_t as

$$C_t = S_t^{1/2} U_t (S_t^{1/2})',$$

where the U_t are a sequence of iid random variables such that $\mathbb{E}(U_t) = I_n$ and $S_t^{1/2}$ is any PD matrix such that $S_t^{1/2} (S_t^{1/2})' = S_t$. Then proceed as follows. Estimate the model on $\{C_t, t = 1, \dots, T\}$, obtain the estimated conditional covariances $\hat{S}_t$ and

1. Generate the estimated standardized residuals

$$\hat{U}_t = \hat{S}_t^{-1/2} C_t (\hat{S}_t^{-1/2})', \quad t = 1, \dots, T.$$

Rescale the $\hat{U}_t$ to have sample mean equal to I_n and used them to generate bootstrap replicates of C_t , as described in Steps 2-4.

2. Let h be the chosen forecast horizon and generate with replacement a bootstrap sample of length h from the empirical CDF of $\hat{U}_t$. We will label these observations $\tilde{U}_{T+j}$ ($j = 1, \dots, h$).
3. Generate S_{T+j} by the following steps. Compute $S_{T+j,ii}$

$$S_{T+j,ii} = (1 - \hat{\gamma}_i - \hat{\delta}_i) + \hat{\gamma}_i C_{T+j-1,ii} + \hat{\delta}_i S_{T+j-1,ii}$$

and D_{T+j} as $D_{T+j} = \{diag(S_{T+j,ii})^{1/2}\}$.

Then, obtain R_{T+j} as

$$R_{T+j} = \{\text{diag}(Q_t)\}^{-1/2} Q_t \{\text{diag}(Q_t)\}^{-1/2}$$

where

$$\begin{aligned} Q_{T+j} &= (1 - \hat{\alpha} - \hat{\beta})\bar{Q} + \hat{\alpha}P_{T+j-1} + \hat{\beta}Q_{T+j-1} \\ \bar{Q} &= \frac{1}{T} \sum_{t=1}^T P_t \\ P_{T+j} &= \{\text{diag}(Q_{T+j})\}^{1/2} D_{T+j}^{-1} C_{T+j} D_{T+j}^{-1} \{\text{diag}(Q_{T+j})\}^{1/2} \end{aligned}$$

and

$$\begin{aligned} C_{T+j} &= (S_{T+j})^{1/2} \tilde{U}_{T+j} ((S_{T+j})^{1/2})' \\ S_{T+j} &= (D_{T+j}) R_{T+j} (D_{T+j}) \end{aligned}$$

The procedure is initialized setting $R_T = \hat{R}_T$, $C_T = \hat{C}_T$, $S_T = \hat{S}_T$, $P_T = \hat{P}_T$.

4. Iterate steps 3 – 4 for $j = 1, \dots, h$.
5. Repeat steps 2-4 for $b = 1, \dots, B$ times.

As a result the algorithm generates a $(T + h) \times B$ matrix of bootstrap samples. The forecast at time $T + h$ can be computed as the mean, over all the bootstrap samples, of values referring to time $T + h$. So, if we denote with $C_t^{(b)}$ the b-th bootstrap sample ($b = 1, \dots, B$), the h-step ahead forecast of S_t is computed as

$$\hat{S}_{T,h} = \frac{1}{B} \sum_{b=1}^B C_{T+h}^{(b)}.$$

■

Conclusions

The present thesis tries to contribute to the existing literature dealing with the modeling and forecasting of economic variables of interest, such as variances, covariances and correlations. Our work heavily builds on the most recent findings in financial econometrics, highlighting the key role played by the so-called *realized measures* in developing more accurate models. Even in very basic frameworks, it is known that they lead to better in-sample fit and to undoubtedly more powerful predictions than models relying on daily data alone. The results of the empirical applications presented in Chapter 2 also point towards these conclusions.

Overall, despite maintaining an homogeneous path throughout the chapters, we face different methodological and computational issues that in turn require to apply a specific approach. This has lead us to investigate several related topics and to deliver each time the most appropriate answer, among those possible, to the case at hand.

To summarize, an initial part of the thesis embrace the common belief of constant long run co-volatility, which turns out to be a convenient feature to develop simple and intuitive model specifications that allow estimating large-dimensional systems with several tens of assets. In order to keep estimation tractable we propose a multistep estimation approach and demonstrate, through simulations, that splitting the likelihood function into steps leads to accurate parameter estimates overall, even when the time series length is limited. Furthermore, this approach is discovered to be reasonable even in cases where the standard one-step estimator is feasible: the multistep estimator we use is less exposed to parameter proliferation and thus less distorted as the cross-sectional dimension increases. To our knowledge, this finding is new to the literature on realized covariance modeling and contributes to make these models as valuable benchmarks for more sophisticated specifications currently used in financial applications.

In the remaining part of the thesis we relax the assumption of constant mean reversion and investigate the usefulness of models allowing for time-varying long run dynamics, which entail more flexible but also more involved parameterizations. The purpose is to improve on the forecasting accuracy in periods of highly unstable market conditions, where the presence of the dynamic long run component is expected to quickly adapt to the changing economic fundamentals. Our detailed out-of-sample analysis performed across Chapters 3 and 4 indeed show the superior effectiveness of these models in producing accurate forecasts, at both one- and multi-step-ahead horizons.

Differently from the constant mean reverting models, those introduced from Chapter 3 onward are more computationally intense and thus, more challenging to use in realistic high dimensional frameworks. The techniques we illustrate in the content of the last chapter in order to overcome these problems have the goal to optimize the trade-off between accuracy and parsimony, yet being readily applicable by practitioners.

The main message that this thesis can provide is very simple. Building a multivariate time series model is a demanding task and features like accuracy, parsimony, efficiency and computational complexity have to be carefully considered. As no perfect model exists for any situation, it is the model designer's capability to find the combination that provides the best solution to the desired goal.

To conclude, we find that our research lends itself to extensions along several directions. Two, in particular, seem straightforwardly applicable.

Firstly, it would be interesting to investigate the existence and nature of the relationship between market volatility and macroeconomic fundamentals, which is crucial in understanding issues relevant to policy makers and institutional investors. This can be achieved by incorporating influential macroeconomic (such as monetary policy news) or economic sentiment variables directly into the specification of the long-term dynamics of the process. These variables are usually available at a monthly frequency and could represent important drivers of the assets long-term co-volatilities or correlations. In particular, over longer horizons (≥ 1 month), the inclusion of low-frequency variables is possibly expected to favorably impact on the overall fit of the model and to further improve its forecasting accuracy. Engle et al. (2013) addressed this topic in the univariate framework but no attempt has been made so far to extend the results to a reasonably large multivariate setting. The main challenge in this respect is represented by the positive definiteness constraint of the high dimensional covariance or correlation matrix. The introduction of exogenous variables do not necessarily guarantees the positive definiteness of the matrix, which has to be regularized using appropriate statistical tools, such as the Fisher-z transformation, shrinkage or eigenvalue cleaning, to name a few. If a novel - easy to apply - approach to handle this issue in a large multivariate setting could be found, then this work would definitely represent a breakthrough in the existing multivariate mixed-frequencies literature.

The second appealing research question deals with the choice of the conditional distribution for the realized covariance matrix. As the literature shows, a substantial body of the existing models make use of the Wishart density function, which can be considered as a multivariate generalization of the χ^2 distribution and possesses a lot of convenient properties. All the models we proposed also rely on the assumption that the series of realized covariance matrices follow a central Wishart distribution with time varying conditional expectation. More recently, some attention has also been paid to other distributions that, similarly to the Wishart, ensure positive-definiteness of the covariance matrix. For example, Jin & Maheu (2013) extend some existing dynamic Wishart models to mixture models of Wishart and inverse-Wishart distributions, used to produce both point and density forecasts. On the other hand, Janus et al. (2014)

develop a new dynamic model for the joint evolution of daily return covariance matrices and realized covariance matrices allowing for both fat tails and outliers. They account for these features by assuming a matrix-Fisher distribution for the realized kernels and a multivariate Student-t distribution for the returns. In both cases the authors claim that the proposed models substantially improve over existing competitive models employing a conditional Wishart density.

We consider this as a very promising area of research. In particular, as there is not a clear documented reason justifying the choice of the Wishart distribution over other competitors, we believe it could be interesting to investigate the models in- and out-of-sample performance under the use of different underlying matrix-variate distributions as to assess if a specific one is found to systematically beat the others. Inspired by the recent works of [Aastveit et al. \(2014\)](#) and [Clark & Ravazzolo \(2015\)](#), the comparison would be performed on the basis of the accuracy of both real-time point and density forecasts, as complete probability distributions over outcomes provide helpful information for making more conscious economic decisions. In order to obtain a more general result, a first analysis could be conducted on simpler (and possibly more popular) model structures (i.e., the one underlying the CAW model) and later on eventually extended to the more involved set of models proposed in Chapter 3.

As we are not aware of other articles which deal with this specific research question, we expect this study to provide a more formal and hopefully useful explanation regarding the choice of one specific matrix-variate distribution.

Appendix A

The Wishart distribution

Intuitive understanding

The Wishart distribution is the multivariate extension of the gamma distribution, although most statisticians use the Wishart in the special case of integer degrees of freedom, in which case it simplifies to a multivariate generalization of the χ^2 distribution. As the χ^2 distribution describes the sums of squares of p draws from a univariate normal distribution, the Wishart distribution represents the sums of squares (and cross-products) of p draws from a multivariate normal distribution.

Mathematical understanding

PDF Let $C \sim W_n(\Sigma; \nu)$, where Σ denotes a positive definite scale matrix (which can be thought of as a covariance matrix from a multivariate normal distribution), ν is the parameter that denotes the degrees of freedom, and n indicates the dimensions of C (i.e., $C \in R^{n \times n}$). Then C is positive definite with probability density function (pdf)

$$f(S) = \frac{|C|^{\frac{\nu-n-1}{2}}}{2^{\frac{\nu n}{2}} |\Sigma|^{\frac{\nu}{2}} \Gamma_n(\frac{\nu}{2})} \exp \left[\frac{1}{2} \text{tr} (\Sigma^{-1} C) \right]$$

where $|A|$ represents the determinant of square matrix A , $\text{tr}(A)$ is the trace of square matrix A (i.e., the sum of the diagonal elements of A), and $\Gamma_n(\nu/2) = (\pi)^{n(n-1)/4} \prod_{i=1}^n \Gamma(\frac{\nu+1-i}{2})$, with $\Gamma(a) = \int_0^\infty x^{a-1} \exp(-x) dx$ for $a > 0$.

Note that we must have $\nu > n - 1$ to ensure that S is invertible. If $\nu > n - 1$ does not hold, then $W_n(\Sigma; \nu)$ is called a Singular Wishart distribution due to Σ being a singular matrix.

Individual Variates Similar to the χ^2 distribution, draws from a Wishart distribution represent sums of squares and not variances. Because the scale matrix Σ can be thought of as a population covariance matrix, individual draws from the Wishart distribution will often be several times the magnitude of the covariance matrix. Note that if $X \sim \chi^2(p)$, the $X = \sum_{i=1}^p Z_i^2$ where $Z_1, \dots, Z_p \sim i.i.d.N(0, 1)$, so the sums of squares extension to the Wishart distribution is necessary.

Sum of Individual Variates A Wishart distribution acts like a distribution for the sums of squares and cross-products. To see the relationship between a Wishart distribution and sums of squares, let $C_1, C_2, \dots, C_i, \dots, C_k$ be independent from k Wishart distributions where $C_i \sim W(\Sigma, \nu_i)$. Then

$$CC = \sum_{i=1}^k S_i \sim W\left(\Sigma, \sum_{i=1}^k \nu_i\right).$$

In other words, C_i can be thought of as the sums of squares matrix for a sub-sample of ν_i scores from a multivariate normal distribution. And $CC = \sum_{i=1}^k C_i$ is the sums of squares matrix across the k sub-samples. However, when the k sub-samples are independent (all with population covariance matrix Σ) then the sums of squares across the k sub-samples is equivalent to the sums of squares across the $N = \sum_{i=1}^k \nu_i$ total scores.

Expected value The expected value of C is

$$\mathbb{E}(C) = \nu \Sigma.$$

Also in parallel with the χ^2 distribution, the expected value of a Wishart distribution depends on number of draws one makes from the multivariate normal distribution. In comparison, the expected value of a $\chi^2(\nu)$ distribution is ν , so that the only differences between a Wishart expectation and a χ^2 expectation are the underlying dimensionality of the data and a scale component.

As for the other moments of order k of the elements of C , they all exist for $k > 0$.

Invariance and decomposition of C The Wishart has two important properties. First, its parameterization is invariant under marginalization, since a $n^* \times n^*$ submatrix C^* , $n^* < n$, follows a Wishart distribution $C^* \sim W_{n^*}(\nu, \Sigma^*)$. Note that the degrees of freedom remain unchanged.

Second, it is invariant under transformation $C \rightarrow ACA'$. Namely, if $C \sim W_p(\Sigma, \nu)$ and A is any $n \times n$ nonsingular matrix, then $ACA' \sim W_n(A\Sigma A'; \nu)$.

Appendix B

Summary of TVLR component models

The following Table B.1 recalls names and main features of the time-varying long run component dynamic models illustrated in Chapter 3, so as to improve their readability and comprehension.

The general rules used to build the models' acronyms are as follows:

- the first letter of the name refers to the basic structure of the model, being either “A” for Additive or “M” for Multiplicative;
- the second letter refers to the filter used to model the long-run component, so it is either “M” if the finite distributed lag specification typical of the MIDAS scheme is used or “K” for the nonparametric kernel smoother (e.g. Nadaraya-Watson estimator);
- when the MIDAS-type filter is used, the following “A”, if present, stands for 'aggregated' and denotes the type of temporal scheme applied to the series of realized variances or covariances in the MIDAS polynomial;
- the following string refers to the realized MGARCH-type specification used to model the short term dynamics common to all models, being one among ReDCC, ReBEKK and ReDECO;
- finally, when present, the “-P” suffix that is added at the very end of the name identifies the use of the purged covariance estimator as described in the content of the thesis (Chapter 3, Section 3.2.1).

This suffix is only present for models belonging to the Additive class.

As for the models used as benchmarks in Chapter 3, names are simplified: the initial letter “C” denotes the constant secular component and the remaining part of the string, again ReDCC, ReBEKK or ReDECO, determines the model being used for the short term component.

There is no need to explain the acronym cReDCC ([Bauwens et al. \(2012\)](#)) as it already exists in the literature (see also Appendix C).

Table B.1: Complete list of TVLR dynamic component models

MODEL	CLASS				
	Additive			Multiplicative	
	Not Aggregated	Aggregated	Purged	Parametric	Semiparametric
1 - AMReDCC	✓				
2 - AMReDECO	✓				
3 - AMAReDCC		✓			
4 - AMAReDECO		✓			
5 - AMReDCC-P	✓		✓		
7 - AMAReDCC-P		✓	✓		
8 - AMAReDECO-P		✓	✓		
9 - MKReDCC					✓
10 - MKReDECO					✓
11 - MKReBEKK					✓
12 - MMReDCC				✓	
13 - MMReDECO				✓	
14 - MMReBEKK				✓	
15 - MMAReDCC				✓	
16 - MMAReDECO				✓	
17 - MMAReBEKK				✓	

Appendix C

Overview of competing models

In the thesis, several models are used as benchmarks in the empirical applications. Herein we briefly sketch their structure and, when necessary, provide some further comments to avoid confusion with their acronyms. The notation used in the following is the same as in the content of each chapter.

The Conditional Autoregressive Wishart (CAW) of Golosnoy et al. (2012)

The CAW model is used in Chapter 3 as benchmark for the Re-cDCC model. It is fitted to the series of realized covariance matrices by assuming an autoregressive moving average structure for the scale matrix of the Wishart distribution. This model accounts for the positive definiteness of the covariance matrix without imposing parametric restrictions and can be estimated by Maximum Likelihood. Assuming a full BEKK-type recursion, the model is sufficiently described by the following two equations:

$$\begin{aligned} C_t &\sim W_n(\nu, S_t/\nu) \\ S_t &= \bar{C}\bar{C}' + AC_{t-1}A' + BS_{t-1}B', \end{aligned}$$

where A, B are n -dimensional matrices of parameters and $\bar{C}$ is a $\{n(n+1)/2\}$ lower triangular matrix of parameters such that $\bar{C}\bar{C}'$ is the $n \times n$ mean reverting constant covariance matrix.

The scalar Realized DCC (ReDCC) of Bauwens et al. (2012) This model is employed in Chapter 3 and Chapter 4. It can be considered the precursor of the Realized consistent DCC (Re-cDCC) model introduced in Chapter 2, and has often been cited with the cRDCC name (see e.g. Boudt et al. (2014)). In this thesis we decided to adopt the ReDCC acronym (or cReDCC for its *consistent* version) in order to avoid confusion with the Rotated DCC (RDCC) model of Noureldin et al. (2012)

(see below). Its structure is given by the following equations:

$$\begin{aligned}
C_t &\sim W_n(\nu, S_t/\nu) \\
S_t &= D_t R_t D_t \\
D_t &= \text{diag}\{\sqrt{S_{ii,t}}\}, \quad i = 1, \dots, n \\
S_{ii,t} &= (1 - \gamma_i - \delta_i) \bar{S}_{ii} + \gamma_i C_{ii,t-1} + \delta_i S_{ii,t-1} \\
Q_t &= \bar{C} \bar{C}' + \alpha P_{t-1} + \beta Q_{t-1} \\
P_t &= \{\text{diag}(C_t)\}^{-1/2} C_t \{\text{diag}(C_t)\}^{-1/2} \\
R_t &= \{\text{diag}(Q_t)\}^{-1/2} Q_t \{\text{diag}(Q_t)\}^{-1/2}
\end{aligned}$$

where $\bar{S}_i$ is the unconditional realized variance of asset i and $\bar{C}$ is a $\{n(n+1)/2\}$ lower triangular matrix of parameters such that $\bar{C} \bar{C}'$ is a $n \times n$ constant matrix (not necessarily a correlation matrix). Regularization of Q_t through R_t is needed to obtain a well defined correlation matrix.

The rotated DCC (RDCC) of Noureldin et al. (2012) The Rotated DCC belongs to the class of multivariate rotated ARCH models introduced by Noureldin et al. (2012). The basic structure is to rotate the returns and then to fit them using a DCC-type parameterization of the time-varying covariance whose long-run covariance is the identity matrix. BEKK-type structures are also possible. The model is usually estimated by QML using covariance targeting, thus it can be used in relatively large dimensional systems. Letting $\mathbf{r}_t, t = 1, \dots, T$ denote the n -dimensional time series of daily asset returns with $\mathbb{E}(\mathbf{r}_t | \mathfrak{S}_{t-1}) = 0$, the basic structure of the model is given as follows:

$$\begin{aligned}
\text{Var}(\mathbf{r}_t) &= \bar{H} = P \Lambda P' \\
\mathbf{r}_t &= \bar{H}^{1/2} \mathbf{e}_t \\
\mathbf{e}_t &= \bar{H}^{-1/2} \mathbf{r}_t = P \Lambda^{-1/2} P' \mathbf{r}_t \\
\text{Var}(\mathbf{e}_t) &= I_n \\
\text{Var}(\mathbf{e}_t | \mathfrak{S}_{t-1}) &= G_t \\
\text{Var}(\mathbf{r}_t | \mathfrak{S}_{t-1}) &= H_t = \bar{H}^{1/2} G_t \bar{H}^{1/2}.
\end{aligned}$$

BEKK or DCC-type dynamics have to be specified for the matrix G_t directly. The structure of these models allows for a three-step (two-step in the BEKK case) estimation strategy to estimate the parameter vector of interest. In the first step, a method of moment estimator is used to obtain $\bar{H}$, while the second-step estimation is based on the maximization of a normal quasi-likelihood function for the time series of rotated returns $\mathbf{e}_t$. This model is conveniently compared to the Re-cDCC in the application of Chapter 2, as the models share many similarities in the implemented step-wise estimation

approach. Nevertheless, the crucial difference between the two relies in the information used to form predictions of the conditional covariance matrix: in the RDCC case only series of daily returns are used, while the Re-cDCC uses realized data computed from exploiting intra-daily return information.

The Constant ReDCC (CReDCC) The CReDCC model is employed as benchmark for the class of multiplicative TVLR component models in Chapter 3, as it implies the same multiplicative structure with a constant long run component in spite of the time-varying MIDAS filter. Extension to the CReDECO and CReBEKK specifications is just straightforward. The set of equations needed is as follows.

$$\begin{aligned}
C_t &\sim W_n(\nu, S_t/\nu) \\
S_t &= LD_t^* R_t^* D_t^{*'} L' \\
M &= LL' \\
C_t^* &= L^{-1} C_t (L')^{-1} \\
D_t^* &= \text{diag}\{(S_{ii,t}^*)^{1/2}\}, \quad i = 1, \dots, n \\
S_{ii,t}^* &= (1 - \gamma_i - \delta_i) + \gamma_i C_{ii,t-1}^* + \delta_i S_{ii,t-1}^* \\
R_t^* &= (1 - \alpha - \beta) I_n + \alpha P_{t-1}^* + \beta R_{t-1}^* \\
P_t^* &= \{\text{diag}(C_t^*)\}^{-1/2} C_t^* \{\text{diag}(C_t^*)\}^{-1/2}.
\end{aligned}$$

L is a $\{n(n+1)/2\}$ lower triangular matrix of parameters such that $M = LL'$ is a $n \times n$ constant covariance matrix. As in the MMReDCC model, it holds $\mathbb{E}(D_t^* R_t^* D_t^{*'}) = I_n$, with I_n the n -dimensional identity matrix. ■

References

- Aastveit, K. A., Foroni, C. & Ravazzolo, F. (2014), ‘Density forecasts with midas models’, *Norges Bank Working Paper 10/2014* .
- Aielli, G. P. (2006), Consistent estimation of large scale dynamic conditional correlations. University of Florence Working Paper.
- Aielli, G. P. (2013), ‘Dynamic conditional correlation: on properties and estimation’, *Journal of Business & Economic Statistics* **31**(3), 282–299.
- Aït-Sahalia, Y. & Hansen, L. P. (2009), *Handbook of Financial Econometrics, Vol 1: Tools and Techniques*, Vol. 1, Elsevier.
- Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2005), ‘How often to sample a continuous-time process in the presence of market microstructure noise’, *Review of Financial Studies* **18**(2), 351–416.
- Andersen, T., Bollerslev, T., Diebold, F. & Ebens, H. (2001), ‘The distribution of realized stock return volatility’, *Journal of Financial Economics* **61**(1), 43–76.
- Andersen, T., Bollerslev, T., Frederiksen, P. & Nielsen, M. (2010), ‘Continuous-time models, realized volatilities and testable distributional implications for daily stock returns’, *Journal of Applied Econometrics* **25**, 233–261.
- Andersen, T. G. & Bollerslev, T. (1998), ‘Answering the skeptics: Yes, standard volatility models do provide accurate forecasts’, *International economic review* pp. 885–905.
- Andersen, T. G., Bollerslev, T. & Diebold, F. X. (2007), ‘Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility’, *The review of economics and statistics* **89**(4), 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2001), ‘The distribution of realized exchange rate volatility’, *Journal of the American statistical association* **96**(453), 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), ‘Modeling and forecasting realized volatility’, *Econometrica* **71**(2), 579–625.
- Bandi, F. M. & Russell, J. R. (2008), ‘Microstructure noise, realized variance, and optimal sampling’, *The Review of Economic Studies* **75**(2), 339–369.

- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008), ‘Designing realized kernels to measure the ex post variation of equity prices in the presence of noise’, *Econometrica* **76**(6), 1481–1536.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2009), ‘Realized kernels in practice: Trades and quotes’, *The Econometrics Journal* **12**(3), C1–C32.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (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**(2), 149–169.
- Barndorff-Nielsen, O. E. & Shephard, N. (2002), ‘Estimating quadratic variation using realized variance’, *Journal of Applied econometrics* **17**(5), 457–477.
- Barndorff-Nielsen, O. E. & Shephard, N. (2004), ‘Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics’, *Econometrica* **72**(3), 885–925.
- Barndorff-Nielsen, O. & Shephard, N. (2001), ‘Normal modified stable processes’, *Theory of Probability and Mathematics Statistics* **65**, 1–19.
- Bauer, G. H. & Vorkink, K. (2011), ‘Forecasting multivariate realized stock market volatility’, *Journal of Econometrics* **160**(1), 93–101.
- Bauwens, L., Braione, M. & Storti, G. (2016a), ‘A dynamic component model for forecasting high-dimensional realized covariance matrices’, *Econometrics and Statistics* (just-accepted).
- Bauwens, L., Braione, M. & Storti, G. (2016b), ‘Forecasting comparison of long term component dynamic models for realized covariance matrices’, *Annals of Economics and Statistics* (just-accepted).
- Bauwens, L., Grigoryeva, L. & Ortega, J.-P. (2015), ‘Non-scalar GARCH models: Composite likelihood estimation and empirical comparisons’, *Working Paper*.
- Bauwens, L., Grigoryeva, L. & Ortega, J.-P. (2016), ‘Estimation and empirical performance of non-scalar dynamic conditional correlation models’, *Computational Statistics & Data Analysis* **100**, 17–36.
- Bauwens, L., Hafner, C. M. & Pierret, D. (2013), ‘Multivariate volatility modeling of electricity futures’, *Journal of Applied Econometrics* **28**(5), 743–761.
- Bauwens, L., Storti, G. & Violante, F. (2012), Dynamic conditional correlation models for realized covariance matrices. CORE Discussion Paper 2012/60.
- Bollerslev, T. (1986), ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of econometrics* **31**(3), 307–327.

- Bollerslev, T. (1990), ‘Modeling the coherence in short-run nominal exchange rates: A multivariate generalized arch model’, *Review of Economics and Statistics* **72**, 498–505.
- Bollerslev, T., Patton, A. J. & Quaadvlieg, R. (2016a), ‘Exploiting the errors: A simple approach for improved volatility forecasting’, *Journal of Econometrics* **192**(1), 1–18.
- Bollerslev, T., Patton, A. J. & Quaadvlieg, R. (2016b), ‘Modeling and forecasting (un) reliable realized covariances for more reliable financial decisions’, *Available at SSRN 2759388*.
- Bonato, M., Caporin, M. & Ranaldo, A. (2009), Forecasting realized (co)variances with a block structure Wishart autoregressive model. Working Papers 2009-03, Swiss National Bank.
- Bonato, M., Caporin, M. & Ranaldo, A. (2012), ‘A forecast-based comparison of restricted wishart autoregressive models for realized covariance matrices’, *The European Journal of Finance* **18**(9), 761–774.
- Boswijk, H. P. (1995), ‘Identifiability of cointegrated systems’, *Tinbergen Institute Working Paper* (95/78).
- Boudt, K., Danielsson, J. & Laurent, S. (2013), ‘Robust forecasting of dynamic conditional correlation GARCH models’, *International Journal of Forecasting* **29**(2), 244–257.
- Boudt, K., Laurent, S., Lunde, A., Quaadvlieg, R. & Orimar, S. (2014), ‘Positive semidefinite integrated covariance estimation, factorizations and asynchronicity’. CREATES Research Paper 2014/5.
- Chiriac, R. & Voev, V. (2011), ‘Modelling and forecasting multivariate realized volatility’, *Journal of Applied Econometrics* **26**(6), 922–947.
- Christensen, K., Kinnebrock, S. & Podolskij, M. (2010), ‘Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data’, *Journal of Econometrics* **159**(1), 116–133.
- Christensen, K. & Podolskij, M. (2007), ‘Realized range-based estimation of integrated variance’, *Journal of Econometrics* **141**(2), 323–349.
- Christoffersen, P. F. (1998), ‘Evaluating interval forecasts’, *International Economic Review* **39**(4), 841–862.
- Clark, T. E. & Ravazzolo, F. (2015), ‘Macroeconomic forecasting performance under alternative specifications of time-varying volatility’, *Journal of Applied Econometrics* **30**(4), 551–575.
- Clements, M. P. & Smith, J. (1997), ‘The performance of alternative forecasting methods for SETAR models’, *International Journal of Forecasting* **13**(4), 463–475.

- Colacito, R. & Engle, R. (2006), 'Testing and valuing dynamic correlations for asset allocation', *Journal of Business & Economic Statistics* **24**(2), 238–253.
- Colacito, R., Engle, R. F. & Ghysels, E. (2011), 'A component model for dynamic correlations', *Journal of Econometrics* **164**(1), 45–59.
- Corsi, F. (2009), 'A simple approximate long-memory model of realized volatility', *Journal of Financial Econometrics* **7**(2), 174–196.
- Cubadda, G., Guardabascio, B., Hecq, A. et al. (2015), A vector heterogeneous autoregressive index model for realized volatility measures, Technical report.
- de Pooter, M., Martens, M. & van Dijk, D. (2008), 'Predicting the daily covariance matrix for s&p 100 stocks using intraday data, but which frequency to use?', *Econometric Reviews* **27**(1-3), 199–229.
- Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business & Economic Statistics* **13**(3), 253–263.
- Engle, R. (2002a), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business & Economic Statistics* **20**(3), 339–350.
- Engle, R. (2002b), 'New frontiers for arch models', *Journal of Applied Econometrics* **17**(5), 425–446.
- Engle, R. F. (2009), 'High dimension dynamic correlations', *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry: A Festschrift in Honour of David F. Hendry* p. 122.
- Engle, R. F. & Gallo, G. M. (2006), 'A multiple indicators model for volatility using intra-daily data', *Journal of Econometrics* **131**(1), 3–27.
- Engle, R. F., Ghysels, E. & Sohn, B. (2013), 'Stock market volatility and macroeconomic fundamentals', *Review of Economics and Statistics* **95**(3), 776–797.
- Engle, R. F. & Russell, J. (1998), 'Autoregressive conditional duration; a new model for irregularly spaced transaction data', *Econometrica* **66**, 1127–1162.
- Engle, R. F., Shephard, N. & Sheppard, K. (2008), 'Fitting vast dimensional time-varying covariance models'.
- Engle, R. & Kelly, B. (2012), 'Dynamic equicorrelation', *Journal of Business and Economic Statistics* **30**(2), 212–228.
- Engle, R. & Kroner, F. (1995), 'Multivariate simultaneous generalized ARCH', *Econometric Theory* **11**(1), 122–150.

- Engle, R. & Lee, G. (1999), *A Permanent and Transitory Component Model of Stock Return Volatility*, Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger, Oxford University Press - R. Engle and H. White eds., pp. 475–497.
- Engle, R. & Manganelli, S. (2004), ‘Caviar: Conditional autoregressive value at risk by regression quantiles’, *Journal of Business & Economic Statistics* **22**(4), 367–381.
- Engle, R. & Rangel, J. (2008), ‘The spline-GARCH model for low-frequency volatility and its global macroeconomic causes’, *Review of Financial Studies* **21**(3), 1187–1222.
- Fan, J., Li, Y. & Yu, K. (2012), ‘Vast volatility matrix estimation using high-frequency data for portfolio selection’, *Journal of the American Statistical Association* **107**(497), 412–428.
- Finger, C. (2005), ‘Back to backtesting’, *RiskMetrics Monthly Research*.
- Francq, C., Horvath, L. & Zakoïan, J.-M. (2011), ‘Merits and drawbacks of variance targeting in GARCH models’, *Journal of Financial Econometrics* **4**, 619–656.
- Francq, C., Horváth, L. & Zakoïan, J.-M. (2014), ‘Variance targeting estimation of multivariate GARCH models’, *Journal of Financial Econometrics* **14**(2), 353–382.
- Fresoli, D. E. & Ruiz, E. (2016), ‘The uncertainty of conditional returns, volatilities and correlations in dcc models’, *Computational Statistics & Data Analysis* (just-accepted).
- Fresoli, D., Ruiz, E. & Pascual, L. (2015), ‘Bootstrap multi-step forecasts of non-gaussian var models’, *International Journal of Forecasting* **31**(3), 834–848.
- Ghysels, E., Santa-Clara, P. & Valkanov, R. (2005), ‘There is a risk-return trade-off after all’, *Journal of Financial Economics* **76**(3), 509–548.
- Ghysels, E., Sinko, A. & Valkanov, R. (2007), ‘Midas regressions: Further results and new directions’, *Econometric Reviews* **26**(1), 53–90.
- Giacomini, R. & White, H. (2006), ‘Tests of conditional predictive ability’, *Econometrica* **74**(6), 1545–1578.
- Golosnoy, V., Gribisch, B. & Liesenfeld, R. (2012), ‘The conditional autoregressive wishart model for multivariate stock market volatility’, *Journal of Econometrics* **167**(1), 211–223.
- Gouriéroux, C., Jasiak, J. & Sufana, R. (2009), ‘The wishart autoregressive process of multivariate stochastic volatility’, *Journal of Econometrics* **150**(2), 167–181.
- Hafner, C. & Linton, B. (2010), ‘Efficient estimation of a multivariate multiplicative volatility model’, *Journal of Econometrics* **159**(1), 55–73.

- Hansen, P. R. & Lunde, A. (2006), ‘Realized variance and market microstructure noise’, *Journal of Business & Economic Statistics* **24**(2), 127–161.
- Hansen, P. R., Lunde, A. & Nason, J. M. (2011), ‘The model confidence set’, *Econometrica* **79**(2), 453–497.
- Härdle, W. (2004), *Nonparametric and semiparametric models*, Springer.
- Hautsch, N., Kyj, L. M. & Malec, P. (2015), ‘Do high-frequency data improve high-dimensional portfolio allocations?’, *Journal of Applied Econometrics* **30**(2), 263–290.
- Hautsch, N., Kyj, L. M. & Oomen, R. C. (2012), ‘A blocking and regularization approach to high-dimensional realized covariance estimation’, *Journal of Applied Econometrics* **27**(4), 625–645.
- Hayashi, T. & Yoshida, N. (2008), ‘Asymptotic normality of a covariance estimator for nonsynchronously observed diffusion processes’, *Annals of the Institute of Statistical Mathematics* **60**(2), 367–406.
- Jacod, J. & Protter, P. (1998), ‘Asymptotic error distributions for the euler method for stochastic differential equations’, *Annals of Probability* pp. 267–307.
- Janus, P., Lucas, A. & Opschoor, A. (2014), ‘New heavy models for fat-tailed returns and realized covariance kernels’.
- Jin, X. & Maheu, J. M. (2013), ‘Modeling realized covariances and returns’, *Journal of Financial Econometrics* **11**(2), 335–369.
- Laurent, S., Rombouts, J. V. & Violante, F. (2012), ‘On the forecasting accuracy of multivariate GARCH models’, *Journal of Applied Econometrics* **27**(6), 934–955.
- Laurent, S., Rombouts, J. V. & Violante, F. (2013), ‘On loss functions and ranking forecasting performances of multivariate volatility models’, *Journal of Econometrics* **173**(1), 1–10.
- Ledoit, O., Santa-Clara, P. & Wolf, M. (2003), ‘Flexible multivariate GARCH modeling with an application to international stock markets’, *The Review of Economics and Statistics* **85**(3), 735–747.
- Lindsay, B. G. (1988), ‘Composite likelihood methods’, *Contemporary mathematics* **80**(1), 221–39.
- Lunde, A., Shephard, N. & Sheppard, K. (2016), ‘Econometrics analysis of vast covariance matrices using composite realized kernels and their application to portfolio choice’, *Journal of Business and Economic Statistics* **34**(4), 504–518.
- Lutkepohl, H. (1996), *Handbook of matrices*, Wiley.
- Martens, M. & Van Dijk, D. (2007), ‘Measuring volatility with the realized range’, *Journal of Econometrics* **138**(1), 181–207.

- Morgan, J. (1994), ‘Introduction to riskmetrics’, *New York: JP Morgan*.
- Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V. & von Weizsäcker, J. E. (1997), ‘Volatilities of different time resolutions, Analyzing the dynamics of market components’, *Journal of Empirical Finance* **4**(2), 213–239.
- Mykland, P. A. & Zhang, L. (2009), ‘Inference for continuous semimartingales observed at high frequency’, *Econometrica* **77**(5), 1403–1445.
- Newey, W. K. & West, K. D. (1987), ‘A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix’, *Econometrica* **55**(3), 703–708.
- Noureldin, D., Shephard, N. & Sheppard, K. (2012), ‘Multivariate high-frequency-based volatility (HEAVY) models’, *Journal of Applied Econometrics* **27**(6), 907–933.
- Noureldin, D., Shephard, N. & Sheppard, K. (2014), ‘Multivariate rotated ARCH models’, *Journal of Econometrics* **179**(1), 16–30.
- Pakel, C., Shephard, N., Sheppard, K. & Engle, R. F. (2014), Fitting vast dimensional time-varying covariance models, Technical report, Working Paper.
- Pascual, L., Romo, J. & Ruiz, E. (2006), ‘Bootstrap prediction for returns and volatilities in GARCH models’, *Computational Statistics & Data Analysis* **50**(9), 2293–2312.
- Patton, A. J. (2011), ‘Volatility forecast comparison using imperfect volatility proxies’, *Journal of Econometrics* **160**(1), 246–256.
- Patton, A. J. & Sheppard, K. (2009), ‘Optimal combinations of realised volatility estimators’, *International Journal of Forecasting* **25**(2), 218–238.
- Pedersen, R. S. & Rahbek, A. (2014), ‘Multivariate variance targeting in the BEKK–GARCH model’, *The Econometrics Journal* **17**(1), 24–55.
- Rice, J. (1984), ‘Bandwidth choice for nonparametric regression’, *The Annals of Statistics* **12**(4), 1215–1230.
- Shephard, N. & Sheppard, K. K. (2010), ‘Realising the future: forecasting with high-frequency-based volatility (heavy) models’, *Journal of Applied Econometrics* **25**(2), 197–231.
- Tse, Y. & Tsui, A. (2002), ‘A multivariate GARCH model with time-varying correlations’, *Journal of Business & Economic Statistics* **20**(3), 351–362.
- Zhang, L. (2011), ‘Estimating covariation: Epps effect, microstructure noise’, *Journal of Econometrics* **160**(1), 33–47.
- Zhang, L. et al. (2006), ‘Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach’, *Bernoulli* **12**(6), 1019–1043.
- Zumbach, G. & Lynch, P. (2001), ‘Heterogeneous volatility cascade in financial markets’, *Physica A: Statistical Mechanics and its Applications* **298**(3), 521–529.