UNIVERSITY OF MESSINA

PHD IN ECONOMICS, MANAGEMENT AND STATISTICS
XXXIV CYCLE
Coordinator: Prof. Edoardo OTRANTO

Department of Economics

Component Volatility Models: A MIDAS Approach

SSD: SECS-S/01

Author:
Luca SCAFFIDI DOMIANELLO

Supervisor:
Prof. Edoardo OTRANTO

A.Y. 2020/2021

Declaration of Authorship

I, Luca SCAFFIDI DOMIANELLO, declare that this thesis titled, "Component Volatility Models: A MIDAS Approach" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: ____

[Luca SCAFFIDI DOMIANELLO]

Place: Messina, Italy

Date: February 2, 2022

"Fragility is the quality of things that are vulnerable to volatility."

Nassim Nicholas Taleb

UNIVERSITY OF MESSINA

Abstract

Department of Economics

Doctor of Philosophy

Component Volatility Models:
A MIDAS Approach

by Luca SCAFFIDI DOMIANELLO

After the seminal paper of Engle (1982) and the generalization provided by Bollerslev (1986), studies on time-varying volatility abound in the literature. Researchers proposed several extensions to capture the so-called *stylized facts*, i.e., the empirical regularities in the series of financial variables like asset returns. One is related to the long-memory behavior, that led to the development of component models to capture, in a parsimonious way, this complex dependence structure. Furthermore, component models based on the MIDAS filter can capture the effect of variables sampled with a lower frequency, such as economic variables, on conditional variance. Then, we can analyze the relationship between financial volatility and economic conditions. So, within this Ph.D. thesis, in the first chapter, we briefly review several univariate and multivariate volatility models, highlighting their drawbacks and the improvements we want to provide with the models we propose in the following two chapters.

Indeed, univariate MIDAS models cannot immediately capture bursts of volatility due to the smoothness of their long-run component. Then, in the second chapter, we propose a new MIDAS model with a markovian dynamic in the short-run component to detect abrupt shifts in the average level of the series. Empirical results indicate that taking into account both abrupt shifts in the average level and economic source of volatility improves the in-sample performance of the model than competitive ones.

In the multivariate framework, component models based on the Cholesky decomposition provide us with a covariance matrix that is sensible to the order of the assets. For this purpose, in the third chapter, we propose a multivariate component model that is invariant to asset ordering with a substantial gain in the estimation time. We provide also a specification, based on the Hadamard exponential function, with time-varying and asset-pair specific parameters. Both in-sample and out-of-sample analyses indicate that the proposed model outperforms the competitive ones.

Acknowledgements

The Ph.D. was a great experience that made me grow from a professional and human point of view, then, I am deeply grateful to the University of Messina and Catania for giving me this opportunity.
Foremost, I would like to express my sincere gratitude to my supervisor, Prof. Edoardo Otranto, for his invaluable expertise and continuous support. I am very proud of my time working with him, I learned a lot from his immense knowledge.
I extend my special thanks to Prof. Fabio Forgione for his help and encouragement during my academic life.
I would also like to thank Prof. Giampiero M. Gallo and Prof. Luc Bauwens for their insightful comments and suggestions on the topics of this dissertation. I am also very grateful to all the professors of the Ph.D. course for their invaluable advice during the academic route.
I want to thank my colleagues with a special mention for Cinzia, Demetrio, Francesca, Sara, Carmelo, and Francesco: it was an honor sharing this amazing experience with all of you!
My appreciation also goes to my friends and all the people who believed in me, for their continuous encouragement throughout my studies.
Last but not least, I express my sincere gratitude to my family for supporting continuously me and all they have done for my education.

Contents

List of Figures

List of Tables

To my mother, my father, and my brother...

Chapter 1

Univariate and Multivariate Volatility Models: a Literature Review

1.1 Introduction

Time-varying volatility models became popular after the pioneering works of Engle (1982) and Bollerslev (1986), who noticed the autoregressive structure of the conditional variance of economic and financial variables (e.g., asset returns). The great success of this class of models is due to their capability of capturing most of the so-called *stilyzed facts* (or empirical regularities) of financial returns: the main one is the volatility clustering, i.e., "large changes tend to be followed by large changes - of either sign - and small changes tend to be followed by small changes" (Mandelbrot, 1963). Another feature of financial returns is related to their leptokurtic unconditional distribution, that is, they have fatter tails than the normal distribution. Nevertheless, several extensions of the baseline model were proposed in the literature, because the latter is not able to capture other stylized facts of financial variables. One is the asymmetric or leverage effect, i.e., negative shocks increase future volatility more than positive shocks of the same size. In addition, financial variables exhibit a long memory behavior, that is the impact of shocks on volatility decays hyperbolically rather than exponentially. Furthermore, in the last two decades, through the availability of ultra-high frequency data, more precise ex-post volatility measures than squared returns were proposed by researchers, and new models based on these measures arose in the literature.

Univariate volatility models were soon followed by their multivariate extension, due to the importance of the correlation between asset returns for financial applications: hedging, asset allocation, pricing, risk management, and so on. Nevertheless, within the multivariate framework, we face the course of dimensionality problem, i.e., the number of parameters to be estimated is, generally, a quadratic function of the number of assets considered, then the estimation of the model becomes unfeasible from a computational point of view. In addition, the models have to ensure positive definite conditional covariance matrices.

Nevertheless, both in a univariate and multivariate framework, the time-varying volatility models described above have some limitations and can be further improved. This is why in the chapter 2 and chapter 3, respectively for the univariate and multivariate framework, we provide some extensions, that try to overcome the drawbacks of the models depicted above.

In this chapter, we provide a brief review of the existing volatility models in literature, highlighting their drawbacks, both in the univariate framework and in the multivariate one. More specifically, section (1.2) analyzes univariate time-varying volatility models that firstly appeared in literature, while section (1.3) examines models that assume a time-varying long-run variance, and section (1.4) introduces models that use the MIxed DAta Sampling (MIDAS) polynomial, a filter that allows handling variables sampled at different frequencies; then, in section (1.5) a new class of models, suitable for volatility measures based on high-frequency data, are presented, with several extensions provided. The second part of the chapter analyzes the multivariate counterpart of univariate volatility models (section (1.6), (1.7), and (1.8). Then, in section (1.9), we introduce multivariate volatility models, that allow the long run matrix to be time-varying, based on the MIDAS filter. Finally, section (1.10) concludes the chapter with some remarks.

1.2 Early conditional volatility models

Engle (1982) specifies the Autoregressive Conditional Heteroskedasticity (ARCH) model as follows:

$$
\begin{aligned}
&r_t = \mu_t + \epsilon_t \\
&\epsilon_t = \sqrt{h_t}\eta_t \qquad \eta_t \sim N(0,1)\ \forall t \\
&E(r_t - \mu_t|\mathcal{I}_{t-1})^2 = E(\epsilon_t^2|\mathcal{I}_{t-1}) = h_t \\
&h_t = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2
\end{aligned}
\tag{1.1}
$$

Where μ_t[1] is the time-varying mean of asset returns (r_t), ϵ_t the residual of the conditional mean equation, $\mathcal{I}_t = \{\epsilon_t, \epsilon_{t-1}, ..., \epsilon_1\}$ the information set available at time t, and η_t is an i.i.d. gaussian disturbance. For what concerns the conditional variance, h_t, it is a linear function of p lagged squared residuals, $\omega > 0$ is a constant, and the α_i ARCH parameters are constrained to be nonnegative in order to ensure the positivity of the process.

Let us now consider the second moment of r_t, the so-called unconditional variance: if the model is covariance-stationary, i.e., the unconditional variance does not depend on t and is equal to a constant, σ^2, then it is specified as follows:

$$\begin{aligned} \sigma^2 = E(h_t) &= \sum_{i=1}^{p} \alpha_i E(\epsilon_{t-i}^2) \\ &= \sum_{i=1}^{p} \alpha_i \sigma^2 \\ &= \frac{\omega}{1 - \sum_{i=1}^{P} \alpha_i} \end{aligned} \tag{1.2}$$

Notice that by the law of iterated expectations $E(\epsilon_t^2) = E[E(\epsilon_t^2|\mathcal{I}_{t-1})] = E(h_t)$ and that, under the assumption of stationarity, $E(\epsilon_t^2) = E(\epsilon_{t-1}^2) = \sigma^2$. Furthermore, the covariance-stationarity of the process requires that $\sum_{i=1}^{P} \alpha_i < 1$. This constraint adds to the positivity constraint discussed above.

Nevertheless, most of the financial variables are persistent, that is the impact of a shock on future variance decays slowly, then a high number of ARCH parameter estimates should be required. At this purpose, Bollerslev (1986) proposed a generalization of the ARCH models (GARCH) with a more flexible, although parsimonious, specification, by taking into account past conditional variances. Let us now consider the variance equation of the GARCH process:

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}^2 \tag{1.3}$$

Here, the conditional variance is a linear function of past squared residuals and conditional variances, with $\omega > 0$, $\alpha_i \geq 0$, and $\beta_i \geq 0$ to guarantee the positivity of the process. Notice that if the GARCH parameters (i.e., β_i) are constrained to be equal to 0, then the GARCH reduces to the ARCH model, in this sense, it is a generalization. Let us now consider the unconditional

[1]The conditional mean of r_t, i.e. μ_t, could be captured through an ARMA model see, e.g., Hamilton (1994), ch. 3.

variance:

$$\begin{aligned} E(h_t) = \sigma^2 &= \sum_{i=1}^{p} \alpha_i E(\epsilon_{t-i}^2) + \sum_{i=1}^{q} \beta_i E(h_{t-i}^2) \\ &= \sum_{i=1}^{p} \alpha_i \sigma^2 + \sum_{i=1}^{q} \beta_i \sigma^2 \\ &= \frac{\omega}{1 - (\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i)} \end{aligned} \tag{1.4}$$

The covariance-stationarity of the process requires that $\sum_{i=1}^{p} \alpha_i + \sum_{i=1}^{q} \beta_i < 1$. In empirical application p and q are very often set equal to 1. Notice that by eq. (1.4), the constant term for the GARCH(1,1) is equal to $(1 - \alpha - \beta)\sigma^2$. Then, the model can be specified as follows:

$$h_t = (1 - \alpha - \beta)\sigma^2 + \alpha\epsilon_{t-1}^2 + \beta h_{t-1} \tag{1.5}$$

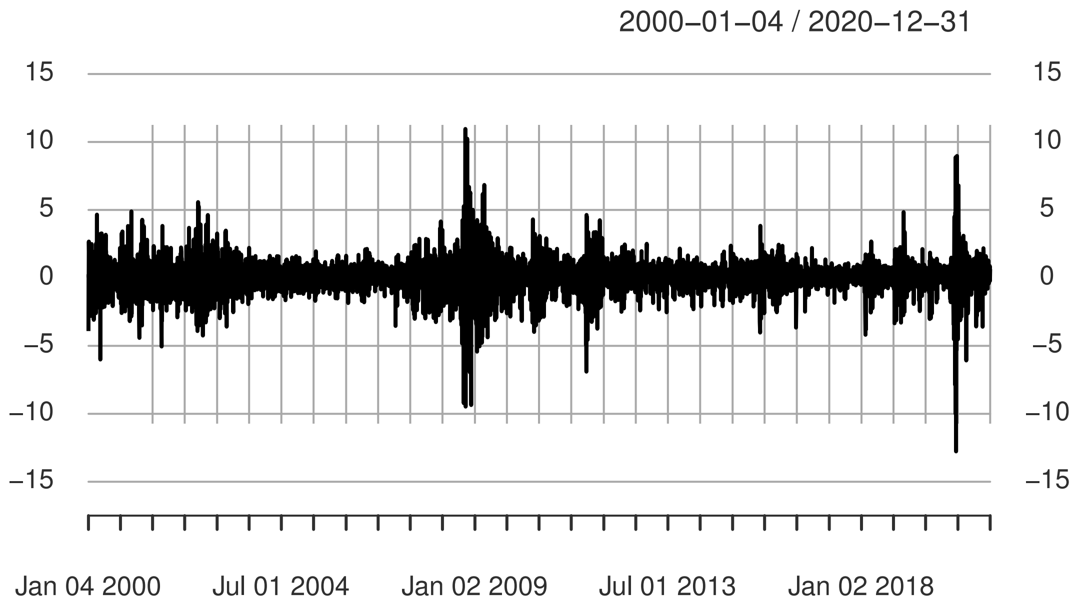
The conditional variance is expressed as a weighted average value of the last squared residual, the unconditional, and the last conditional variance. Notice that, by recursive substitution, the GARCH(1,1) process admits an ARCH(∞) representation, thus confirming the flexibility of the model:

$$\begin{aligned} h_t &= \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2 \\ &= \omega + \alpha\epsilon_{t-1}^2 + \beta(\omega + \alpha\epsilon_{t-2}^2 + \beta h_{t-2}^2) \\ &= \omega + \alpha\epsilon_{t-1}^2 + \beta[\omega + \alpha\epsilon_{t-2}^2 + \beta(\omega + \alpha\epsilon_{t-3}^2 + \beta h_{t-3}^2)] \\ &= \omega \sum_{i=1}^{k} \beta^{i-1} + \alpha \sum_{i=1}^{k} \beta^{i-1}\epsilon_{t-i}^2 + \beta^k h_{t-k}^2 \\ &= \frac{\omega}{1-\beta} + \alpha \sum_{i=1}^{\infty} \beta^{i-1}\epsilon_{t-i}^2 \end{aligned} \tag{1.6}$$

where $\lim_{k\to\infty} \beta^k h_{t-k}^2 = 0$ due to the covariance-stationarity of the process.

The great success of the GARCH(1,1) model is due to its capability of capturing the so-called *stilyzed facts*: the main one is the so-called volatility clustering. Indeed, as we can see from fig. (1.1), returns are "clustered", i.e., it is possible to distinguish between quiet and turmoil periods. Another feature of financial returns is related to their leptokurtic unconditional distribution: they have fatter tails than the normal distribution. Notice that, even if the GARCH(1,1) process assumes that returns are conditionally normally distributed, their unconditional distribution is not gaussian. Let us consider the

FIGURE 1.1: S&P 500 index daily log returns. Sample period: 4 January 2000, 31 December 2020.



Notes: Log returns are expressed in percentage scale.

kurtosis of ϵ_t:

$$\begin{aligned} \frac{E(\epsilon_t)^4}{\sigma^4} &= \frac{E(h_t^2)E(\eta_t^2)}{\sigma^4} = \frac{3E(h_t^2)}{\sigma^4} = 3\frac{1-(\alpha+\beta)^2}{1-2\alpha^2-(\alpha+\beta)^2} \\ E(h_t^2) &= E[(\omega+\alpha\epsilon_{t-1}^2+\beta h_{t-1}^2)^2] = \sigma^4\frac{1-(\alpha+\beta)^2}{1-2\alpha^2-(\alpha+\beta)^2} \end{aligned} \tag{1.7}$$

It easiliy checked that the unconditional kurtosis is greater than 3, in fact $1-(\alpha+\beta)^2 > 1-2\alpha^2-(\alpha+\beta)^2$ if the fourth moment of ϵ_t exists[2].

Nevertheless, the GARCH(1,1) process is not able to take into account the asymmetric response of volatility to a past shock, the so-called asymmetric or leverage effect: Black (1976) found that future volatility and current returns are negatively correlated, i.e. negative shocks increase future volatility more than positive shocks of the same size. One explanation of the asymmetric effect is that a drop in stock prices implies that the firm becomes more leveraged, then it is riskier. Another explanation is related to the volatility feedback hypothesis[3]: assume there is a large piece of good news about future dividends. Other pieces of news will follow, due to the persistence of

[2]The existence of the fourth moment requires that $2\alpha^2+(\alpha+\beta)^2 < 1$. In addition, for the calculation of the fourth moment, see Appendix A.

[3]See, for example, Campbell and Hentschel (1992).

volatility, thus increasing future expected volatility. Nevertheless, increasing expected volatility requires higher expected stock returns, then the stock price drops, dampening the positive impact of good news about future dividends. Whereas, if we consider a large piece of bad news about future dividends, volatility amplifies the effect of bad news, then returns are correlated with future volatility.

To capture the asymmetric effect, several extensions of the GARCH process have been proposed in the literature: here, we consider the Exponential GARCH (EGARCH) model of Nelson (1991) and the GJR GARCH one of Glosten, Jagannathan, and Runkle (1993)[4].

The EGARCH model is parameterized through the exponential function, so the volatility is ensured to be positive without constraints on the parameters:

$$\ln h_t = \omega + \alpha \left(\frac{|\epsilon_{t-1}|}{\sqrt{h_{t-1}}} - \frac{2}{\pi} \right) + \beta \ln h_{t-1} + \gamma \frac{\epsilon_{t-1}}{\sqrt{h_{t-1}}} \tag{1.8}$$

The leverage effect is captured by the parameter γ, expected to be negative, with the impact of an unexpected negative return equals to $\alpha - \gamma$, while, when the unexpected return is positive, the effect is equal to $\alpha + \gamma$.

Whereas, the GJR GARCH is specified as follows:

$$h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1} + \gamma \mathbb{1}_{(\epsilon_{t-1}<0)}\epsilon_{t-1}^2 \tag{1.9}$$

where $\mathbb{1}_{(\epsilon_t<0)}$ is a dummy variable equals to 1 when the return, r_t, at time t is negative, 0 otherwise. The coefficient γ, if positive, captures the asymmetric effect, in fact when the past unexpected return is negative, the impact is equal to $\alpha + \gamma > \alpha$. Let us consider the unconditional variance implied by the model:

$$\begin{aligned} E(h_t) = \sigma^2 &= \alpha E(\epsilon_{t-1}^2) + \beta E(h_{t-1}^2) + \frac{1}{2}\gamma E(\epsilon_{t-1}^2) \\ &= \alpha\sigma^2 + \beta\sigma^2 + \frac{1}{2}\gamma\sigma^2 \\ &= \frac{\omega}{1-(\alpha+\beta+\gamma/2)} \end{aligned} \tag{1.10}$$

Notice that it is assumed that the unexpected returns have a 0 median, and their sign is uncorrelated with the squared residuals. Moreover, the process is covariance stationary if $\alpha + \beta + \gamma/2 < 1$.

[4]GJR is the acronym of the authors of the paper.

1.3 Component Volatility Models

In empirical applications, the estimated persistence of the models is very high, i.e., we have a sum of α and β close to 1. For this purpose Engle and Bollerslev (1986) proposed the Integrated GARCH (IGARCH), i.e., a GARCH(1,1) with the constraint $\alpha + \beta = 1$[5].

However, Ding, Granger, and Engle (1993) showed that the autocorrelation function of squared returns decays hyperbolically rather than exponentially, as implied by a GARCH process, also the IGARCH one[6]. To capture this long memory property, fractionally integrated GARCH processes (FIGARCH) have been proposed in literature[7].

Another approach was pursued by Ding and Granger (1996): by inspecting the behavior of the sample autocorrelation function of squared residuals, they thought that the volatility process is affected, at least, by two different components: one has a short-run effect, while the other one has a long-run effect. They considered the following model:

$$\begin{aligned} h_t &= \omega h_{1t} + (1-\omega) h_{2t} \\ h_{1t} &= \alpha_1 \epsilon_{t-1}^2 + (1-\alpha_1) h_{1t-1} \\ h_{2t} &= \sigma^2 (1 - \alpha_2 - \beta_2) + \alpha_2 \epsilon_{t-1}^2 + \beta_2 h_{2t-1} \end{aligned} \tag{1.11}$$

where h_{1t} and h_{2t} are the two volatility components, while ω and $1-\omega$ are the weights of each component, that sum up to one. The first component has an IGARCH structure, while the second one is a GARCH(1,1) process. They found that their model captures the sample autocorrelation of squared residuals better than the simple GARCH(1,1).

Engle and Lee (1999) proposed a component model that is the sum of a more persistent component and a shorter-lived one. Let us consider the alternative parameterization of the GARCH(1,1):

$$h_t = \sigma^2 + \alpha(\epsilon_{t-1}^2 - \sigma^2) + \beta(h_{t-1} - \sigma^2)$$

By allowing the constant long-run level, σ^2, to be time-varying, and to evolve

[5] Note that with this parameterization the unconditional variance does not exist, although the process is strictly stationary, see Nelson (1990).

[6] See Ding and Granger (1996).

[7] See, for example, the FIGARCH of Baillie, Bollerslev, and Mikkelsen (1996).

through an autoregressive structure, their proposed model is specified as follows:

$$\begin{aligned} h_t - q_t &= \alpha(\epsilon^2_{t-1} - q_{t-1}) + \beta(h_{t-1} - q_{t-1}) \\ q_t &= \omega + \rho q_{t-1} + \phi(\epsilon^2_{t-1} - h_{t-1}) \end{aligned} \tag{1.12}$$

By rearranging the terms, we can identify the two components of the model:

$$\begin{aligned} h_t &= q_t + s_t \\ q_t &= \omega + \rho q_{t-1} + \phi(\epsilon^2_{t-1} - h_{t-1}) \\ s_t &= (\alpha + \beta)s_{t-1} + \alpha(\epsilon^2_{t-1} - h_{t-1}) \end{aligned} \tag{1.13}$$

where s_t is the short-run component with a zero mean, that captures transitory shocks, and q_t is the long-run component, that follows an AR process[8], aimed at capturing permanent effects. The usually constraints are required in order to ensure positiveness and covariance-stationarity of the process: $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta < 1$, $\phi > 0$ and $0 < \rho < 1$. In addition, identification of the model requires that $0 < \alpha + \beta < \rho < 1$, that is, the long-run component is more persistent. Their empirical results show that the long-run component is much less sensitive to volatility shocks ($\phi < \alpha$), while the short-run component is much less persistent than the long-run component $(\alpha + \beta) << \rho$. Interestingly, they found that the 1987 crash had only transitory effects, in fact, the market reached the pre-crash level very quickly.

1.4 GARCH-MIDAS

Component volatility models discussed above are characterized by the summation of two components, while Engle, Ghysels, and Sohn (2013) proposed a multiplicative component model, called GARCH-MIDAS, in which the conditional variance is multiplicatively decomposed into a short-run component and a long-run one[9]. Let $\epsilon_{i,t}$ be our residual for the *i-th* day of the lower frequency period t (a week, a month, a quarter, etc, with N_t the number of days for that period), then the conditional variance of the GARCH-MIDAS model

[8]Notice that $\epsilon^2_{t-1} - h_{t-1}$ is a martingale difference sequence (MDS).

[9]Here, we do not consider models with a deterministic long-run component (see, for example, Engle & Rangel, 2008).

is defined as follows:

$$
\begin{aligned}
h_{i,t} &= g_{i,t}\tau_t \\
g_{i,t} &= 1-\alpha-\beta-\gamma/2+\alpha\frac{\epsilon^2_{i-1,t}}{\tau_t}+\beta g_{i-1,t}+\gamma\mathbb{1}_{(\epsilon_{i-1,t}<0)}\frac{\epsilon^2_{i-1,t}}{\tau_t} \\
\tau_t &= exp\left\{m+\theta\sum_{k=1}^{K}\varphi_k(\lambda_1,\lambda_2)X_{t-k}\right\} \\
\varphi_k(\lambda_1,\lambda_2) &= \frac{(k/K)^{\lambda_1-1}(1-k/K)^{\lambda_2-1}}{\sum_{j=1}^{K}(j/K)^{\lambda_1-1}(1-j/K)^{\lambda_2-1}}
\end{aligned}
\tag{1.14}
$$

where $g_{i,t}$ is the short-run component representing the known volatility clustering and the daily fluctuations. The parameterization of the constant term in $g_{i,t}$ implies that its unconditional value is equal to $E(g_{i,t}) = 1$ as identifying condition. While, τ_t is the long-run component driven by a low frequency stationary variable, X_t. Notice that if τ_t is assumed to be constant, the model reduces to a simple GARCH. For what concerns the long-run component, the exponential form is used to ensure its positiveness, because the low-frequency variable can assume, potentially, both positive and negative values. Moreover, note that τ_t is predetermined with respect to the information set available at time $t-1$, in fact $E(x_t|\mathcal{I}_{t-1}) = \tau_t E(g_{i,t}|\mathcal{I}_{t-1}) = \tau_t$[10] and it is constant throughout the whole period t. That is, the long-run component is the changing level around which conditional variance fluctuates.

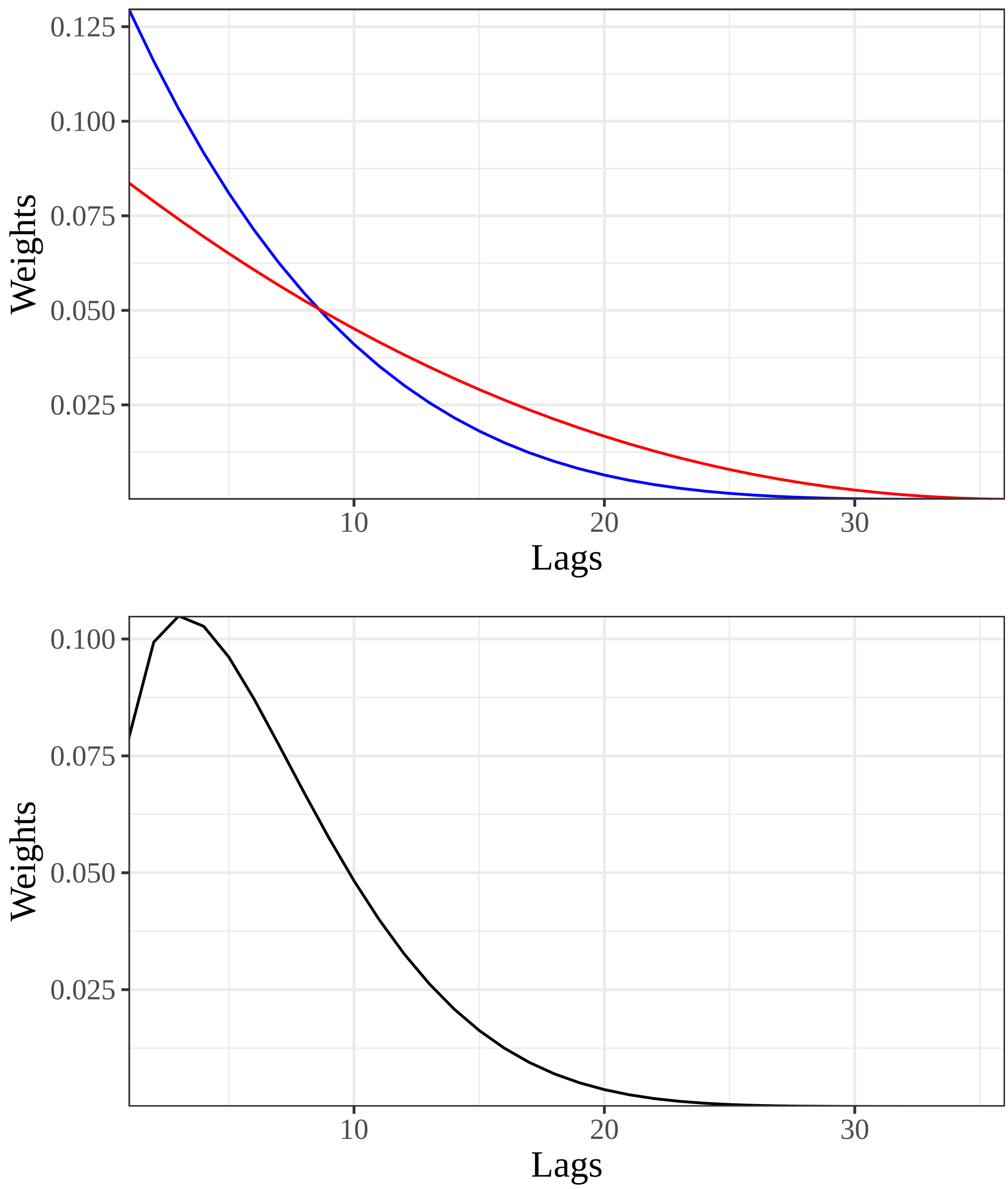
The MIDAS filter is $\varphi_k(\lambda_1,\lambda_2)$, a weighting function of the past K values of X_t, with the weights that sum up to one. This filter, based on the beta function, is called beta polynomial, and it is quite flexible, allowing us to handle in the same equation variables sampled at different frequencies. As we can see at the top of fig. (1.2), if $\lambda_1 = 1$ and $\lambda_2 > 1$, the MIDAS function is monotonically decreasing, as fast as λ_2 increases. While, at the bottom, the weights show a hump-shaped pattern[11]. The effect of the weighted sum of the lagged values of X_t is captured through the coefficient θ.

The low-frequency variable could be financial (VIX index, monthly Realized Volatility, etc.) or economic (Industrial Production, Inflation rate, Unemployment rate, etc.). Besides the level of the economic variables, one could also consider the volatility of the low-frequency variable. In addition, it could be interesting to augment the model with future values (or expected values) of the low-frequency variable, to analyze the so-called lead-lag relationship (see

[10]If we assume that the expectation of the short-run component is equal to its unconditional value, that is 1 (See Engle, Ghysels, and Sohn, 2013).

[11]For a complete discussion on the shapes of the beta lag and the other smoothing functions, see Ghysels, Sinko, and Valkanov (2007).

FIGURE 1.2: Beta polynomial Midas weights



Notes: In both panels the weights of the Beta polynomial MIDAS filter are plotted on 36 lags. At the top, the weights of the red line and the blue one both share a $\lambda_1 = 1$, while the $\lambda_2 s$ are respectively equal to 3 and 5. At the bottom the weights of the black line are characterized by a $\lambda_1 = 1.6$ and a $\lambda_2 = 7.5$

Engle, Ghysels, and Sohn, 2013). In this case, the specification of the long-run component is the following:

$$\tau_t = exp\left\{ m + \theta^b \sum_{k=1}^{K^b} \varphi_k(\lambda_1, \lambda_2) X_{t-k} + \theta^f \sum_{k=-K^f}^{0} \varphi_k(\lambda_1, \lambda_2) X_{t-k} \right\} \tag{1.15}$$

where for $k = 1, ..., K^b$ we have the lagged values of the explanatory variable, X_t, while for $k = 0, ..., K^f$ we have the current and future values. Note that the coefficient of past and future (expected) values of X_t is allowed to change, that is the impact can be different (θ^b and θ^f), but the filter weights are the same, so the model remains parsimonious.

Generally, the GARCH-MIDAS is estimated with one exogenous variable, X_t, to avoid a proliferation of parameters. Interestingly, when Engle, Ghysels, and Sohn (2013) consider the Industrial Production (IP) growth rate as a low-frequency variable, they find a negative estimated coefficient θ, that is, a contraction of IP increases conditional variance, the so-called countercyclical pattern of financial volatility documented by Officer (1973) and Schwert (1989), i.e., volatility is high during a recession and low during an expansion phase. In addition, models in which the long-run component is driven by economic variables tend to outperform the other models for a long-term horizon.

After the work of Engle, Ghysels, and Sohn (2013), different studies concentrated on the GARCH-MIDAS, aiming at improving its performance. For example, Asgharian, Hou, and Javed (2013) applied a principal component analysis (PCA) to capture the information contained in different macroeconomic variables, thus avoiding the convergence problem due to the inclusion of several exogenous variables. Their study showed that the inclusion of economic information, through the PCA, improves the long-term forecasting capability of the model. Whereas, Conrad and Loch (2015) concentrated on the lead-lag relationship discussed above, by employing a two-sided filter. Differently from Engle, Ghysels, and Sohn (2013), they used expectations about future macroeconomic variables to employ a feasible two-sided filter. More interestingly, they identify some macroeconomic variables as leading indicators (such as the term spread and housing starts) while other variables as lagging indicators (Industrial Production and unemployment rate).

Amendola, Candila, and Gallo (2019) proposed an Asymmetric GARCH MIDAS (A-GARCH MIDAS), which allows the macroeconomic variables to have a different impact on financial volatility according to their sign. Then, the long-run component is specified as follows:

$$\tau_t = exp\left\{ m + \theta^+ \sum_{k=1}^{K} \varphi_k(\lambda_1^+, \lambda_2^+) X_{t-k} \mathbb{1}_{(X_{t-k} \geq 0)} + \theta^- \sum_{k=1}^{K} \varphi_k(\lambda_1^-, \lambda_2^-) X_{t-k} \mathbb{1}_{(X_{t-k} < 0)} \right\} \tag{1.16}$$

Notice that we have two kind of asymmetry: one related to the sign specific parameters (θ^+ and θ^-) and the other related to the shape of the weighting function ($\varphi_k(\lambda_1^+, \lambda_2^+)$ and $\varphi_k(\lambda_1^-, \lambda_2^-)$). Their empirical application confirms the asymmetric response of volatility on the macroeconomic variable, with their model outperforming the competitive ones, both in-sample and out-of-sample.

1.5 Multiplicative Error Model

In a GARCH framework, that is in all the models discussed above, conditional volatility is extracted from the daily demeaned squared returns. In the last two decades, researchers, through the availability of ultra-high frequency data (i.e., intraday returns), proposed new ex-post daily volatility measures, that became the target to evaluate forecasts of volatility models. Due to the high liquidity of financial markets, we can think of asset price and return series as discrete observations from a continuous-time process. Let us consider the instantaneous return:

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \tag{1.17}$$

where $dp(t)$ is the continuous log-price increment, μ_t the deterministc local drift, $\sigma(t)$ the spot volatility and $dW(t)$ a standard brownian motion. From this specification, it derives that the one-period return is:

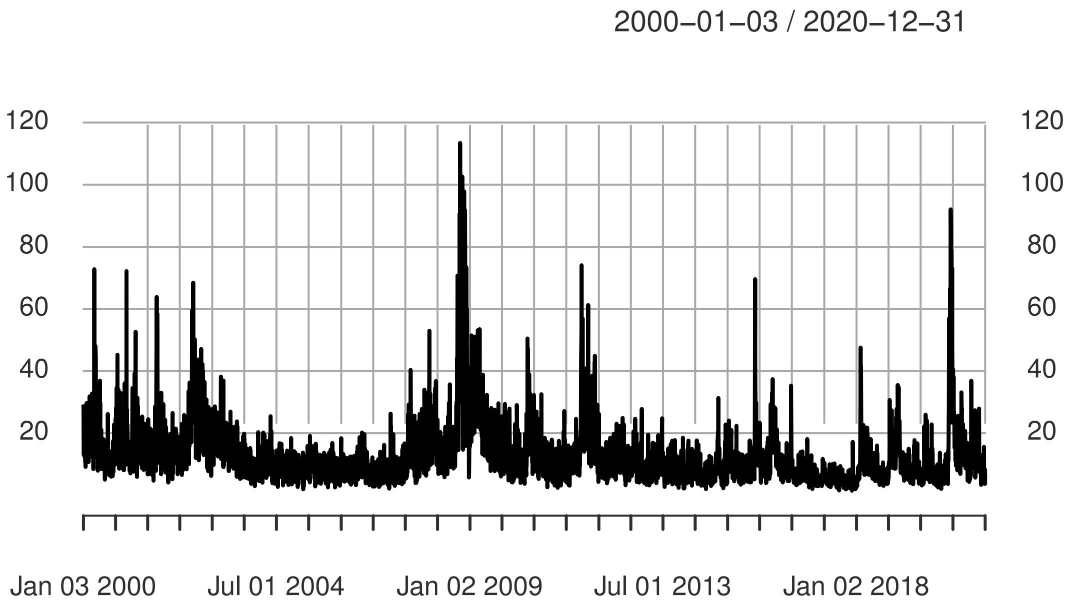
$$r_t = p(t) - p(t-1) = \int_{t-1}^{t} \mu(s)ds + \int_{t-1}^{t} \sigma(s)dW(s) \tag{1.18}$$

and $\int_{t-1}^{t} \sigma(s)d(s)$ is the Integrated Variance (IV). Notice that the latter is not observable, but it can be approximated, according to Andersen and Bollerslev (1998), through the Realized Variance (RV), that is the following discrete sum of squared high-frequency returns over $[t-1, t]$:

$$RV_t = \sum_{i=1}^{M} r_{i,t}^2 \tag{1.19}$$

where $r_{i,t}$ is the high frequency return and M the number of intraday returns considered within the period of length $[t-1, t]$. Andersen and Bollerslev (1998) proved that the Realized Variance is a much more precise estimator of the Integrated Variance than the daily squared returns. Although squared returns are an unbiased estimator, they are very noisy. In addition, the RV estimator keeps the information provided by intraday returns, which is excluded by the GARCH models. As we can see from fig. (1.3), RV measure

FIGURE 1.3: S&P 500 index annualized Realized kernel Volatility. Sample period: 3 January 2000, 31 December 2020.



Notes: Annualized Realized kernel volatility is expressed in percentage scale.

shares the feature of long-run dependence. For this purpose, the Multiplicative Error Model (MEM) (Engle, 2002b; Engle and Gallo, 2006) a class of time-series models for positive valued processes (e.g., Realized Variance[12]), could be employed to directly model Realized Volatility and to obtain more accurate forecasts than the GARCH models. In the MEM framework, the volatility process, $\{x_t\}$, is the product of a time-varying scale factor and a standard positive valued random error. The advantage of this specification is to provide the conditional expectation of the variable of interest rather than the expectation of the log. The model is specified as follows:

$$\begin{aligned} x_t &= \mu_t \epsilon_t, \\ \epsilon_t|\mathcal{I}_{t-1} &\sim Gamma\,(a, 1/a)\ \forall t \\ \mu_t &= \omega + \alpha x_{t-1} + \beta \mu_{t-1} + \gamma \mathbb{1}_{(r_{t-1}<0)} x_{t-1} \end{aligned} \tag{1.20}$$

where μ_t is a positive quantity predetermined conditionally at the information set of the previous period, $\mathcal{I}_{t-1} = \{x_{t-1}, x_{t-2}, ..., x_1\}$, while the error term, ϵ_t, follows a Gamma distribution with a unit mean[13]. The coefficient γ captures the asymmetric effect (negative returns tend to increase future

[12]It is possible to consider also, for example, durations - see Engle and Russell (1998) - and volumes - see Manganelli (2005).

[13]In order to ensure the nonnegativeness of x_t, the error term is defined on a positive support.

volatility more than positive ones). Moreover, to ensure the positiveness and the stationarity of the process, the following constraints are applied: $\omega > 0$, $\alpha > 0$, $\beta \geq 0$, $\gamma \geq 0$ and $\alpha + \beta + \gamma/2 < 1$.

The specification in eq. (1.20) implies that the conditional mean of x_t is equal to:

$$E\left(x_t|\mathcal{I}_{t-1}\right) = \mu_t \tag{1.21}$$

while, due to the stationarity of the process, the unconditional mean is equal to:

$$E\left(x_t\right) = \mu = \frac{\omega}{1-\alpha-\beta-\gamma/2} \tag{1.22}$$

whereas, the conditional variance of the process is specified as follows:

$$Var\left(x_t|\mathcal{I}_{t-1}\right) = \frac{\mu_t^2}{a} \tag{1.23}$$

Furthermore, also the unconditional variance of the process is derived[14]:

$$Var\left(x_t\right) = \mu^2 \frac{1/a(1-2\alpha\beta-\beta^2)}{1-[(\alpha+\beta)^2+(1/a)\alpha^2]} \tag{1.24}$$

In the spirit of Engle and Lee (1999), Brownlees, Cipollini, and Gallo (2012) proposed the Asymmetric Composite MEM, thus providing an extension to the component models[15]. As for the GARCH(1,1), from (1.22) $\omega = \mu(1-\alpha-\beta-\gamma/2)$, then the MEM can be specified as follows:

$$\begin{aligned} \mu_t &= \mu + \xi_t \\ \xi_t &= \alpha x_{t-1}^{(\xi)} + \beta\xi_{t-1} + \gamma \mathbb{1}_{(r_{t-1}<0)} x_{t-1}^{(\xi-)} \end{aligned} \tag{1.25}$$

where $x_{t-1}^{(\xi)} = x_t - \mu$ and $x_{t-1}^{(\xi)} = \mathbb{1}_{(r_{t-1}<0)} x_{t-1} - \mu/2$. Brownlees, Cipollini, and Gallo (2012) allows the unconditional variance, μ, to be time-varying:

$$\mu_t = \chi_t + \xi_t \tag{1.26}$$

where the dynamic of χ_t is specified as:

$$\chi_t = \omega^{(\chi)} + \alpha^{(\chi)} x_{t-1}^{(\chi)} + \beta^{(\chi)} \chi_{t-1} \tag{1.27}$$

[14]See appendix B.

[15]For other component models within the MEM class, see, e.g., the P-Spline MEM (Brownlees & Gallo, 2010) with a deterministic long-run component, and the Spillover AMEM (Otranto, 2015) in which the conditional variance is the sum of one component, representing the proper volatility dynamics, and another one representing the volatility transmitted from the other market (the so-called spillover effect).

with $x_{t-1}^{(\chi)} = x_t - \xi_t$. Identifying condition requires that $\alpha^{(\chi)} + \beta^{(\chi)} > \alpha + \beta + \gamma/2$, so that the long run component is more persistent. They found that the Composite MEM solves the problem of residual autocorrelations of the baseline MEM.

Amendola et al. (2020) applied the logic of the MIDAS approach to the MEM framework, by exploiting the relation between economics and financial volatility through the Realized Volatility estimator, rather than the demeaned squared returns like the GARCH-MIDAS. As for the latter, their component model, called MEM MIDAS, is a multiplicative component volatility model in which the conditional variance is split into a short-run component and a long-run one. Let $x_{i,t}$ be the Realized Volatility for the *i-th* day of the lower frequency period t, with N_t the number of days for that period, then the MEM-MIDAS model is specified as follows:

$$
\begin{aligned}
& x_{i,t} = g_{i,t}\tau_t\epsilon_{i,t} = \mu_{i,t}\epsilon_{i,t} \\
& \epsilon_{i,t}|\mathcal{I}_{i-1,t} \sim Gamma\left(a, \frac{1}{a}\right) \ \forall i = 1, ..., N_t \\
& g_{i,t} = 1 - \alpha - \beta - \gamma/2 + \alpha\frac{x_{i-1,t}}{\tau_t} + \beta g_{i-1,t} + \gamma \mathbb{1}_{(r_{i-1,t}<0)}\frac{x_{i-1,t}}{\tau_t} \\
& \tau_t = exp\left\{m + \theta\sum_{k=1}^{K}\varphi_k(\lambda_1,\lambda_2)X_{t-k}\right\} \\
& \varphi_k(\lambda_1,\lambda_2) = \frac{(k/K)^{\lambda_1-1}(1-k/K)^{\lambda_2-1}}{\sum_{j=1}^{K}(j/K)^{\lambda_1-1}(1-j/K)^{\lambda_2-1}}
\end{aligned}
\tag{1.28}
$$

where $g_{i,t}$ is the short-run component representing the known volatility clustering and the daily fluctuations. Whereas τ_t is the long-run component, that is the changing level around which conditional variance fluctuates, driven by a low frequency stationary variable, X_t. As for the GARCH-MIDAS, if τ_t is assumed to be constant, the model reduces to the simple MEM. For what concerns the MIDAS filter, $\varphi_k(\lambda_1, \lambda_2)$, it is defined as in the GARCH-MIDAS discussed above. Their empirical analysis showed that the MEM-MIDAS outperforms the respective GARCH-MIDAS, though the long-run component of the MEM-MIDAS model is not able to capture immediately bursts of volatility and generally abrupt shifts in the average level due to its smooth pattern.

1.6 Multivariate Garch

The extension of univariate time-varying volatility models to the multivariate framework is of primary interest for researchers due to the importance of asset returns correlation for financial applications: hedging, asset allocation, pricing, risk management, and so on. Nevertheless, as the number of assets increases, we face the course of dimensionality problem, i.e., the number of parameters to be estimated are, generally, a quadratic function of the number of assets considered, then the estimation of the model becomes unfeasible from a computational point of view. In summary, we have to provide a specification flexible enough without imposing too strong constraints on the model parameters. In addition, we have to specify a model that ensures positive definite conditional covariance matrices, then additional constraints on the parameters could be required.

A direct generalization of the GARCH model to the multivariate framework is the VECH model of Bollerslev, Engle, and Wooldridge (1988), where each variance and covariance is a linear function of past variances, covariances, squared residuals, and cross-product residuals. Let ϵ_t be a n-dimensional vector of residuals, that conditionally at the information set at time $t-1$, $\mathcal{I}_{t-1}$, follows a n-dimensional Normal distribution. Then, the VECH(1,1) model is specified as follows:

$$\begin{aligned}
&\epsilon_t = H_t^{1/2}\eta_t \qquad \eta_t \sim N\left(0, I_n\right)\ \forall t \\
&Var(\epsilon_t|\mathcal{I}_{t-1}) = H_t^{1/2}Var(\eta_t|\mathcal{I}_{t-1})(H_t^{1/2})' = H_t \\
&vech(H_t) = w + Avech(\epsilon_{t-1}\epsilon_{t-1}^{'}) + Bvech(H_{t-1})
\end{aligned} \tag{1.29}$$

where $vech(\cdot)$ is the vector half operator that stacks the lower portion of a symmetric matrix, $H_t^{1/2}$ is the Cholesky factor of the conditional covariance matrix, H_t, such that $H_t^{1/2}(H_t^{1/2})' = H_t$. $w = vech(W)$ is a $N(N+1)/2$ column parameter vector, while A and B are square parameter matrices of order $N(N+1)/2$. Notice the high number of parameters to be estimated, that is, $N(N+1)(N(N+1)+1)/2$. Moreover, to ensure the positive definiteness of H_t strong restrictions have to be imposed on the parameters. Covariance stationarity of the process requires that the eigenvalues of $A+B$ are less then 1 in modulus. Let $\Sigma = E(\epsilon_t\epsilon_t^{'})$. Then by covariance stationarity $vech(\Sigma) = (I_{n(n+1)/2} - A - B)^{-1}w$.

To reduce the number of parameters to be estimated, Bollerslev, Engle, and Wooldridge (1988) proposed the diagonal version of the VECH model, in which the matrices A and B are assumed to be diagonal, so that each covariance, h_{ij}, depends only on its past values and innovations, $\epsilon_{it-1}\epsilon_{jt-1}$. Sufficient conditions to the positive definiteness of the conditional covariance matrix can be easily derived if we consider the Hadamard representation of the diagonal vech:

$$H_t = W + A \odot \epsilon_{t-1}\epsilon_{t-1}^{'} + B \odot H_{t-1} \tag{1.30}$$

where $\odot$ is the element-wise (Hadamard) product[16], W, A and B are symmetric $N \times N$ matrices. For this parameterization the number of parameters is equal to $(3N(N+1))/2$ and it ensures positive definite covariance matrices if A, B, W, and the initial value of H_t, H_0, are positive definite[17],[18]. The initial value of H_t can be set equal to the sample covariance matrix, that is $H_0 = T^{-1}\sum_{t=1}^{T}\epsilon_t\epsilon_t^{'}$, while we can ensure the positive definiteness of A, B, W through the Cholesky decomposition. A more restricted model assumes the matrices A and B to be rank one, that is $A = aa^{'}$, where a is a n-dimensional vector. The rank-one version of the diagonal Vech requires $N(N+5)/2$ parameters to be estimated.

Engle and Kroner (1995) proposed the BEKK model, that through the use of the quadratic forms, ensures the positive definiteness of the conditional covariance matrix, without additional constraints on the model parameters:

$$H_t = W + A\epsilon_{t-1}\epsilon_{t-1}^{'}A^{'} + BH_{t-1}B' \tag{1.31}$$

where $W = CC'$ with C an order n lower triangular matrix to ensure the positive definiteness of W, while A and B are two order n square matrices. The above parameterization ensure positive definite H_t if H_0 is positive definite. Furthermore, identifying conditions require that a_{11}, b_{11}, and the diagonal elements of C must be positive. The parameters in the matrices A and B do not represent the direct impact of past variances, covariances, squared residuals, and cross-product residuals on variances and covariances. Let us consider,

[16]The Hadamard product of two matrices $A = (a_{ij})$ and $B = (b_{ij})$ with the same dimension is defined as $A \odot B = (a_{ij}b_{ij})$

[17]Note that the matrix $\epsilon_{t-1}\epsilon_{t-1}^{'}$ is positive semidefinite.

[18]See Ding and Engle (2001).

for example, two assets and assume that there is not GARCH effect:

$$\begin{aligned}
h_{11,t} &= c_{11}^2 + a_{11}^2\epsilon_{1,t-1}^2 + 2a_{11}a_{12}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{12}^2\epsilon_{2,t-1}^2 \\
h_{12,t} &= c_{11}c_{21} + a_{11}a_{21}\epsilon_{1,t-1}^2 + (a_{12}a_{21} + a_{11}a_{22})\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{12}a_{22}\epsilon_{2,t-1}^2 \\
h_{22,t} &= c_{22}^2 + a_{21}^2\epsilon_{1,t-1}^2 + 2a_{22}a_{21}\epsilon_{1,t-1}\epsilon_{2,t-1} + a_{22}^2\epsilon_{2,t-1}^2
\end{aligned} \tag{1.32}$$

Then, excluding constant terms and the GARCH effect, instead of the 9 parameters required for the VECH model, we have only 4 parameters. This is because the parameters governing the dynamics of the covariance are the product of the parameters governing the dynamics of the variances. Any BEKK model can be expressed in the VECH form and it is unique[19]. If we vectorize, $vech(\cdot)$, eq. (1.31) we obtain[20]:

$$vech(H_t) = D_n^+ W + D_n^+(A \bigotimes A)D_n vech(\epsilon_{t-1}\epsilon_{t-1}^{'}) + D_n^+(B \bigotimes B)D_n vech(H_{t-1}) \tag{1.33}$$

where $\otimes$ is the Kronecker product[21], D_n is an order $n^2 \times n(n+1)/2$ matrix, called duplication matrix, while D_n^+ is its generalized inverse, that is $D_n^+ = (D_n^{'}D_n)^{-1}D_n^{'}$. Then, covariance stationarity requires that the eigenvalues of $D_n^+(A \otimes A)D_n + D_n^+(B \otimes B)D_n$ are less than one in modulus.
So, $vech(\Sigma) = \left[I_{n(n+1)/2} - D_n^+(A \otimes A)D_n + D_n^+(B \otimes B)D_n\right]^{-1} vech(W)$. Notice that the constant parameters matrix of the BEKK can be specified as: $W = \Sigma - A\Sigma A^{'} - B\Sigma B^{'}$. Then the model can be rewritten as follows:

$$H_t = \Sigma - A\Sigma A^{'} - B\Sigma B^{'} + A\epsilon_{t-1}\epsilon_{t-1}^{'}A^{'} + BH_{t-1}B^{'} \tag{1.35}$$

This is the variance targeting representation, where Σ can be estimated through the sample covariance, that is $\hat{\Sigma} = T^{-1}\sum_{t=1}^{T}\epsilon_t\epsilon_t^{'}$. Thus the number of parameters to be estimated is $2N^2$ instead of $2N^2 + N(N+1)/2$. Notice that we have to ensure the positive definiteness of the matrix $\Sigma - A\Sigma A^{'} - B\Sigma B^{'}$.

A constrained version of the BEKK is the diagonal BEKK, where the matrices A and B are diagonal. Assume that a and b are two n-dimensional vectors,

[19]See Engle and Kroner (1995).
[20]Notice that $vec(ABC) = (C^{'} \otimes A)vec(B)$; $vec(A) = D_n vech(A)$; $vech(A) = D_n^+ vec(A)$.
[21]Given a $(k \times l)$ matrix A and a $(m \times n)$ matrix B, the $(km \times ln)$ matrix $A \otimes B$ is specified as follows:

$$A \bigotimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1l}B \\ a_{21}B & a_{22}B & \cdots & a_{2l}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}B & a_{k2}B & \cdots & a_{kl}B \end{bmatrix} \tag{1.34}$$

then the diagonal BEKK is specified as follows:

$$\begin{aligned} H_t &= W + diag(a)\epsilon_{t-1}\epsilon_{t-1}^{'}diag(a) + diag(b)H_{t-1}diag(b) \\ &= W + aa^{'} \odot \epsilon_{t-1}\epsilon_{t-1}^{'} + bb^{'} \odot H_{t-1} \end{aligned} \tag{1.36}$$

where $diag(a)$ and $diag(b)$ are two $N \times N$ diagonal matrices whose elements on the diagonal are given, respectively, by the vectors a and b. Note that the diagonal BEKK is equivalent to the rank-one version of the diagonal VECH. A more constrained version of the BEKK model is the scalar BEKK, which assumes $A = \alpha I$ and $B = \beta I$, where I is an order n Identity matrix. That is, the parameters governing the dynamics of the process are the same for each series of variances and covariances. This is a strong condition, although it reduces the number of parameters.

Above, we have models for variances and covariances, while, Bollerslev (1990) pursued another approach and proposed the Constant Conditional Correlation (CCC) model, based on the decomposition of the conditional covariance matrix into variances and correlations:

$$H_t = D_t R D_t \tag{1.37}$$

where D_t is a diagonal matrix with each diagonal elements, $d_{ii,t}$, equal to the conditional standard deviation of the $i-th$ asset, $h_{i,t}^{1/2}$, and R is a symmetric constant parameter matrix, representing the constant correlation, with diagonal elements, r_{ii}, equal to 1 by construction. The conditional variance of each asset, separately modeled, is specified as a GARCH(1,1) process:

$$h_{i,t} = w_i + \alpha_i \epsilon_{i,t-1}^2 + \beta_i h_{i,t-1} \tag{1.38}$$

Notice that the matrix H_t is positive definite and it is a proper conditional covariance matrix if the conditional standard deviations are positive and the constant correlation matrix, R, is positive definite. Notice that this model requires $N(N+5)/2$ parameters to be estimated.

Engle (2002a) proposed the Dynamic Conditional Correlation (DCC) Model, thus allowing the correlation matrix, R, to be time varying, R_t:

$$\begin{aligned} H_t &= D_t R_t D_t \\ R_t &= diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \\ Q_t &= (1-\alpha-\beta)S + \alpha\xi_{t-1}\xi_{t-1}^{'} + \beta Q_{t-1} \end{aligned} \tag{1.39}$$

where $diag(Q_t)$ is a diagonal matrix whose elements are the diagonal elements of Q_t, where the latter is the so called quasi conditional correlation matrix[22]. S is the unconditional correlation matrix with diagonal elements equal to 1, and $x_{i,t}$ is the vector of the degarched residuals, that is $\xi_t = \epsilon_t \oslash h_t^{1/2}$, where $h_t^{1/2}$ is the vector of the diagonal elements of D_t[23]. α and β are non negative scalar parameters with $\alpha + \beta < 1$ to ensure the mean reverting property. Notice that the matrix Q_t is ensured to be positive definite if Q_0 is positive definite, while $\xi_{t-1}\xi_{t-1}^{'}$ is positive semidefinite. Then we recover the conditional correlation matrix, R_t through the standardisation provided in the second line of eq. (1.39).

1.7 Invariant Property

A useful feature that a model should share is the so-called invariant property (see, e.g., Francq and Zakoian, 2019 and Bauwens, Laurent, and Rombouts, 2006). Indeed, If $Var(\epsilon_t|\mathcal{I}_{t-1}) = H_t$, then $Var(Z\epsilon_t|\mathcal{I}_{t-1}) = \hat{H}_t = ZH_tZ'$, where Z is an invertible square matrix. It requires that the conditional variance $\hat{H}_t$ belongs to the same class of specification for H_t for any choice of Z. Then, if ϵ_t is a vector of asset returns, $Z\epsilon_t = \hat{\epsilon}_t$ is a vector of portfolios using the same assets, with each row of Z representing the weights of each asset. So, if the invariant property for a class of model holds, it is not necessary to re-estimate the model. Let us consider now the invariant property for the BEKK model: Assumes that $H_t = W + A\epsilon_{t-1}\epsilon_{t-1}^{'}A' + BH_{t-1}B'$, then:

$$\hat{H}_t = ZWZ' + ZA\epsilon_{t-1}\epsilon_{t-1}^{'}A'Z' + ZBH_{t-1}B'Z' = \hat{W} + \hat{A}\hat{\epsilon}_{t-1}\hat{\epsilon}_{t-1}^{'}\hat{A}^{'} + \hat{B}\hat{H}_{t-1}\hat{B}^{'} \tag{1.40}$$

where $\hat{W} = ZWZ'$, $\hat{A} = ZAZ^{-1}$, and $\hat{B} = ZBZ^{-1}$. Note that for the scalar BEKK:

$$\hat{H}_t = \hat{W} + \alpha\epsilon_{t-1}\epsilon_{t-1}^{'} + \beta\hat{H}_{t-1} \tag{1.41}$$

This property does not hold for the diagonal BEKK because if A is diagonal, ZAZ^{-1} is not diagonal, that is $\hat{\epsilon}_t$ does not follow a diagonal BEKK. The same applies for the DCC model, indeed if $H_t = D_tR_tD_t$, ZD_t is not a diagonal matrix. It is useful to stress that if this property does not hold, it does not mean that we cannot empirically use the model, but it could be useful in financial applications.

[22]The off-diagonal elements of Q_t are not ensured to range between -1 and 1, then it is not a proper correlation matrix.

[23]The element-wise division of two vectors a and b with the same dimension is defined as $a \oslash b = (a_i/b_i)$.

1.8 Conditional Autoregressive Wishart Model

Through the availability of ultra-high frequency data (i.e., intraday returns), econometricians compute more precise ex-post covariance matrices measures (e.g., Realized Covariances), and models that directly model Realized Covariance matrices arose in literature to evaluate whether models based on high-frequency data provides a better performance, both in-sample and out of sample, than models based on daily data.

At this purpose, Golosnoy, Gribisch, and Liesenfeld (2012) proposed the Conditional Autoregressive Wishart (CAW) model, in which the Realized Covariance matrix is assumed to follow a Wishart distribution, while the scale matrix, representing the conditional expected value of the Realized Covariance matrix, follows an autoregressive moving average process, i.e., it depends on past realized and conditional covariance matrices. More specifically, for the scale matrix, they proposed a BEKK-type recursion (call it ReBEKK).

Let C_t be an order n positive definite symmetric (PDS) square matrix, here the realized covariance (RC) one, that conditionally at the information set available at time $t-1$, is assumed to follow a n-dimensional Wishart distribution:

$$C_t|\mathcal{I}_{t-1} \sim W_n\left(\nu, S_t/\nu\right), \quad \forall\, t = 1, ..., T \tag{1.42}$$

where $\nu(> n-1)$ are the degrees of freedom, S_t is a PDS scale matrix and it is the conditional expectation of the Realized Covariance matrix (C_t), i.e., the conditional covariance matrix:

$$E(C_t|\mathcal{I}_{t-1}) = S_t \tag{1.43}$$

The conditional covariance matrix, S_t, follows a BEKK type recursion:

$$S_t = CC^{'} + AC_{t-1}A^{'} + BS_{t-1}B^{'} \tag{1.44}$$

where, as for the BEKK model, C is a lower triangular matrix, A and B are order n square matrices. The model is covariance stationary if the eigenvalues of $A^* + B^*$ are less than one in modulus, with $A^* = D_n^+(A \otimes A)D_n$ and $B^* = D_n^+(B \otimes B)D_n$. Then the unconditional mean of C_t exists and it is equal to $\Sigma = (I_n - AA^{'} - B\ B^{'})^{-1}CC^{'}$. The ReBEKK guarantees that S_t is positive definite if so is S_0.

Bauwens, Storti, and Violante (2012) proposed a different dynamics for the scale matrix, S_t, more specifically they assume that it follows a DCC-type dynamic:

$$\begin{aligned}
H_t &= D_t R_t D_t \\
R_t &= diag(Q_t)^{-\frac{1}{2}} Q_t diag(Q_t)^{-\frac{1}{2}} \\
Q_t &= (1-\alpha-\beta)S + \alpha P_{t-1} + \beta Q_{t-1} \\
P_t &= D_t^{-1} C_t D_t^{-1}
\end{aligned} \tag{1.45}$$

where D_t is a diagonal matrix with each diagonal elements, $d_{ii,t}$, equals to the conditional standard deviation of the $i-th$ asset, $h_{i,t}^{1/2}$. The conditional variance of each asset, separately modeled, is specified as a MEM(1,1) process:

$$S_{ii,t} = \omega_i + \alpha_i C_{ii,t-1} + \beta S_{ii,t-1} \tag{1.46}$$

In eq. (1.45), $diag(Q_t)$ is a diagonal matrix whose elements are the diagonal elements of Q_t, where the latter is the so called quasi conditional correlation matrix. S is the unconditional correlation matrix with diagonal elements equal to 1. Finally, by standardisation of Q_t we recover the conditional correlation matrix, R_t.

1.9 Component Multivariate Volatility Models

Variances and covariances share the feature of long-run dependence. Component models, with a parsimonious specification, can capture this long memory behavior, differently from single-component models discussed above. Moreover, there are different sources of volatility with short-run or long-run effects, that a component model can take into account. For this purpose, Colacito, Engle, and Ghysels (2011) proposed the DCC-MIDAS, that is a combination of the DCC of Engle (2002b), the component GARCH of Engle and Lee (1999), and the GARCH-MIDAS of Engle, Ghysels, and Sohn (2013). More specifically, following the decomposition of Σ_t assumed by the DCC, variances are expressed as a GARCH-MIDAS, while in the correlation part, short-lived effects are captured by the DCC autoregressive scheme, where the intercept of the latter is assumed to be time-varying, reflecting long-run movements extracted through the MIDAS filter. Let us introduce two time-scales, that is, t indicates the high-frequency period (a day), while τ indicates the low-frequency period (a month, a quarter, and so on), that is constituted by N_v days. Then the model is specified as follows:

$$
\begin{aligned}
H_t &= D_t R_t D_t \\
D_t &= diag(h_{i,t}) \qquad \forall i = 1,..,n \\
h_{i,t} &= g_{i,t} m_{i,\tau} \\
g_{i,t} &= 1 - \alpha_i - \beta_i + \alpha_i \frac{\epsilon^2_{i,t-1}}{m_{i,\tau}} + \beta_i g_{i,t-1} \\
m_{i,\tau} &= m_i + \theta_i \sum_{k=1}^{K_v} \varphi_k(\lambda_2) RV_{i,\tau-k} \\
RV_{i,\tau} &= \sum_{j=(\tau-1)N_v+1}^{\tau N_v} r^2_{i,j} \\
R_t &= diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \\
Q_t &= (1-\alpha-\beta)P_t + \alpha \xi_{t-1}\xi'_{t-1} + \beta Q_{t-1} \\
P_t &= \sum_{k=1}^{K_c} \varphi_k(\omega_2) c^{(m)}_{t-k} \\
c^{(m)}_t &= diag\left(\sum_{k=t-N_c}^{N_c} \xi^2_k\right)^{-1/2} \left(\sum_{k=t-N_c}^{N_c} \xi_k \xi'_k\right) diag\left(\sum_{k=t-N_c}^{N_c} \xi^2_k\right)^{-1/2}
\end{aligned}
\tag{1.47}
$$

The volatility series, $h_{i,t}$, are modeled via a GARCH-MIDAS, that is a multiplicative component volatility model, in which the short-run component, $g_{i,t}$, follows a unit mean GARCH(1,1) process, while the long-run component $m_{i,\tau}$ is driven by past low frequency Realized Volatilities through the MIDAS filter. The quasi correlation matrix is driven by past degarched residuals, that is $\xi_t = \epsilon_t \oslash h_t^{1/2}$ and contains a time-varying intercept, P_t, that is a weighted average of low frequency standardized degarched residuals[24]. Note that the decay pattern of the weighting function is the same for all the series both in the short-run component and in the long-run component. Let us rewrite the dynamic assumed for the quasi correlation matrix, Q_t:

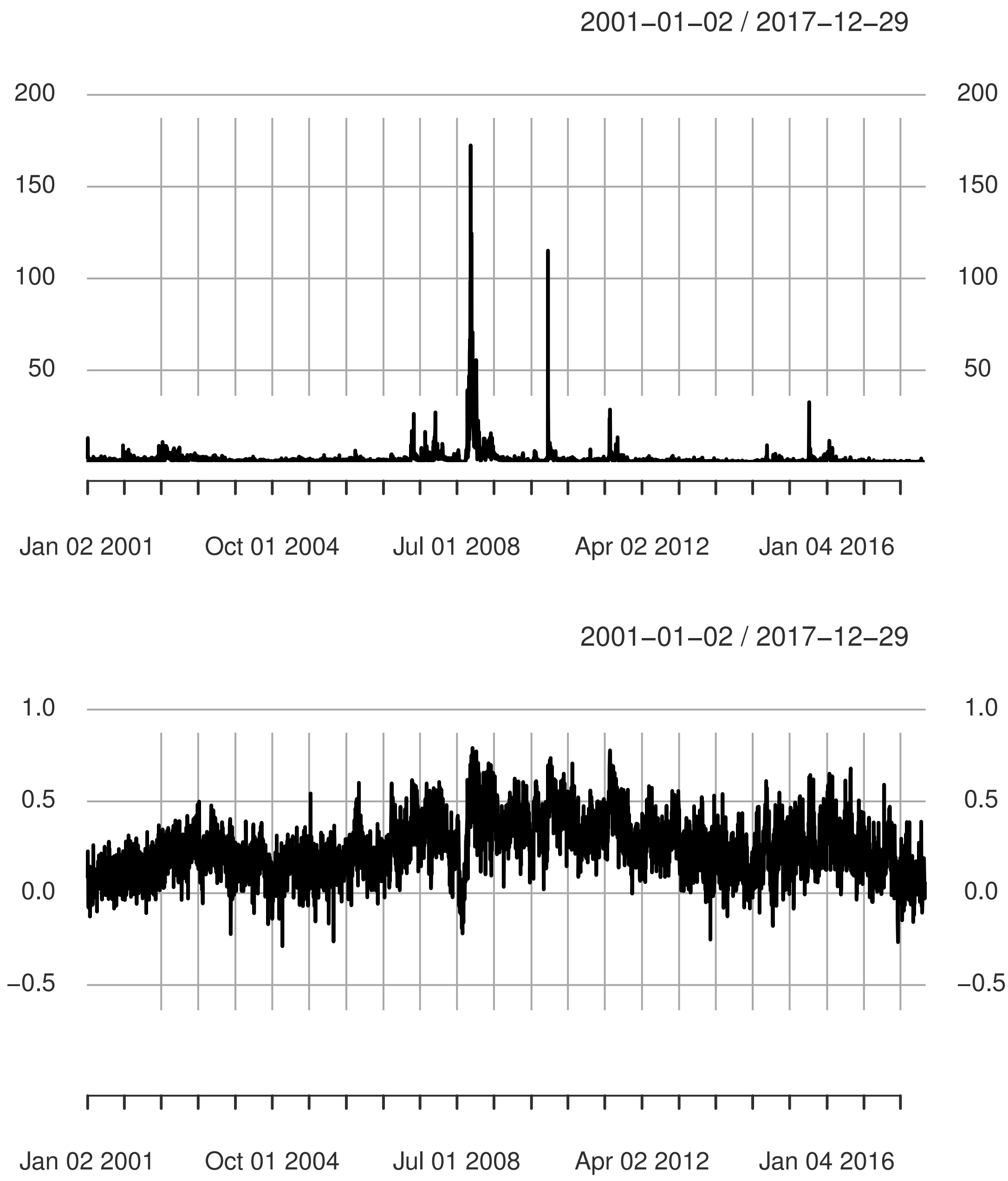
$$
Q_t - P_t = \alpha(\xi_{t-1}\xi'_{t-1} - P_t) + \beta(Q_{t-1} - P_t) \tag{1.48}
$$

From this specification, the fluctuation of Q_t around the long-run intercept, P_t, can be easily understood. Their empirical analysis reveals the better performance, both in-sample and out of sample, of the DCC-MIDAS than the standard DCC.

For what concerns Realized Covariance matrices, Golosnoy, Gribisch, and

[24]Standardization is used to ensure the positive semidefiniteness of P_t, see Engle (2002a).

FIGURE 1.4: Apple & Chevron annualized Realized Covariance (top) and Correlation (bottom). Sample Period: 2 January 2001, 29 December 2017.



Notes: Annualized Realized covariance is expressed in percentage scale.

Liesenfeld (2012) extended their base CAW model, by proposing a component model in which the Scale matrix, S_t, is multiplicatively decomposed into a short-run component, S_t^*, following a BEKK-type recursion, and a long-run component, M_t, specified as a function of past monthly Realized Covariance matrices through the MIDAS filter:

$$
\begin{aligned}
S_t &= M_t^{1/2} S_t^* (M_t^{1/2})^{'} \\
M_t &= M_t^{1/2} (M_t^{1/2})^{'} \\
S_t^* &= (I_n - AA^{'} - BB^{'}) + AC_{t-1}^* A^{'} + BS_{t-1}^* B^{'} \\
C_t^* &= (M_t^{1/2})^{-1} C_t (M_t^{1/2})^{'-1} \\
M_t &= \bar{\Lambda} + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) C_{t-k}^{(m)} \\
C_{t-k}^{(m)} &= \sum_{\tau=t-mk}^{t-m(k-1)-1} C_\tau \\
\varphi_k(\omega_1, \omega_2) &= \frac{(k/K)^{\omega_1 - 1} (1 - k/K)^{\omega_2 - 1}}{\sum_{j=1}^{K} (j/K)^{\omega_1 - 1} (1 - j/K)^{\omega_2 - 1}}
\end{aligned}
\tag{1.49}
$$

where C_t^* is the Realized Covariance matrix purged by its long-run component and $M_t^{1/2}$ is the Cholesky factor of the long-run component. Λ is a constant parameter matrix, that can be decomposed through the Cholesky factor to ensure its positive definiteness, that is $\Lambda = LL^{'}$ with L a lower triangular matrix. $C^{(m)}$ is the monthly Realized Covariance matrix, i.e., the aggregation of daily Realized Covariance matrices for 20 trading days. $\varphi_k(\omega_1, \omega_2)$ is the multivariate MIDAS filter with $\omega_1 = 0$ and $\omega_2 > 1$ to ensure a decreasing decay pattern of the weights. θ is a non-negative scalar that captures the whole effect of $C^{(m)}$, while A and B are two order n square parameter matrices. Notice that differently from the DCC-MIDAS here a unique long-run component is specified for variances and covariances, rather than different ones for variances and correlations, thus preserving parsimony. As we can see from fig. (1.4) the long-run level of covariances is not constant, more specifically it is much higher in a period of market downturns, then allowing it to be time-varying is empirically justified. Their work showed a better in-sample and out of sample performance of the MIDAS extension than the baseline model. Moreover, the persistence of the BEKK component, as usual for the models with the MIDAS filter, is much lower than the base CAW.

Bauwens, Braione, and Storti (2016) proposed a multiplicative component volatility model in which, differently from Golosnoy, Gribisch, and Liesenfeld (2012), the short-run component follows a DCC-type dynamic:

$$\begin{aligned} S_t^* &= D_t^* R_t D_t^* \\ S_{ii,t}^* &= (1-\alpha_i-\beta_i) + \alpha_i C_{ii,t-1} + \beta_i S_{ii,t-1} \\ R_t^* &= (1-\alpha-\beta) I_n + \alpha P_{t-1}^* + \beta R_{t-1}^* \\ P_t^* &= diag(C_t^*)^{-1/2} C_t^* diag(C_t^*)^{-1/2} \end{aligned} \tag{1.50}$$

where D_t^* is a diagonal matrix with each diagonal elements, $d_{ii,t}^*$ equals to the short run conditional standard deviation of the $i-th$ asset $(S_{ii,t}^*)^{1/2}$. The short run conditional variance of each assets, separately modeled, is specified as a MEM(1,1) process, as for the DCC of Bauwens, Storti, and Violante (2012). As for the CAW of Golosnoy, Gribisch, and Liesenfeld (2012), C_t^* is the short run Realized Covariance matrix, while P_t^* is the short run Realized Correlation matrix. They found that the models with a time-varying long-run component outperform their corresponding constant long-run component, both in-sample and out-of-sample.

1.10 Concluding Remarks

In this chapter, we provide a brief literature review of time-varying volatility models, both in a univariate and multivariate framework. More specifically, we examine models in which conditional variance is extracted from daily returns (GARCH), and models based on Realized Volatility measures (MEM). Particular attention is devoted to component models based on MIxed DAta Sampling (MIDAS) filter. Component volatility models can reproduce, in a parsimonious way, the long memory behavior of asset returns, by decomposing conditional variance into a long-run component, representing the changing average level that evolves smoothly, and a short-run component representing the daily fluctuations. The MIDAS filter is a weighting function, that allows us to link variables sampled at different frequencies, then it is possible to measure the influence of the economics on financial volatility.

Nevertheless, the long-run component specified as a function of economic variables through the MIDAS filter, cannot capture abrupt shifts in the level of the series, due to its smooth pattern. This is why in the second chapter, we propose a MIDAS model with a Markovian dynamic added to the short-run component. So, the model can detect discrete changes of the volatility process, that are not taken into account by a single regime model.

For what concerns the multivariate framework, models discussed above apply a decomposition of the conditional covariance matrix that is sensible to the order of the assets. Then, in the third chapter, we propose a multivariate component volatility model, that decomposes the conditional covariance matrix into a short-run component, following a scalar BEKK type recursion, and a long-run one, that is a function of monthly Realized Covariances matrices, through the use of the MIDAS filter. We use a parameterization that is not sensible to the order of the assets and its estimation is easier from a computational point of view. Moreover, we also provide an extension that, with one parameter more than the base model, allows us to have a more flexible dynamic for each covariance: the specification admits asset-pair specific and time-varying ARCH parameters.

Chapter 2

Stock Market Volatility, Macroeconomic Fundamentals and Regime Switching

2.1 Introduction

After the seminal paper of Engle (1982), econometric models on time-varying financial volatility abound in the literature. Whereas, the relationship between macroeconomic variables and volatility has not been exploited enough. Officer (1973) and Schwert (1989) tried to analyse the economic sources of volatility, documenting the so-called countercyclical pattern of stock market volatility, that is, volatility is high during a recession and is low during an expansion phase.

Engle, Ghysels, and Sohn (2013) gave a new pulse to this strand of literature by introducing the GARCH-MIDAS model: a multiplicative component model in which the conditional variance is decomposed into a short-run component and a long-run one. Component models, introduced by Engle and Lee (1999), can capture the long-run dependence of volatility through a parsimonious structure. The GARCH-MIDAS specifies the short-run component as a unit mean GARCH (1,1) process that represents the known volatility clustering and the daily fluctuations. The long-run component, driven by macroeconomic and/or financial variables, represents the average level of volatility, that evolves smoothly. The appeal of this new model is that it allows linking variables sampled at different frequencies, then it is possible to examine directly the linkage between economics and financial volatility.

Within the GARCH framework, volatility is extracted from the daily demeaned squared returns. In the last two decades, researchers, through the availability of ultra-high frequency data, have proposed more precise ex-post daily volatility measures, like Realized Volatility (RV), which is the sum of intraday squared returns. Note that the RV estimator keeps the information provided by intraday returns, while it is excluded by the GARCH models. A drawback of this estimator is that returns sampled at very high frequencies are auto-correlated, due to the microstructure noise. Barndorff-Nielsen et al. (2008) proposed an estimator robust to such market microstructure noise, that is the realized kernel volatility. In addition, RV measures, as for squared returns, share the features of long-run dependence and volatility clustering, thus providing the devolopment of new models, like the Multiplicative Error Model (MEM, Engle, 2002b; Engle and Gallo, 2006). The MEM is a class of time series models for positive valued processes, in which the dependent variable is the product of a time-varying scale factor and a standard positive valued random error. One of the advantages of this specification is to provide the conditional expectation of the variable of interest, rather than the expectation of the log.

Amendola et al. (2020) extended MIDAS volatility models to the MEM framework and proposed the MEM-MIDAS to exploit the relationship between economics and financial volatility by considering the Realized Volatility estimator, rather than the demeaned squared returns as in the GARCH-MIDAS. As pointed out by the authors, the long-run component of the MEM-MIDAS model is not able to capture immediately bursts of volatility and generally abrupt shifts in the average level, due to its smooth pattern.

Then, in the model which we propose (call it MS-MEM-MIDAS), a Markovian dynamic is added to the short-run component to detect abrupt changes in the average level[1]. So, it can capture both smooth and discrete changes in the long-run level of volatility. Furthermore, the MS-MEM-MIDAS model aims to distinguish between the persistence caused by macroeconomic factors, the regime persistence, and the persistence of past values of the realized volatility.

The structure of the chapter is organized as follows: Section (2.2.1) provides a brief review of the different Markov Switching volatility models known in the literature, while Section (2.2.2) presents the model we propose in this Chapter. Section (2.4) analyzes the finite sample properties of the estimator through a Monte Carlo simulation, while in Section (2.3) statistical inference is discussed. Section (2.5) presents the data employed in the empirical analysis (2.5.1), estimation results (2.5.2), an in-sample models comparison (2.5.3) and an out-of-sample exercise (2.5.4). Finally, section (2.6) offers some concluding remarks.

[1] See Pan et al. (2017) in a GARCH framework.

2.2 Theoretical Framework

2.2.1 Markov Switching Volatility models

Lamoureux and Lastrapes (1990) found that shifts in the average level of volatility could be the source of the high persistence implied by a GARCH(1,1) process. For example, it is known that the average level of volatility is higher during turbulent periods than quiet ones. For this purpose, Hamilton and Susmel (1994) proposed a nonlinear process, that is an ARCH(p) model with a time-varying scale parameter. More specifically, their model is regime-dependent, and the switching mechanism is governed by a Markov process[2]. If the regimes are persistent, that is if at time t the process is in regime 1, it is likely that at $t+1$ the process will be in the same regime, then there are at least two sources of volatility persistence: one related to the regimes, while the other one is related to shocks. This is why, in a single regime GARCH(1,1), the persistence due to shocks is overestimated, that is the estimate of $\alpha+\beta$ is close to one. Let us consider the Markov Switching ARCH(p) of Hamilton and Susmel (1994):

$$
\begin{aligned}
&\epsilon_t = \sqrt{g_{s_t} h_t}\eta_t \qquad \eta_t \sim N(0,1) \;\; \forall t \\
&E(\epsilon_t^2 | s_t, \Phi_{t-1}) = g_{s_t} h_t \\
&\Phi_{t-1} = (s_{t-1}, ..., s_{t-p}, \epsilon_t, \epsilon_{t-1}, ..., \epsilon_{t-p}) \\
&h_t = \omega + \sum_{i=1}^{p} \alpha_i \frac{\epsilon_{t-i}^2}{g_{s_{t-i}}}
\end{aligned}
\tag{2.1}
$$

where s_t is a discrete latent random variable representing the regime of the process at time t and it can take on the values on the range $\{1, 2, .., N\}$, with N the number of regimes. The switching scale parameter is g_{s_t}, with $g_1 = 1$ and $g_j \geq 1$ for $j = 2, ..., N$. The dynamic of the latent variable, s_t, is governed by a first order Markov chain, that is:

$$
P\{s_t = j | s_{t-1} = i\} = P\{s_t = j | s_{t-1} = i, s_{t-2}...\} = p_{ij} \tag{2.2}
$$

where the transition probability, p_{ij}, is the probability that at time t the process is in the state j given that at time $t-1$ it was in the state i. Note that only the most recent value of the latent variable, s_{t-1}, influences the current value, s_t. Transition probabilities are collected in an order N square matrix,

[2]For an exhaustive description of Markov switching models see, Hamilton (1994), ch. 22.

called transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}$$

Note that the elements of each column add up to 1, that is $\sum_{j=1}^{N} p_{ij} = 1$.

As said above, the discrete regime variable, s_t, is not observed, then we can only infer in which regime the process is at each point in time t. The forecast depends on the information set we use, indeed if it is based on the observed data up to time $t-1$, we have the predicted probability, $P(s_t = j|\mathcal{I}_{t-1})$, with $\mathcal{I}_{t-1} = \{\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_1\}$, up to time t we have the filtered probabilities $P(s_t = j|\mathcal{I}_t)$, and through the full sample observations we have the smoothed probabilites, $P(s_t = j|\mathcal{I}_T)$[3].

By applying their model to returns on the New York Stock Exchange, they interestingly found that the regimes are very persistent (that is $p_{jj} > 0.9$) while the persistence of shocks within each regime is much lower than single regime models. The main drawback of the Markov Switching ARCH is related to the absence of GARCH effects. This is due to the so-called path dependence problem, that is, in a GARCH(1,1) process the conditional variance depends on the whole history of the latent variable s_t. Indeed, a regime-switching GARCH(1,1) is specified as follows:

$$h_{t,s_t} = \omega_{s_t} + \alpha\epsilon_{t-1}^2 + \beta h_{t-1,s_{t-1}} \tag{2.3}$$

Notice that h_{t,s_t} depends on the current regime through the constant, ω_{s_t}, and on the previous regime through $h_{t-1,s_{t-1}}$. In turn $h_{t-1,s_{t-1}}$ depends on s_{t-1} through $\omega_{s_{t-1}}$, and on s_{t-2} through $h_{t-2,s_{t-2}}$, and so on until h_{1,s_1}. In a few words the conditional variance at time t depends on the whole history of the latent variable up to time t, $\{s_t, s_{t-1}, s_{t-2}, ..., s_1\}$, thus rendering the estimation of the parameters infeasible. Gray (1996) proposed a solution, by aggregating the regime dependent conditional variance at each t, so the

[3] All the probabilities discussed above are calculated through the Hamilton filter and the Kim one, explained in appendix C.

variance depends only on the current regime:

$$\begin{aligned} h_{t,s_t} &= \omega_{s_{t-1}} + \alpha\epsilon_{t-1}^2 + \beta h_{t-1} \\ h_{t-1} &= p_{t,1}h_{t-1,s_1} + (1-p_{t,1})h_{t-1,s_2} \\ p_{t,1} &= P(s_t = 1|\mathcal{I}_{t-1}) \\ \mathcal{I}_{t-1} &= \epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_1 \end{aligned} \tag{2.4}$$

where $p_{t,1}$ is the predicted probability of the process to be in the state 1 at time t conditionally at the information set at time $t-1$. Importantly, notice that h_{t,s_t} depends only on the current regime because, through the aggregation, h_{t-1} is no more path dependent. Differently from Hamilton and Susmel (1994), he allows the ARCH and the GARCH coefficients to be time-varying, not only the level of volatility. Thus, by applying a 2 regime-switching GARCH model to the US short term interest rate, his results show that in the high volatility regime, conditional variance is more sensitive to recent shocks ($a_2 > a_1$), while the inertial effect is lower than low volatility regime ($b_1 > b_2$). This is an interesting feature that a single regime GARCH model is not able to capture. To circumvent the path dependence problem, Dueker (1997) proposed another solution in the spirit of the collapsing procedure introduced by Kim (1994)[4], then as in Gray (1996), the estimation of the parameters is feasible. The results show that the model he proposed has a better performance than the simple GARCH(1,1) and the Markov Switching ARCH of Hamilton and Susmel (1994).

A third solution to the path dependence problem was offered by Haas, Mittnik, and Paolella (2004), who specified a model which preserves the economic interpretation according to which shocks drive volatility[5]:

$$h_{j,t} = \omega_j + \alpha_j\epsilon_{t-1}^2 + \beta_j h_{j,t-1} \qquad j = 1, ..., N \tag{2.5}$$

Notice that the above parameterization avoids the path dependence problem in that $h_{j,t}$ is separately specified for each regime. Moreover, in all the other regime-switching models discussed above, there is not a straightforward correspondence between the variance of each regime and the related parameters, because each state-dependent variance is affected by the other component variances and potentially could fluctuate as the mixing probabilities vary. Due to the analytical tractability of the model, differently from

[4] See below for an explanation of the collapsing procedure proposed by Kim.

[5] If you consider the ARCH(∞) representation of a GARCH(1,1) process it is clear the role of shocks in driving the volatility process.

the other ones, the dynamic properties and stationarity conditions are derived. Their results confirm the ability of the model to adequately capture the volatility pattern.

Within the Realized Volatility paradigm, Gallo and Otranto (2015) extended the MEM (Engle, 2002b; Engle & Gallo, 2006) to the Markov Switching framework, thus allowing the model to capture shifts in the average level of variance and/or different dynamics in the series of the process. Then, their model, the MS-AMEM is specified as follows:

$$\begin{aligned} x_t &= \mu_{t,s_t}\epsilon_t \\ \epsilon_t|s_t &\sim Gamma\left(a_{s_t}, \frac{1}{a_{s_t}}\right) \ \forall t \\ \mu_{t,s_t} &= \omega_{s_t} + \alpha x_{t-1} + \beta\mu_{t-1,s_{t-1}} + \gamma\mathbb{1}_{(r_{t-1}<0)}x_{t-1} \end{aligned} \tag{2.6}$$

where ω_{s_t} is the switching constant parameter, with $\omega_j \geq \omega_{j-1}$ for $j = 2, 3, ..., N$, so that the level of volatility increases through the states. Note that only the constant is state-dependent, because we want to detect only shifts in the average level. The latter, within each regime j, with $j = 1, .., N$ is equal to:

$$\mu_j = \frac{\omega_j}{1-\alpha-\beta-\gamma/2} \tag{2.7}$$

As for the regime-switching GARCH, the conditional variance, μ_{t,s_t}, depends on the whole history of the latent variable, s_t. To circumvent this path dependence problem, they adopt the following collapsing procedure proposed by Kim (1994):

$$\hat{\mu}_{t,s_t} = \frac{\sum_{i=1}^{N} P\{s_t = j, s_{t-1} = i|\mathcal{I}_t\}\hat{\mu}_{t,s_t,s_{t-1}}}{P\{s_t = j|\mathcal{I}_t\}} \tag{2.8}$$

where $P\{s_t = j, s_{t-1} = i|\mathcal{I}_t\}$ and $P\{s_t = j|\mathcal{I}_t\}$ are the filtered probabilities. At each time t, the N^2 possible values of μ_t (depending on s_t and s_{t-1}) are collapsed into N values (depending on s_t only) by a weighted average, thus obtaining an approximation of the conditional variance at time t, $\hat{\mu}_{t,s_t}$.

2.2.2 The Markov Switching MEM MIDAS

A drawback of the MEM-MIDAS model is that the combination of the long-run component, τ_t, which represents the influence of the real economy, and the short-run component $g_{i,t}$, representing the daily fluctuations, are not sufficient to describe the pattern of financial volatility. Indeed, as we can see at the top-right of fig. (2.2), we have a long-run component, according to Amendola et al. (2020), that lags behind volatility bursts, or generally is not able to capture peaks of volatility. Then, similarly to Pan et al. (2017) in a GARCH context, we propose to add a markovian dynamic to the short-run component, to detect abrupt shifts in the average level of volatility. So, the model we propose takes into account both macroeconomic sources of volatility (through the MIDAS structure) and discrete shifts in the average level (through state-dependent parameters). Let $x_{i,t}$ be the realized volatility, and $s_{i,t}$ a discrete-time latent variable of the *i-th* day of the low-frequency period t (a week, a month, a quarter, etc, with N_t the number of days for that period) then the Markov Switching MEM-MIDAS model (MS-MEM-MIDAS) is specified as follows:

$$\begin{aligned}
x_{i,t} &= \mu_{i,t,s_{i,t}}\epsilon_{i,t} = g^*_{i,t,s_{i,t}}\tau^*_t\epsilon_{i,t} \\
\epsilon_{i,t}|s_{i,t} &\sim Gamma\left(a_{s_{i,t}}, 1/a_{s_{i,t}}\right) \forall\, i = 1, ..., N_t \\
g^*_{i,t,s_{i,t}} &= \omega_{s_{i,t}} + \alpha\frac{x_{i-1,t}}{\tau^*_t} + \beta g_{i-1,t,s_{i-1,t}} + \gamma\mathbb{1}_{(r_{i-1,t}<0)}\frac{x_{i-1,t}}{\tau^*_t} \\
\tau^*_t &= exp\left\{\theta\sum_{k=1}^{K}\varphi_k(\lambda_1,\lambda_2)X_{t-k}\right\} \\
\varphi_k(\lambda_1,\lambda_2) &= \frac{(k/K)^{\lambda_1-1}(1-k/K)^{\lambda_2-1}}{\sum_{j=1}^{K}(j/K)^{\lambda_1-1}(1-j/K)^{\lambda_2-1}}
\end{aligned} \tag{2.9}$$

where $g^*_{i,t,s_{i,t}}$ is the fast-moving volatility component[6] with a MS-MEM type structure and τ^*_t is the slow-moving one, driven by a low frequency stationary variable, X_t. We use the exponential form to ensure its positivity, because the low-frequency variable can assume both positive and negative values. The MIDAS filter is $\varphi_k(\lambda_1,\lambda_2)$, a weighting function of the past K values of X_t, with the weights that sum up to one. This filter, based on the beta function, is called beta polynomial, and is quite flexible, allowing us to link variables sampled at different frequencies. If $\lambda_1 = 1$ and $\lambda_2 > 1$, the function is monotonically decreasing, as fast as λ_2 increases, that is the most recent observations have more influence on the long-run component.

Notice that the fast-moving component exhibits the path dependence problem discussed above, then in the spirit of Gallo and Otranto (2015) we use the following collapsing procedure:

$$\hat{g}^*_{i,t,s_{i,t}} = \frac{\sum_{l=1}^{N}P\{s_{i,t}=j, s_{i-1,t}=l|\mathcal{I}_{i,t}\}\hat{g}^*_{i,t,s_{i,t},s_{i-1,t}}}{P\{s_{i,t}=j|\mathcal{I}_{i,t}\}} \tag{2.10}$$

Then, by averaging the N^2 possible values of the fast-moving component

[6]Notice that when $i = 1$, $(i-1,t) = (N_{t-1}, t-1)$.

$g^*_{i,t,s_{i,t}}$, with the weights equal to the corresponding filtered probabilities, the model is no more path-dependent. Other approximations of the likelihood have been proposed in the literature, like the deterministic particle filter of Augustyniak, Boudreault, and Morales (2018). Nevertheless, as showed by the Monte Carlo experiment discussed below[7], the approach we use in the approximation of the likelihood is satisfactory and the bias is very small as the length of the series increases.

With respect to the MEM-MIDAS model, the specification of the volatility components is slightly different[8]: indeed, the regime dependent constant appears in the fast-moving component to take into account abrupt shifts in the level of volatility, while it is removed from the slow-moving component, otherwise, there would be a parameter identification problem. Only the parameters $\omega_{s_{i,t}}$ and $a_{s_{i,t}}$ are state-dependent, due to non-convergence problem in the estimation procedure caused by the increasing number of parameters (Guérin & Marcellino, 2013; Pan et al., 2017).

With the parameterization proposed in eq. (2.9), the slow-moving component loses its interpretation of a trend around which conditional volatility fluctuates. Moreover, models with regime-switching provides the long-run component, τ_t, as the average level of volatility of the inferred regime, through the smoothed probabilities, $P\{s_{i,t} = j|\mathcal{I}_{N_T,T}\}$, as in Gallo and Otranto (2015). Keeping it in mind, to recover the interpretation of the long-run component as the average level of volatility, we employ, differently to Pan et al. (2017), the following ex-post transformation:

$$\begin{aligned} \tau_{i,t} &= \tau^*_t \frac{\omega_{s_{i,t}}}{1-\alpha-\beta-\gamma/2} \\ g_{i,t,s_{i,t}} &= \frac{g^*_{i,t,s_{i,t}}}{\frac{\omega_{s_{i,t}}}{1-\alpha-\beta-\gamma/2}} \end{aligned} \tag{2.11}$$

where the imputation of the regime at each point in time, t, is based on the value of the smoothed probabilities. Now, the new fast moving component, $g_{i,t,s_{i,t}}$, is the ratio of $g^*_{i,t,s_{i,t}}$ and its expected value, $E(g^*_{i,t,s_{i,t}})$, then its unconditional value is equal to 1, that is, $E(g_{i,t,s_{i,t}}) = 1$[9], like the MEM-MIDAS model. At the same time, multiplying τ_t by the expected value of the fast moving

[7]See sec. (2.4).

[8]See Pan et al. (2017).

[9]Note that $E(g^*_{i,t}|s_{i,t}) = \frac{\omega_{s_{i,t}}}{1-\alpha-\beta-\gamma/2}$. See Timmermann (2000) for the moments of Markov Switching models.

component, $E(g^*_{i,t,s_{i,t}})$, we obtain the long-run component, $\tau_{i,t}$, representing the time varying average level of volatility[10].

2.3 Quasi Maximum likelihood

The approximate log-likelihood function is a by-product of the Kim's filter[11] and it is specified as follows:

$$\begin{aligned}
LL(\Xi) &= \sum_{t=1}^{T}\sum_{i=1}^{N_t} ll_{i,t} = \sum_{t=1}^{T}\sum_{i=1}^{N_t} f(x_{i,t}|\mathcal{I}_{i-1,t};\Xi) \\
f(x_{i,t}|\mathcal{I}_{i-1,t};\Xi) &= \sum_{j=1}^{N} P(s_{i,t}=j|\mathcal{I}_{i-1,t};\Xi) f(x_{i,t}|\mathcal{I}_{i-1,t}, s_{i,t}=j;\Xi) \\
f(x_{i,t}|\mathcal{I}_{i-1,t}, s_{i,t}=j;\Xi) &= \left[\frac{a_{s_{i,t}}{}^{a_{s_{i,t}}} \mu_{t,s_{i,t}}^{-a_{s_{i,t}}} x_t^{a_{s_{i,t}}-1} \exp\left(-a_{s_{i,t}}\frac{x_{i,t}}{\mu_{i,t,s_{i,t}}}\right)}{\Gamma(a_{s_{i,t}})}\right]
\end{aligned} \tag{2.12}$$

where Ξ is the vector of parameters to be estimated, while $\Gamma(\cdot)$ is the gamma function.

Within the regime-switching framework, the parameters are estimated through the Quasi Maximum Likelihood estimation (QMLE) method. Let us consider the single regime AMEM: note that, if x_t is assumed to follow a Gamma distribution, as in the third line of eq. (2.12), the average score is equal to:

$$\bar{s}_t = aT^{-1}\sum_{t=1}^{T}\left(\frac{x_t-\mu_t}{\mu_t^2}\right)\nabla_\Xi \mu_t \tag{2.13}$$

That is, the expected value of the score, if the conditional expectation is correctly specified, is equal to 0 irrespective of the value of the Gamma parameter, a. Then, the likelihood function has a quasi-likelihood interpretation and $\hat{\Xi}$ is the QML estimator[12]. Consequently, the standard errors of the estimated coefficients are calculated through the sandwich estimator of the variance-covariance matrix, that is the latter is estimated as follows [13]:

$$Var(\hat{\Xi}) = H^{-1}OPH^{-1} \tag{2.14}$$

[10] Note that $\tau_{i,t}$, differently from τ_t, can change within the low frequency period, due to the dependence on the regime.

[11] See Kim (1994) and Kim and Nelson (1999) for an exhaustive description of Kim's filter.

[12] See Brownlees, Cipollini, and Gallo (2012).

[13] See Bollerslev and Wooldridge (1992) and White (1982).

where H^{-1} is the inverse of the Hessian matrix with the opposite sign and OP is the matrix of the Outer Product of the scores, that is:

$$\begin{aligned} H &= -\frac{\partial^2 LL(\Xi)}{\partial\Xi\partial\Xi'} \\ OP &= \sum_{t=1}^{T}\sum_{i=1}^{N_t} \frac{\partial ll_{i,t}(\Xi)}{\partial\Xi\partial\Xi'}\frac{\partial ll_{i,t}(\Xi)}{\partial\Xi\partial\Xi'}' \end{aligned} \tag{2.15}$$

Empirically, the model parameters are estimated numerically on the software R through a personal code[14].

2.4 Monte Carlo Simulation

In this section, we analyze the finite sample properties of the QML estimator of the MS-MEM MIDAS through a Monte Carlo exercise. This experiment is useful to evaluate its performance as the sample size increases. Then, we generate the data from the MS-MEM-MIDAS model for 600 series of length $T = \{1500, 2100, 2700, 3300, 9900\}$.

The true value of the parameters is that obtained in the empirical analysis of the MS-MEM-MIDAS with 2 regimes[15]. The low-frequency variable, X_t, is assumed to follow a gaussian AR(4) process, while the daily return a gaussian AR(2) one, and the error terms of eq. (2.9) a state-dependent Gamma distribution with unit mean. The number of days, N_t, for each month is equal to 30, then the low-frequency variable is constant for this period. In the long-run component, as in the empirical section, the number of lags selected is equal to 36. Table (2.1) tells us that, generally, the root mean squared error of the estimates decreases as the length of the series increases and the estimator distribution approaches gaussian distribution. The switching constants and the parameters of the Gamma distributions exhibit a small bias, while the estimators of other parameters are, apparently, unbiased. Then, the degree of approximation of the likelihood is acceptable. Overall, the results we found are in line with the Monte Carlo experiment of Gallo and Otranto (2015) for the MS-MEM model.

14 https://www.r-project.org/.

15 See section (2.5.2).

TABLE 2.1: Monte Carlo simulation of the MS MEM MIDAS with 2 regimes

	True	T=1500			T=2100			T=2700			T=3300			T=9900		
		Average	RMSE	Normality	Average	RMSE	Normality	Average	RMSE	Normality	Average	RMSE	Normality	Average	RMSE	Normality
ω_0	0.585	0.606	0.074	0.022	0.599	0.062	0.073	0.597	0.055	0.973	0.602	0.049	0.529	0.598	0.03	0.164
ω_1	1.677	1.778	0.242	0.032	1.769	0.215	0.000	1.757	0.168	0.322	1.756	0.151	0.293	1.722	0.088	0.884
α	0.051	0.046	0.015	0.356	0.049	0.012	0.511	0.048	0.01	0.493	0.050	0.009	0.918	0.051	0.005	0.787
β	0.836	0.839	0.017	0.699	0.837	0.013	0.989	0.838	0.011	0.255	0.836	0.011	0.955	0.835	0.006	0.82
γ	0.124	0.125	0.009	0.924	0.125	0.008	0.816	0.125	0.007	0.923	0.124	0.006	0.753	0.125	0.004	0.958
θ	-0.177	-0.177	0.009	0.892	-0.177	0.007	0.827	-0.177	0.006	0.767	-0.177	0.006	0.304	-0.177	0.003	0.978
a_1	7.677	7.732	0.256	0.463	7.712	0.219	0.853	7.689	0.187	0.999	7.673	0.162	0.999	7.666	0.098	0.587
a_2	7.990	8.321	0.992	0.000	8.154	0.781	0.000	8.026	0.631	0.357	8.064	0.577	0.019	7.950	0.296	0.129
p_{11}	0.994	0.992	0.003	0.000	0.993	0.003	0.000	0.993	0.002	0.011	0.993	0.002	0.087	0.994	0.001	0.048
p_{22}	0.956	0.942	0.026	0.000	0.946	0.022	0.000	0.949	0.017	0.000	0.951	0.015	0.000	0.955	0.007	0.030

Notes: 600 series of different length, $T = \{1500, 2100, 2700, 3300, 9900\}$. Average estimates, root mean squared errors (RMSE), and p-value of the Kolmogorov Smirnov test. We do not report λ_2 because it has a RMSE equal to 0, then it follows a degenerating distribution.

TABLE 2.2: Descriptive Statistics of annualized Realized kernel volatility, log returns and Industrial Production growth rate

	Real. Vol.	**Ret.**	**Ind. Prod.**
Mean	12.593	0.174	0.223
Median	10.102	0.862	0.394
Min	1.506	-148.444	-15.023
Max	113.455	162.241	5.255
St. Dev.	9.515	17.458	2.320
Skewness	3.239	-0.199	-1.861
Kurtosis	17.691	9.376	9.190

Notes: The table reports the Mean, the Median, the Minimum (Min), the Maximum (Max), the Standard Deviation (St. Dev.), the Skewness and the Kurtosis. All the variables are expressed in annualized percentage terms. Sample period: 2 January 2002, 31 December 2019.

2.5 Empirical Analysis

2.5.1 Data

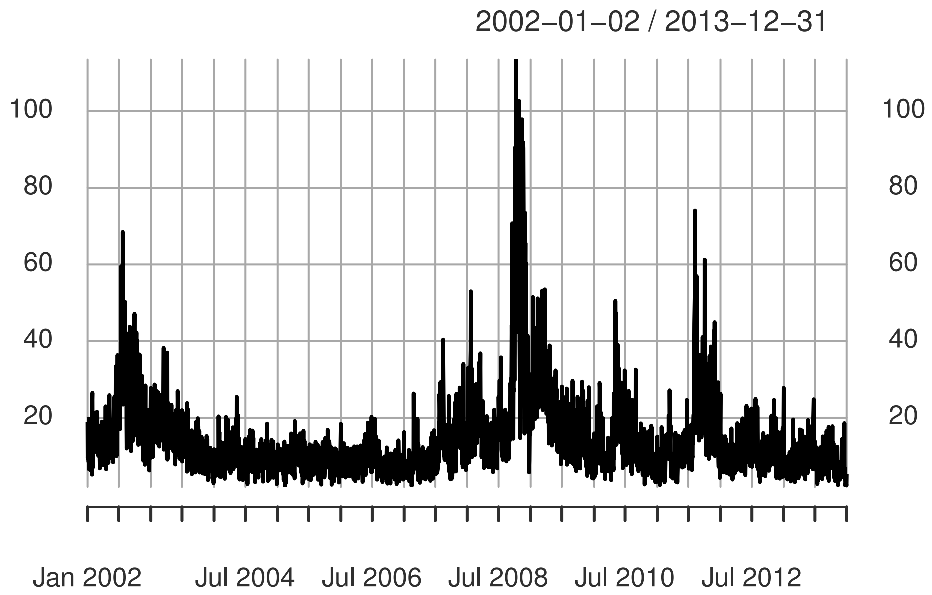
The high-frequency variable is the daily realized kernel volatility[16] of the S&P 500 index, expressed in percentage annualized scale (that is the square root of the realized kernel variance multiplied by $100\sqrt{252}$). Whereas, the low-frequency variable is the Industrial Production growth rate[17], available with a monthly frequency, expressed in percentage annualized change (that is the relative change multiplied by $100\sqrt{12}$). The full sample period covers the interval between 2 January 2002 and 31 December 2019 with a number of 3055 daily observations. In table (2.2), we report some descriptive statistics of the variables used in the empirical analysis. The negative skewness and the high kurtosis of returns reflect some of the stylized facts, that is: large

[16] Data are taken from Oxford-Man institute's Realized Library (https://realized.oxford-man.ox.ac.uk/data/download).

[17] Data are available at Federal Reserve Economic Data data service (https://fred.stlouisfed.org/series/INDPRO).

negative returns are more probable than large positive ones and the returns unconditional distribution has fatter tails than the normal one.

FIGURE 2.1: S&P 500 index Realized Volatility. Sample Period: 2 January 2002, 31 December 2013.



Notes: Annualized Realized volatility is expressed in percentage scale.

Figure (2.1) shows the behavior of the S&P 500 realized volatility for the whole sample period. It is so evident the known volatility clustering (another stylized fact of asset returns) and the discrete shifts in the average level, during some periods of market downturns, that a Markov switching model should be able to capture.

2.5.2 Estimation Results

In the table (2.4), I report the parameter estimation of the models we employ in the empirical analysis. The first in-sample period covers the interval between January 2002 and December 2013.

The parameter estimates for the AMEM are in line with those of the previous studies: the persistence implied by the model is very high ($\alpha + \beta + \gamma/2 \approx 0.96$), the coefficient γ is statistically significant, thus confirming the asymmetric effect, that is negative returns tend to increase future volatility more than the positive ones. The unconditional variance implied by the models is equal to 13.20.

TABLE 2.3: Model Specifications

Model	Functional Form
MEM	$x_t = \mu_t \epsilon_t$ $\mu_t = \omega + \alpha x_{t-1} + \beta \mu_{t-1} + \gamma \mathbb{1}_{(r_{t-1}<0)} x_{t-1}$
MEM-MIDAS	$x_{i,t} = g_{i,t} \tau_t \epsilon_{i,t} = \mu_{i,t} \epsilon_{i,t}$ $g_{i,t} = 1 - \alpha - \beta - \gamma/2 + \alpha \frac{x_{i-1,t}}{\tau_t} + \beta g_{i-1,t} + \gamma \mathbb{1}_{(r_{i-1,t}<0)} \frac{x_{i-1,t}}{\tau_t}$ $\tau_t = exp\left\{\theta \sum_{k=1}^{K} \varphi_k(\lambda_1, \lambda_2) X_{t-k}\right\}$
MS(2) MEM	$x_t = \mu_{t,s_t} \epsilon_t$ $\mu_{t,s_t} = \omega_{s_t} + \alpha x_{t-1} + \beta \mu_{t-1,s_{t-1}} + \gamma \mathbb{1}_{(r_{t-1}<0)} x_{t-1}$ $P\{s_t = j \mid s_{t-1} = i\} = p_{ij}$, with $i, j = 1, 2$
MS(2) MEM-MIDAS	$x_{i,t} = \mu_{i,t,s_{i,t}} \epsilon_{i,t} = g^*_{i,t,s_{i,t}} \tau^*_t \epsilon_{i,t}$ $g^*_{i,t,s_{i,t}} = \omega_{s_{i,t}} + \alpha \frac{x_{i-1,t}}{\tau^*_t} + \beta g_{i-1,t,s_{i-1,t}} + \gamma \mathbb{1}_{(r_{i-1,t}<0)} \frac{x_{i-1,t}}{\tau^*_t}$ $\tau^*_t = exp\left\{\theta \sum_{k=1}^{K} \varphi_k(\lambda_1, \lambda_2) X_{t-k}\right\}$ $P\{s_t = j \mid s_{t-1} = i\} = p_{ij}$, with $i, j = 1, 2$
MS(3) MEM	$x_t = \mu_{t,s_t} \epsilon_t$ $\mu_{t,s_t} = \omega_{s_t} + \alpha x_{t-1} + \beta \mu_{t-1,s_{t-1}} + \gamma \mathbb{1}_{(r_{t-1}<0)} x_{t-1}$ $P\{s_t = j \mid s_{t-1} = i\} = p_{ij}$, with $i, j = 1, 2, 3$
MS(3) MEM-MIDAS	$x_{i,t} = \mu_{i,t,s_{i,t}} \epsilon_{i,t} = g^*_{i,t,s_{i,t}} \tau^*_t \epsilon_{i,t}$ $g^*_{i,t,s_{i,t}} = \omega_{s_{i,t}} + \alpha \frac{x_{i-1,t}}{\tau^*_t} + \beta g_{i-1,t,s_{i-1,t}} + \gamma \mathbb{1}_{(r_{i-1,t}<0)} \frac{x_{i-1,t}}{\tau^*_t}$ $\tau^*_t = exp\left\{\theta \sum_{k=1}^{K} \varphi_k(\lambda_1, \lambda_2) X_{t-k}\right\}$ $P\{s_t = j \mid s_{t-1} = i\} = p_{ij}$, with $i, j = 1, 2, 3$

Notes: The table reports the functional form for the MEM, MEM-MIDAS, MS(2) MEM, MS(2) MEM-MIDAS, MS(3) MEM, and MS(3) MEM-MIDAS specifications.

For the MEM-MIDAS, the θ coefficient is negative, that is a deterioration in economic conditions increases long-term volatility, while the value of the MIDAS filter parameter, λ_2, is larger than one, so the most recent observations of the low-frequency variable have more influence on the long-run component. To get a deeper insight into the impact of macroeconomic variables on the long-run component, it is useful to calculate its Relative Marginal effect (RME):

$$RME = \exp\left[\theta \cdot \varphi_k(\lambda_2) \cdot \Delta X_{t-k}\right] - 1 \tag{2.16}$$

Then, with $\theta = -0.208$ and $\varphi_1 = 0.107$, a unit decrease in IP growth rate will increase next quarter long-run component by $2,25\%$. Furthermore, a unit decrease in the previous quarter of IP, with $\varphi_2 = 0.098$, would rise two quarters ahead long-run component by $2,06\%$ with a cumulative effect after two quarter of a unit decrease of about $4,31\%$. Furthermore, let us analyze the behaviour of τ_t: it is possible to define the weighted sum, at time t, of the lagged macroeconomic variables as $WSUM_t = \sum_{k=1}^{K} \varphi_k(\lambda_1, \lambda_2) X_{t-k}$ and that at time $t-1$ as $WSUM_{t-1} = \sum_{k=2}^{K+1} \varphi_k(\lambda_1, \lambda_2) X_{t-k}$. What matters for an increase or decrease in τ_t is the difference between $WSUM_t$ and $WSUM_{t-1}$, that is a large recent negative value does not imply per se an increase in

TABLE 2.4: Estimation results of six MEM based models for Realized Volatility of S&P 500

	MEM	MEM-MIDAS	MS(2) MEM	MS(2) MEM-MIDAS	MS(3)MEM	MS(3)MEM-MIDAS
w_1	0.514^{***}	0.589^{***}	0.281^{***}	0.585^{***}	0.699^{***}	1.037^{***}
	(0.105)	(0.093)	(0.053)	(0.178)	(0.111)	(0.155)
w_2			1.829^{***}	1.677^{***}	1.046^{***}	1.516^{***}
			(0.368)	(0.264)	(0.198)	(0.369)
w_3					2.46^{***}	3.26^{***}
					(0.354)	(0.526)
α	0.112^{***}	0.098^{***}	0.069^{***}	0.051^{***}	0.043^{***}	0.014
	(0.016)	(0.014)	(0.013)	(0.014)	(0.013)	(0.025)
β	0.790^{***}	0.798^{***}	0.852^{***}	0.836^{***}	0.814^{***}	0.814^{***}
	(0.016)	(0.016)	(0.014)	(0.028)	(0.022)	(0.021)
γ	0.119^{***}	0.123^{***}	0.107^{***}	0.124^{***}	0.138^{***}	0.143^{***}
	(0.011)	(0.011)	(0.01)	(0.011)	(0.013)	(0.014)
θ		-0.208^{***}		-0.177^{***}		-0.187^{***}
		(0.02)		(0.023)		(0.023)
λ_2		4.069^{***}		4.452^{***}		3.148^{***}
		(1.143)		(1.540)		(0.721)
a_1	7.126^{***}	7.290^{***}	7.647^{***}	7.677^{***}	7.562^{***}	7.591^{***}
	(0.235)	(0.207)	(0.208)	(0.227)	(0.263)	(0.315)
a_2			6.694^{***}	7.990^{***}	7.804^{***}	8.402^{***}
			(1.588)	(0.817)	(0.419)	(0.560)
a_3					8.891^{***}	8.627^{***}
					(0.955)	(0.893)
p_{11}			0.993^{***}	0.994^{***}	0.998^{***}	0.994^{***}
			(0.004)	(0.004)	(0.001)	(0.007)
p_{12}					0.001	0.006
					(0.001)	(0.007)
p_{13}					0.001	0.000
					(0.001)	(0.000)
p_{21}					0.003	0.01
					(0.002)	(0.015)
p_{22}			0.848^{***}	0.956^{***}	0.984^{***}	0.978^{***}
			(0.080)	(0.040)	(0.008)	(0.021)
p_{23}					0.013^{*}	0.012^{**}
					(0.008)	(0.005)
p_{31}					0.000	0.000
					(0.000)	(0.000)
p_{32}					0.041	0.04^{**}
					(0.029)	(0.016)
p_{33}					0.959^{***}	0.960^{***}
					(0.029)	(0.016)

Notes: For the specification of the models see tab. (2.3). Dependent variable: annualized Realized kernel Volatility. Low frequency variable: annualized Industrial Production growth rate. Number of lagged annualized Industrial Production growth rates: k=36. First in-sample period: January 2002, December 2013. Daily observations: 3055. Monthly observations of the exogenous variable: 144. P-value: $*** < 0.01$, $** < 0.05$, $* < 0.1$. In brackets, standard errors based on sandwich matrix.

the long run component. For what concerns Markov Switching models, as Gallo and Otranto (2015), we consider three regimes, but, to make a comparison, we also estimate the model with a number of states equal to two[18]. The results of the MS(2)-MEM are in line with those of the previous studies, with the two volatility regimes highly persistent. The expected duration in each regime (calculated as $\frac{1}{1-p_{jj}}$) is about 139 days in the low volatility state, while it is about 7 days for the high volatility state. By inspecting the graph of Realized Volatility (fig. 2.1) it could be interesting to consider an additional regime that can capture periods of very high volatility. Indeed, the persistence of each regime is very high, that is p_{jj} for $j = 1, 2, 3$ close to 1, so the model captures this source of volatility persistence. Consequently,

[18]See Gallo and Otranto (2015) for a description of the procedure they apply to choose the number of regimes.

the persistence of shocks implied by the MS(3) MEM is much lower than the other estimated models ($\alpha + \beta + \gamma/2 \approx 0.922$). Notice that, while in all the regimes close to 1, the transition probabilities of remaining in the same regime, p_{jj}, are decreasing as we pass from lower volatility regimes to the higher ones. It is a consolidated result in the literature that turmoil periods have a shorter duration than quiet ones. In addition, the average level[19] of low, medium and high volatility regime, are respectively equal to 9.45, 14.14, and 33.24, with the average level of regime 3 more than three times that of the regime 1. The models that combine the MIDAS filter with the markovian dynamic in the short run component confirm the results exhibited by the single regime counterpart, that is the parameters enter the model with the expected sign and magnitude. Moreover, estimated component models exhibit a lower level of persistence due to shocks, a recurrent feature of component models[20]. Indeed, the sum of α, β, and $\gamma/2$ is lower than the corresponding one-component model. Finally, if you look at the estimated switching probabilities, it is interesting to notice that, in the Markov switching model with 3 states, we have an interaction between medium and high volatility regime, while the low volatility state, being the most persistent, shows a very low probability of switching to the other ones.

2.5.3 Model comparison

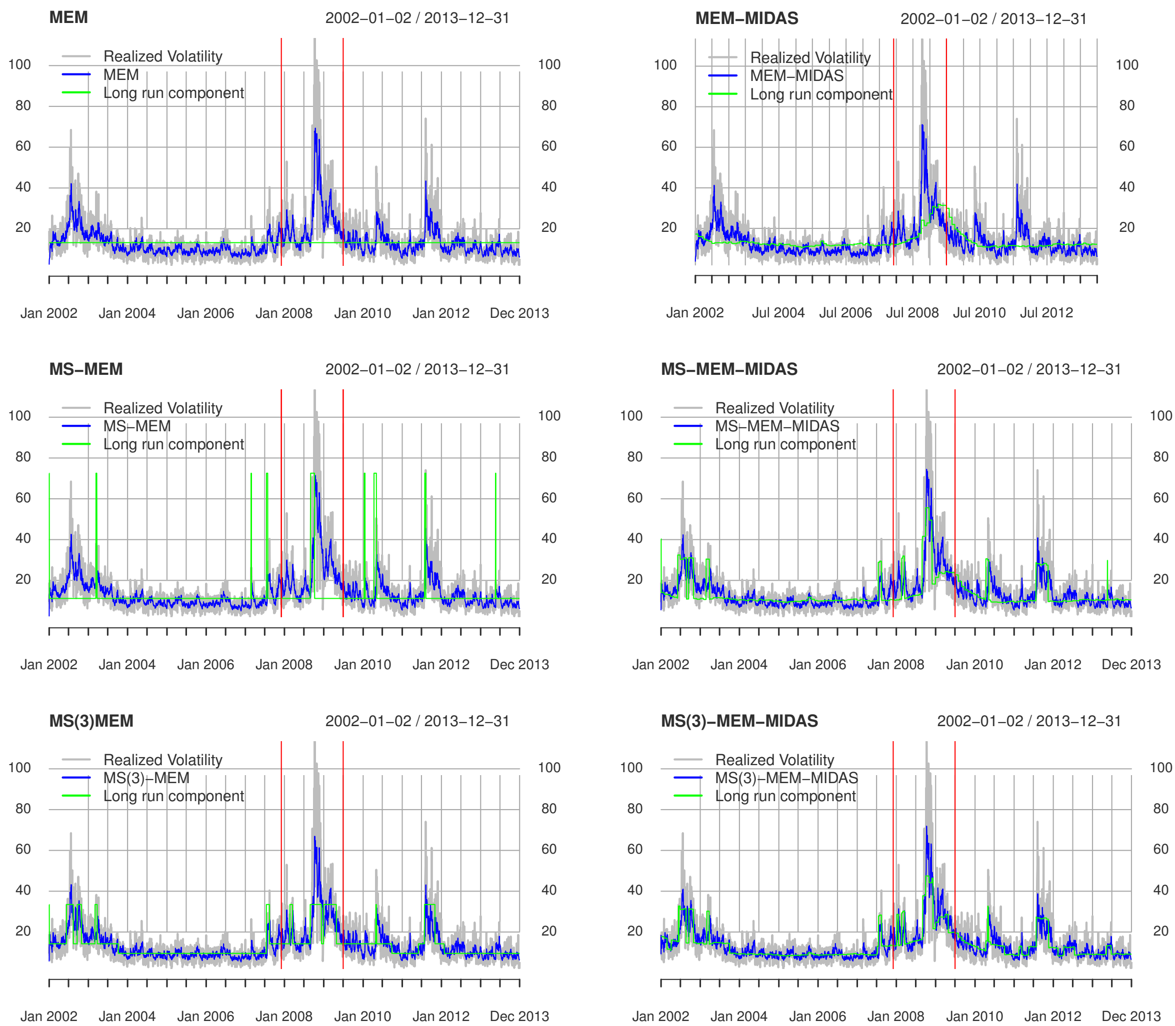
As we can see from fig. (2.2), it seems that the MS-MEM MIDAS model overcomes the main drawback of the single regime MEM MIDAS, that is its lag about explosions of volatility: indeed, a Markov switching model can capture abrupt changes in the level of the series. Conversely, by construction, the long-run component of the MEM-MIDAS evolves smoothly. Furthermore, graphically, the long-run component of the regime-switching MEM MIDAS seems to adequately represent the changing average level around which conditional variance fluctuates.

By dividing the series of the realized volatility by the inferred average level (see fig. 2.3) it is evident that the MS-MEM-MIDAS model can capture abrupt shifts in the average level of volatility, corresponding to periods of market downturns. Looking at fig. (2.6) we can appreciate the ability of the MS(3)-MEM MIDAS to distinguish among low, medium, and high volatility state. There are many switches between states 2 and 3 until mid-2003, reflecting the stock market downturns began in March 2002, while until the first half

[19]See eq. (2.7).
[20]See, for example, Engle, Ghysels, and Sohn (2013).

of 2007 there is a period of moderation, fully captured by the regime of low volatility. Then a period of medium volatility starts with a persistent switch to the regime of high volatility in September 2008 until the end of the year, following the bankruptcy of Lehman Brothers in 2008. The years 2009-2011 are characterized by switches between regimes 1 and 2, with the occasional shift to regime 3 in May 2010, reflecting the flash crash of 6 May. In the second half of 2011, there is a change to state 3 until the end of the year, a period of market downturns related to credit rating downgrade of the United States sovereign debt. From 2012 to December 2013, there is a long period of very low volatility.

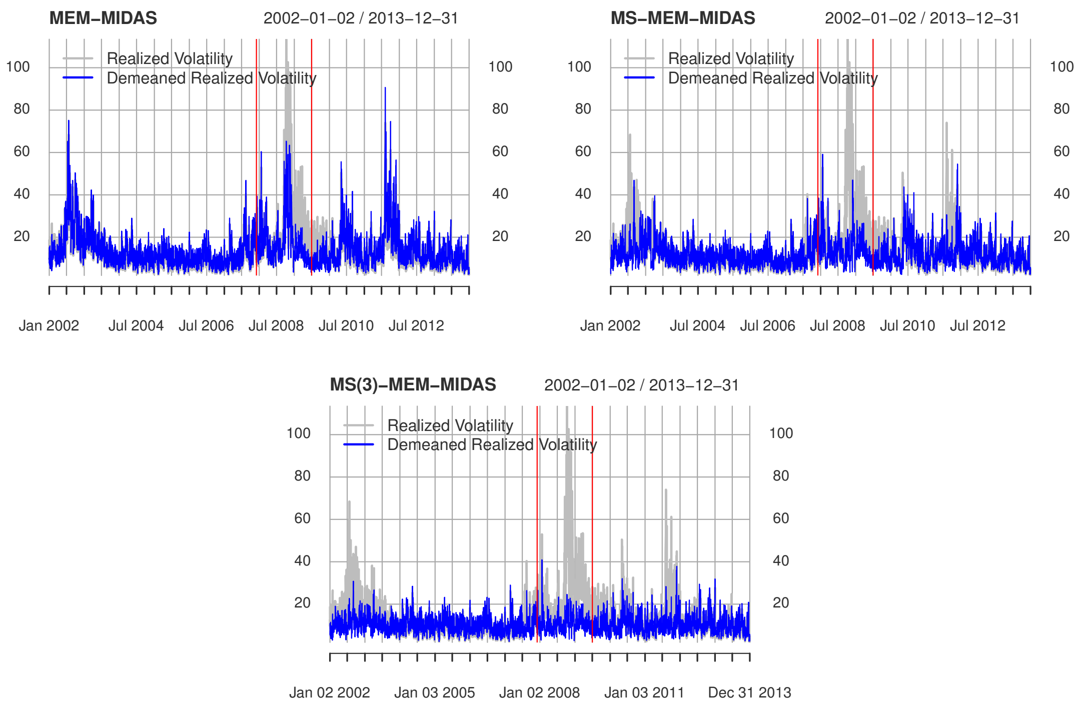
FIGURE 2.2: Estimated Conditional Volatility. Sample Period: 2 January 2002, 31 December 2013



Notes: Realized Volatility (gray line), estimated conditional volatility (blue line), and estimated long-run component (green line). The areas between red vertical bars represent US recession periods date by NBER.

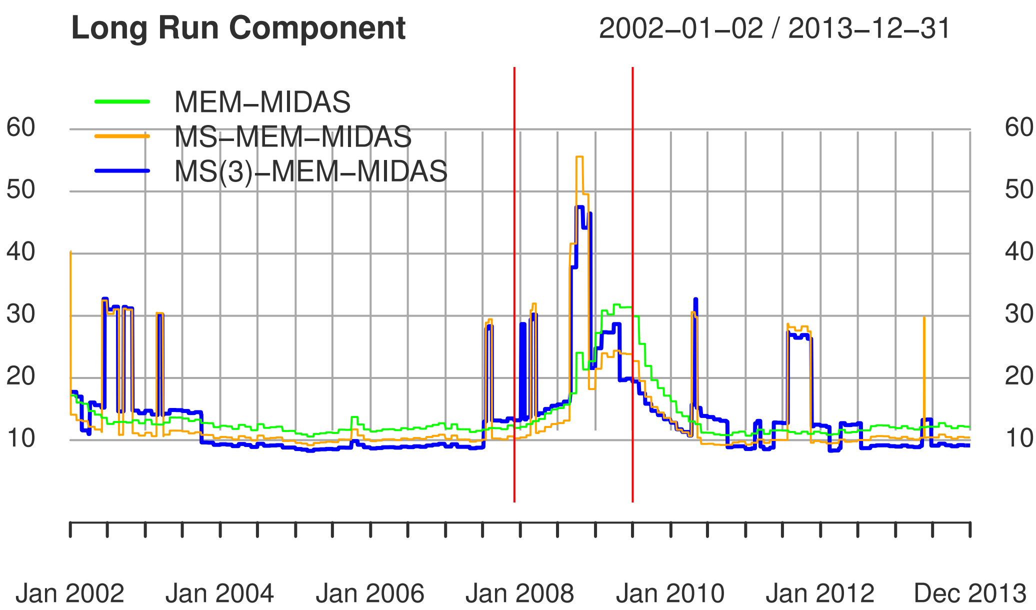
In terms of in-sample performance, table (2.5) shows us that the MS-MEM-

FIGURE 2.3: De-averaged Realized Volatility. Sample Period: 2 January 2002, 31 December 2013



Notes: Realized Volatility (gray line) and the deaveraged Realized Volatility (blue line).

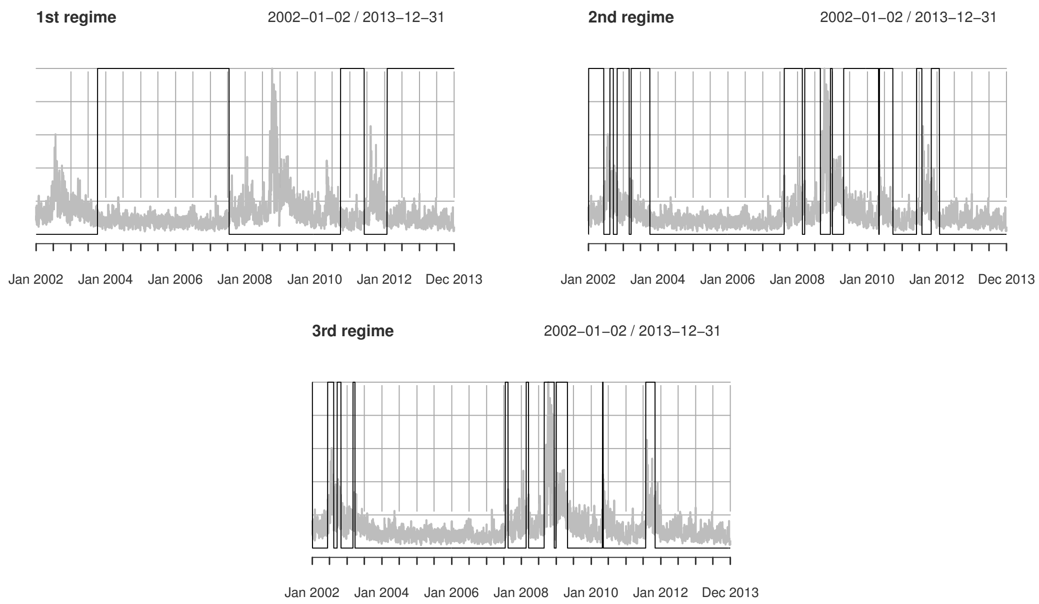
FIGURE 2.4: Long Run Component. Sample Period: 2 January 2002, 31 December 2013.



Notes:Estimated long-run component of MEM MIDAS (green line), MS(2) MEM-MIDAS (orange line), and MS(3) MEM-MIDAS (blue line). The areas between red vertical bars represent US recession periods date by NBER.

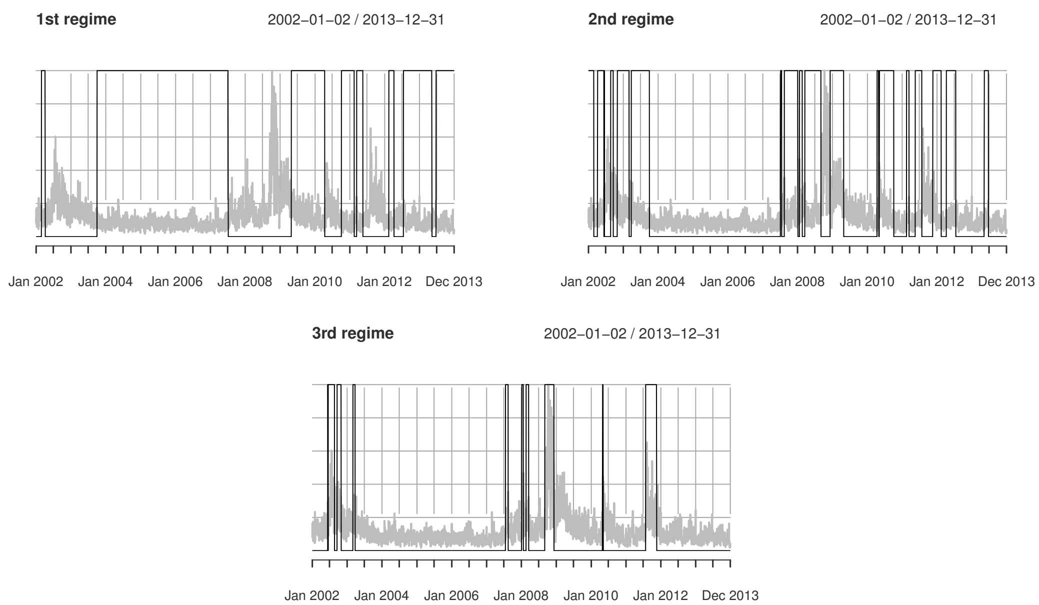
MIDAS has the best value concerning the information criteria, and the two robust loss functions considered here, the Quasi Likelihood (QLIKE) and the

FIGURE 2.5: Smoothed Probabilities MS(3)-MEM. Sample Period: 2 January 2002, 31 December 2013



Notes: Realized Volatility (gray line) and Smoothed Probabilities (black line).

FIGURE 2.6: Smoothed Probabilities MS(3)-MEM-MIDAS. Sample Period: 2 January 2002, 31 December 2013



Notes: Realized Volatility (gray line) and Smoothed Probabilities (black line).

Mean Squared Errors (MSE)[21].

$$\begin{aligned} QLIKE &= \sum_{t=1}^{T}\sum_{i=1}^{N_t} \frac{x_{i,t}}{\hat{x}_{i,t}} - log\hat{x}_{i,t} \\ MSE &= \sum_{t=1}^{T}\sum_{i=1}^{N_t} (x_{i,t} - \hat{x}_{i,t})^2 \end{aligned} \tag{2.17}$$

[21] Accordingly to Patton (2011) a loss function is robust if the ranking of the forecast would not change if we used the true conditional variance in place of its unbiased proxy (remember that the conditional variance is not observable).

TABLE 2.5: In sample performance

	MEM	MEM-MIDAS	MS MEM	MS MEM-MIDAS	MS(3)MEM	MS(3)MEM-MIDAS
Log-lik	-8771.325	-8735.356	-8709.78	-8695.281	-8688.82	**-8673.29**
AIC	5.821	5.799	5.783	5.775	5.774	**5.765**
BIC	5.832	5.813	5.802	5.797	5.804	**5.799**
MSE	36.676	35.854	36.091	35.086	36.123	**34.91**
QLIKE	0.072	0.070	0.071	0.069	0.068	**0.068**
LB_5	0.000	0.001	0.031	0.01	0.008	0.003
LB_{10}	0.000	0.002	0.117	0.046	0.028	0.011
LB_{15}	0.000	0.007	0.294	0.165	0.109	0.058
LB_{20}	0.000	0.007	0.268	0.164	0.079	0.049

Notes: Loglik: maximized value of the log-likelihood. AIC: Akaike Information Criterion. BIC: Bayesian Information Criterion. QLIKE: Quasi-Likelihood function. MSE: mean squared error. In bold the best value for each function. For the QLIKE also the models entering the MCS, $\alpha = 0.25$. Volatility Proxy: annualized Realized kernel Volatility. LB represents the p-values of the Ljung-Box test at each lag for the standardized residuals. Sample period: 2 January 2002-31 December 2013.

where $x_{i,t}$ is the volatility proxy, that is the Realized kernel volatility, while $\hat{x}_{i,t}$ is the estimated conditional volatility.

Then, I compare the performance of the models through the model confidence set (MCS)[22] procedure of Hansen, Lunde, and Nason (2011). Under the null hypothesis, the different models have an equal predictive ability:

$$H_0 : E(d_{ij}) = 0 \qquad i \neq j \quad i, j \in \mathcal{M} \tag{2.18}$$

where d_{ij} is the difference among the loss function of the model i and j, and $\mathcal{M}$ is the set of the compared models. To test the null hypothesis the following test statistic is employed:

$$T_R = \max_{i,j \in \mathcal{M}} \left| \frac{\bar{d}_{ij}}{\sqrt{\hat{V}ar(d_{ij})}} \right| \tag{2.19}$$

where $\bar{d}_{ij}$ is the sample counterpart of $E(d_{ij})$, that is $T^{-1} \sum_{t=i}^{T} d_{ij,t}$ and $\hat{V}ar(d_{ij})$ is the estimated variance of d_{ij} thorugh a bootstrap procedure. When the null is rejected the worst model is eliminated and the test is repeated for the remaining models until the null hypothesis is not rejected, so providing the set of models with equivalent predictive ability. For the MCS procedure[23], I use the quasi-likelihood loss function, due to its highest power[24] since MSE is more sensitive to extreme observations. Indeed, table (2.5) shows that MS-MEM-MIDAS is the only model entering the MCS. From a residual diagnostic point of view, it is evident, again, from table (2.5), that the models with a markovian dynamic have better residual properties by conducting the Ljung-box test.

[22]See Hansen, Lunde, and Nason (2011) for an exhaustive description of the MCS procedure.

[23]To implement the MCS procedure, I use the R package MCS of Bernardi and Catania (2018).

[24]See Patton and Sheppard (2009).

2.5.4 Out of sample analysis

TABLE 2.6: Out of sample results: QLIKE

	MEM	MEM-MIDAS	MS MEM	MS MEM-MIDAS	MS(3) MEM	MS(3) MEM-MIDAS
2014/2015	**0.069**	**0.069**	0.071	0.071	**0.067**	**0.069**
2016/2017	**0.08**	0.088	0.082	0.091	0.088	0.105
2018/2019	**0.062**	**0.062**	**0.062**	**0.062**	0.064	**0.063**
FULL	**0.071**	0.073	0.072	0.075	0.073	0.079

Notes: QLIKE: Quasi-Likelihood function. In bold, the models entering the MCS, $\alpha = 0.25$. Volatility Proxy: annualized Realized Volatility. Rolling window: 12 years. Refitting frequency: 2 years.

TABLE 2.7: Out of sample results: MSE

	MEM	MEM-MIDAS	MS MEM	MS MEM-MIDAS	MS(3) MEM	MS(3) MEM-MIDAS
2014/2015	**18.05**	18.13	19.33	18.87	18.21	18.49
2016/2017	**10.78**	11.62	11.06	12.01	11.82	13.42
2018/2019	16.46	16.39	16.33	**16.29**	16.86	16.32
FULL	**15.09**	15.38	15.58	15.72	15.63	16.08

Notes: MSE: mean squared error. In bold, the best value for each function. Volatility Proxy: annualized Realized Volatility. Rolling window: 12 years. Refitting frequency: 2 years.

The out-of-sample period covers the interval between January 2014 and December 2019. By remembering that the first in-sample period ends in December 2013, we obtain one step ahead forecasts for the following two years (2014-2015) keeping fixed the parameter estimates. At the end of the two years (2015), the model is re-estimated (keeping fixed the window size, 2004-2015), so we generate the one-step-ahead forecasts, keeping fixed the parameter estimates, and so on until the end of the sample (2019). The forecasting exercise (see tab. 2.6 and 2.7) shows us that Markov Switching models do not perform as well as they do in-sample. It is a known result in literature: the better in-sample fit of nonlinear models than parsimonious models does not imply a better out of sample performance[25].

2.6 Concluding Remarks

In this chapter, we mix the properties of the component models, that accommodate smooth changes in the long-run component, and those of the Markov Switching ones, that take into account abrupt shifts in the average volatility level, by proposing a new model, called Markov Switching MEM-MIDAS.

The models with a markovian dynamic show better residual properties and in-sample fitting with respect to the other ones. Graphically, we can appreciate the performance of the model, indeed the MS-MEM-MIDAS model is

[25]See Hansen (2010).

able to capture shifts in the average level of volatility that are not driven by economic variables, thus justifying the approach we use. From an out-of-sample point of view, results suggest that the parsimonious model is the best. Indeed, it is accepted that the nonlinear models have a good in-sample performance, but they could have a worse out-of-sample result.

The proposed model could be extended by considering shifts in the dynamic parameters of the short-run component to account also for different dynamics of the series, or state-dependent coefficients of the low-frequency variable. Moreover, the empirical analysis could be extended to other market indexes, and the set of competing models could be enlarged by considering, e.g., HAR-based models (Corsi, 2009).

Chapter 3

A Multivariate component volatility model for Realized Covariance matrices

3.1 Introduction

Modeling and forecasting the co-movements of financial assets has important implications for finance applications in which the covariance of asset returns matters, such as pricing, risk management, asset allocation, and other related issues. Early multivariate volatility models, started with the VECH of Bollerslev, Engle, and Wooldridge (1988)[1], assumed a constant average (or long-run) level of covariances, while now it is widely accepted that the long-run level is time-varying. So, multivariate component volatility models, that decompose conditional covariance into a short-run and a long-run component are widespread in the literature. Indeed, the component structure can capture, in a parsimonious way, the long-memory behavior of covariance matrices, by distinguishing between transitory and permanent effects of the different sources of covariances. For this purpose, in their DDC-MIDAS, Colacito, Engle, and Ghysels (2011) modeled variances as a GARCH-MIDAS, while the correlations have DCC-type dynamics. More in detail, they assume a time-varying intercept, reflecting long-run movements extracted through the MIDAS filter. The idea of a time-varying long-run level was later extended to models based on Realized Covariance matrices. The latter is a more precise ex-post estimator of the conditional covariance matrix than cross-product returns, by exploiting the information of intraday data. For this purpose, Golosnoy, Gribisch, and Liesenfeld (2012), in their Conditional Autoregressive Wishart (CAW) model, multiplicatively decomposed the conditional covariance matrix into a short-run component, in which the realized

[1] The VECH model is a natural multivariate extension of the GARCH one.

covariance matrix, purged by the long-run component, follows a BEKK-type recursion, and a long-run component that, through the use of the MIDAS filter, is a linear function of past monthly Realized Covariance matrices. Conversely, Bauwens, Braione, and Storti (2016) specified the short-run component through a DCC-type recursion. Nevertheless, the Cholesky decomposition of the covariance matrix is, at least potentially, sensible to the order of the assets, then a model invariant to the ordering should be preferable. Then, in the spirit of the DCC-MIDAS of Colacito, Engle, and Ghysels (2011), we propose a model for Realized Covariance matrices without using the Cholesky decomposition of the covariance matrix, thus ensuring the invariance to the order of the assets. Moreover, any type of decomposition is not needed, so there is a great advantage from a computational point of view. More specifically, the short-run component of the model follows a scalar BEKK-type recursion, while the long-run component is a function of past monthly realized covariance matrices. An extension is also presented, by using the Hadamard exponential function of Bauwens and Otranto (2020a), which allows asset-pair specific and time-varying parameters with one coefficient more than the scalar parameterization. Then a more flexible model is provided but avoiding parameter proliferation.

The structure of the chapter is organized as follows: Section (3.2) describes the new model proposed, while in Section (3.2.1) also an extension of the base specification is provided. In Section (3.3) statistical inference is discussed, while Section (3.4) presents the data employed in the empirical analysis (3.4.1), estimation results (3.4.2), an in-sample models comparison (3.4.3), and an out-of-sample exercise (3.4.4). Finally, Section (3.5) offers some concluding remarks.

3.2 Theoretical model

Let C_t be an order n positive definite symmetric (PDS) matrix, here the realized covariance (RC) one, that conditionally at the information set at time $t-1$, is assumed to follow a n-dimensional Wishart distribution:

$$C_t|\mathcal{I}_{t-1} \sim W_n\left(\nu, S_t/\nu\right), \quad \forall\, t = 1, ..., T \tag{3.1}$$

where,

- $\nu > n-1$ are the degrees of freedom;

- S_t is a PDS scale matrix and it is the conditional expectation of the Realized Covariance matrix, (C_t), i.e., the conditional covariance matrix:

$$E(C_t|\mathcal{I}_{t-1}) = S_t \tag{3.2}$$

Let us consider a scalar BEKK-type recursion for the conditional covariance matrix, S_t:

$$S_t = M(1-\alpha-\beta) + \alpha C_{t-1} + \beta S_{t-1} \tag{3.3}$$

where M is the unconditional covariance, a PDS matrix, and the usual constraints to ensure the stationarity are imposed: $\alpha > 0$, $\beta \geq 0$, $\alpha + \beta < 1$. Similarly to Colacito, Engle, and Ghysels (2011), I allow the intercept matrix, M, to be time varying[2]. Then, the conditional covariance matrix, S_t is additively decomposed into a smoothly time-varying long-run component and a short-lived one. Then the proposed model, call it Realized BEKK MIDAS (ReBEKKMIDAS), is specified as follows:

$$S_t = M_t(1-\alpha-\beta) + \alpha C_{t-1} + \beta S_{t-1} \tag{3.4}$$

The long-run component, M_t, that takes into account secular movements in the level of conditional variances and covariances, is specified as follows:

$$\begin{aligned} M_t &= \bar{\Lambda} + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) C_{t-k}^{(m)} \\ C_{t-k}^{(m)} &= \sum_{\tau=t-mk}^{t-m(k-1)-1} C_\tau \\ \varphi_k(\omega_1,\omega_2) &= \frac{(k/K)^{\omega_1-1}(1-k/K)^{\omega_2-1}}{\sum_{j=1}^{K}(j/K)^{\omega_1-1}(1-j/K)^{\omega_2-1}} \end{aligned} \tag{3.5}$$

where,

- $\bar{\Lambda} = LL'$ and L is an order n lower triangular matrix, with positive diagonal entries as identifying condition;
- $C_t^{(m)}$ is an order n matrix of monthly Realized Covariances, i.e., the aggregation of daily Realized Covariances over m = 22 days. Note that with the above parameterization, the matrix changes day by day, while in principle it could be constant for the whole low frequency period, here 22 days;

[2] Colacito, Engle, and Ghysels (2011) assumed a DCC-type dynamic for the quasi conditional correlation matrix, by replacing the constant intercept matrix with a time-varying one.

- θ is a non-negative scalar parameters, while $\varphi(\omega_1, \omega_2)$ is a weighting function of the past K values of $C_t^{(m)}$, with the weights that sum up to one; if, in general, $\omega_1 = 1$ and $\omega_2 > 1$, this function is monotonically decreasing, as fast as ω_2 increases.

For what concerns the long-run component, notice that in eq. (3.4) we have a unique weighting scheme and slope coefficient for variances and covariances, while in Colacito, Engle, and Ghysels (2011) there is a different specification for the long-run component variances and correlations[3]. Furthermore, as we can see from fig. (3.1) it seems to be a reasonable compromise to assume the same dynamic of the long-run component for all the assets, thus avoiding the proliferation of parameters: periods of high volatility, then with an increase in the average level, are very similar for all the series.
Short-lived effects are captured by the dynamic of S_t: rewriting eq. (3.4)

$$S_t - M_t = \alpha(C_{t-1} - M_t) + \beta(S_{t-1} - M_t) \tag{3.6}$$

We can notice the short-run component fluctuating around the long-run one, M_t.
Note that with the above parameterization, the estimation is invariant to the order of the assets. Indeed, the use of the Cholesky decomposition for the long-run component in other models, such as the Multivariate MIDAS Aggregated Realized BEKK (MMAReBEKK) of Bauwens, Braione, and Storti (2016)[4], renders the short-run Realized Covariance matrix[5] sensible to assets ordering. Moreover, we do not need to calculate for each t the Cholesky factor of the long-run component, $M_t^{1/2}$, then there is a great advantage from a computational point of view: indeed, at each step of the optimization procedure, we have to compute the Cholesky factor and its inverse[6]. Practically, in the empirical analysis in Section (3.4), we obtain a 30 percent time gain in the estimation of the model.

3.2.1 The Hadamard exponential function

The scalar ReBEKKMIDAS imposes the same dynamic for each asset, that is, the coefficients are the same for each series, but this could be a strong

[3]See eq. (1.47) in chapter 1.

[4]The MMAReBEKK is equivalent to the CAW MIDAS described in chapter 1 if we assume a scalar specification for the matrices of dynamic parameters, that is $A = \alpha I$ and $B = \beta I$, where I is the Identity matrix, while α and β are two scalar coefficients.

[5]The short-run Realized Covariance matrix is the the Realized Covariance matrix purged by the long-run component.

[6]See eq. (1.49) in chapter 1.

condition. For this purpose, the model above could be extended through the Hadamard exponential function used by Bauwens and Otranto (2020a), which allows the parameters to be asset-pair specific and time-varying, with one more parameter than the scalar parameterization. Let us consider the Hadamard parameterization of the scalar ReBEKK[7]:

$$S_t = M(1-\alpha-\beta) + \alpha J_n \odot C_{t-1} + \beta J_n \odot S_{t-1} \tag{3.7}$$

where,

- J_n is an order n square matrix of ones;
- $\odot$ is the element-by-element (Hadamard) product, that is, if A and B are two square matrices of order n, then $A \odot B = (a_{ij}b_{ij})$.

Then, the Hadamard exponential extension of the Realized BEKK MIDAS, call it Hadamard Exponential Realized BEKK MIDAS (HEReBEKK-MIDAS), is specified as follows:

$$S_t = M_t(1-\bar{\alpha}_t-\bar{\beta}_t) + A_t \odot C_{t-1} + B_t \odot S_{t-1} \tag{3.8}$$

where,

- A_t and B_t are the matrix of asset-pair specific and time-varying parameters, parameterized through the Hadamard Exponential function;
- $\bar{\alpha}_t$ and $\bar{\beta}_t$ are two scalars representing the average value of the elements in the matrix A_t and B_t, respectively.

However, in empirical studies, only one matrix is parameterized through the Hadamard exponential function: e.g., Bauwens and Otranto (2020a) found that the coefficients relative to the Realized Covariance matrix were time-varying for the models with a BEKK-type recursion. More specifically, the elements of the matrix A_t depend on the lagged conditional (or Realized) correlations:

$$A_t = \alpha\, exp^{\odot}[\phi_A(N_{t-1} - J_n)] = \alpha \frac{exp^{\odot}[\phi_A(N_{t-1})]}{exp(\phi_A)} \tag{3.9}$$

where,

- $exp^{\odot}$ is the Hadamard exponential function, or entry-wise exponential operator, that is, if A square matrif of order n, then $exp^{\odot} A = (exp(a_{ij}))$;

[7]Note that the model in eq. (3.7) is equivalent to the specification in eq. (3.3).

- ϕ_A is a non-negative scalar parameter;
- $N_t = P_t$ or R_t[8];
- $P_t = \{diag(C_t)\}^{-1/2} C_t \{diag(C_t)\}^{-1/2}$ is the Realized correlation matrix;
- $R_t = \{diag(S_t)\}^{-1/2} S_t \{diag(S_t)\}^{-1/2}$ is the conditional correlation matrix.

The diagonal elements of A_t are equal to α (N_{t-1} is a matrix of ones along the main diagonal by construction), while the off-diagonal elements are between 0 and α (the off-diagonal elements of N_{t-1} are between -1 and 1). Notice that if $\phi_A = 0$, then $A_t = \alpha J_n$, i.e., the model reduces to the scalar specification, then we have a constant parameters matrix. Importantly, the entry-wise exponential operator preserves the positive definiteness of a PD matrix[9]. Moreover, it is important to stress that an exact parameterization of the time-varying intercept term should be: $M_t \odot (J_n - A_t - bJ_n)$. Nevertheless, this matrix would not guarantee the positive definiteness of S_t. Then, like Bauwens and Otranto (2020a), we use the approximation proposed by Hafner and Franses (2009) for the intercept, that is $M_t(1 - \bar{\alpha}_t - \beta)$.

Bauwens and Otranto (2020a) justify the dependence of the coefficients in the matrix A_t on the past conditional (or Realized) correlation matrix, through the volatility clustering, a stylized fact of financial returns[10]. Indeed, when there is a period of high market volatility, correlations and their persistence increase, but differently, at least potentially, for each pair of assets. They capture this phenomenon with the parameterization above: so, when correlation increases, the impact of the lagged covariance on the next conditional covariance is stronger.

3.3 Quasi Maximum Likelihood estimation

The parameters of the proposed model are estimated via the Quasi Maximum Likelihood (QML) method in one-step. Indeed, models with a time-varying average level, cannot be directly estimated through the covariance

[8]The choice is an empirical question.
[9]See Appendix D for the proof.
[10]See ch. 1.

targeting[11], a two-step procedure[12]. In that case, before maximizing the log-likelihood, the unconditional covariance matrix can be estimated in the first step by a consistent, though inefficient, estimator based on the sample covariance, that is, $\hat{M} = T^{-1}\sum_{t=1}^{T} C_t$.

Based on the Wishart assumption, what follows is the quasi log-likelihood (omitting the part that does not depend on Φ):

$$LL(\Phi) = \sum_{t=1}^{T} ll_t(\Phi) = -\nu/2 \sum_{t=1}^{T} \left[\ln|S_t| + tr\left(S_t^{-1} C_t\right) \right] \tag{3.10}$$

where Φ is the vector of parameters to be estimated. Let us consider the score of each parameter, ϕ_i:

$$\frac{\partial ll_t(\Phi)}{\partial \phi_i} = -\nu/2 \left[tr\left(S_t^{-1} \frac{\partial S_t}{\partial \phi_i} - C_t S_t^{-1} \frac{\partial S_t}{\partial \phi_i} S_t^{-1} \right) \right] \tag{3.11}$$

Then, let us compute the conditional expected valued of the score[13]:

$$E_{t-1} \frac{\partial ll_t(\Phi)}{\partial \phi_i} = -\nu/2 \left[tr\left(S_t^{-1} \frac{\partial S_t}{\partial \phi_i} - \frac{\partial S_t}{\partial \phi_i} S_t^{-1} \right) \right] = 0 \tag{3.12}$$

That is, the score is a martingale difference sequence and, by the law of iterated expectations, $E\frac{\partial ll_t(\Phi)}{\partial \phi_i} = 0$[14]. As you can see, the parameter of the Wishart distribution does not influence the parameter estimates because the first-order conditions are a linear function of ν, then the latter can be set equal to 1 during the estimation. This is an important result because, even if the assumed distribution is misspecified, the estimator is consistent, given that the conditional expectation of C_t is correct, then its QML interpretation. Consequently, the standard errors are calculated through the Sandwich matrix[15], that is the variance-covariance matrix of Φ is estimated as follows:

$$Var(\hat{\Phi}) = H^{-1} OP H^{-1} \tag{3.13}$$

[11]See, for example, Bauwens, Storti, and Violante (2012).

[12]Nevertheless, Bauwens, Braione, and Storti (2017) proposed an algorithm that iteratively maximizes the moment-based QML function (Iterative Moment-based Profiling estimator - IMP) thus rendering the estimation feasible for a growing number of assets. However, IMP is less efficient than QML estimator, that maximizes in one step.

[13]See Comte and Lieberman (2003) for multivariate GARCH and Bauwens, Storti, and Violante (2012) for models based on Realized Covariance matrices.

[14]Remember that $E_{t-1}(C_t) = S_t$.

[15]See White (1982).

where H^{-1} is the inverse of the Hessian matrix with the opposite sign and OP is the matrix of the Outer Products of the scores, that is:

$$\begin{aligned} H &= -\frac{\partial^2 LL(\Phi)}{\partial\Phi\partial\Phi'} \\ OP &= \sum_{t=1}^{T} \frac{\partial ll_t(\Phi)}{\partial\Phi\partial\Phi'} \frac{\partial ll_t(\Phi)}{\partial\Phi\partial\Phi'}' \end{aligned} \tag{3.14}$$

Empirically, the model parameters are estimated numerically on the software R through a personal code[16].

3.4 Empirical Analysis

3.4.1 Dataset

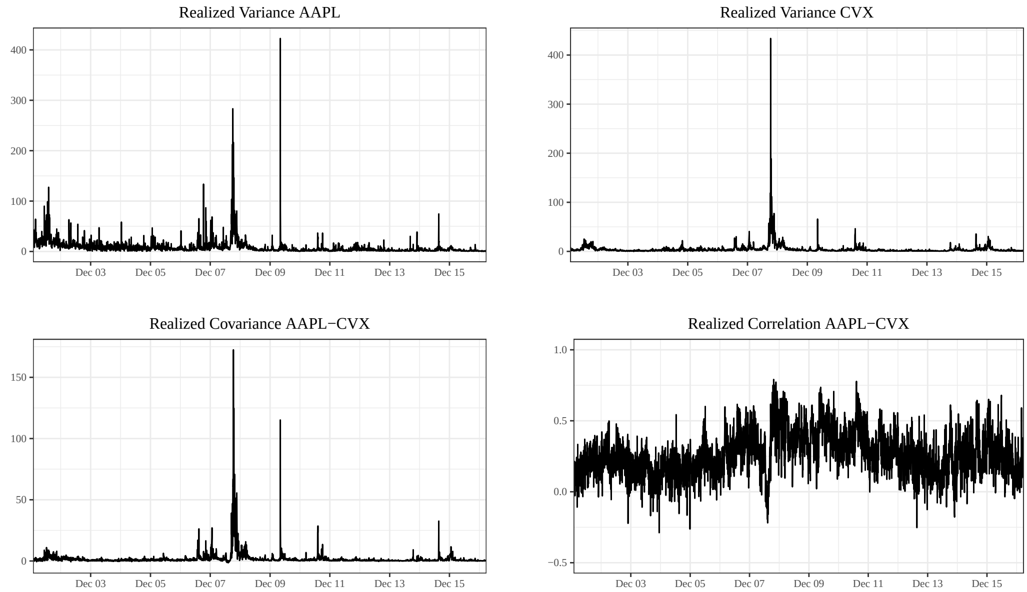
In the empirical analysis, we use a time series of daily Realized Covariance matrices of nine assets belonging to the Dow Jones Industrial Average Index: Apple Inc. (AAPL), Chevron Corporation (CVX), The Walt Disney Company (DIS), The Goldman Sachs Group, Inc. (GS), Home Depot Inc. (HD), International Business Machines Corporation (IBM), Intel Corporation (INTC), 3M Company (MMM), and Exxon Mobil Corporation (XOM). The data we employ is a subset of the dataset used by Bauwens and Otranto (2020b)[17]. This subset is obtained through a grouping algorithm that identifies assets with similar coefficients: in a few words, each group is formed by applying the Wald test to the parameter estimates, and then the model is estimated by constraining the parameters to be the same within each group[18]. The daily Realized Covariance matrix is constructed as the sum of one-minute intraday log returns. By looking at the descriptive statistics in tab. (3.1) and (3.2), variances and covariances are positively skewed and their empirical distribution has a much higher level of excess of kurtosis than the normal distribution. Furthermore, as we can see from fig. (3.1), the long-run level of covariances is not constant, more specifically it is much higher during periods of market downturns, such as the crisis in the years 2007/2008 and the flash crash on May 6, 2010. Then, to allow the average level to be time-varying is empirically justified. Finally, by comparing variances and covariances, we can say

[16] https://www.r-project.org/.

[17] The dataset for the DJIA companies was provided by Lyudmila Grigoryeva (University of Konstanz) and Juan-Pablo Ortega (University of St.Gallen), while Oleksandra Kukharenko (University of Konstanz) computed the Realized Covariance matrices.

[18] See Bauwens and Otranto (2020b) for an exhaustive description of the grouping algorithm.

FIGURE 3.1: Apple & Chevron annualized Realized Variances, Covariance and Correlation. Sample Period: 1 January 2001-28 March 2017.



Notes: Annualized Realized Variances and Covariance are expressed in percentage scale.

TABLE 3.1: Descriptive Statistics of Realized Variances

	Mean	Median	Min	Max	St. Dev.	Skewness	Kurtosis
AAPL	7.657	4.051	0.203	422.73	14.197	11.913	254.389
CVX	4.253	2.352	0.042	433.818	10.270	22.181	802.621
DIS	5.074	2.381	0.061	240.486	9.141	8.382	143.369
GS	7.743	3.031	0.048	1200.654	31.163	24.011	766.83
HD	5.157	2.53	0.097	346.298	10.233	12.789	332.964
IBM	3.299	1.67	0.023	188.96	7.011	11.237	201.174
INTC	6.251	3.622	0.093	205.325	9.131	6.647	87.749
MMM	3.063	1.69	0.048	405.997	8.552	29.842	1282.886
XOM	3.917	2.029	0.076	444.402	10.160	24.001	917.213

Notes: The table reports the Mean, the Median, the Minimum (Min), the Maximum (Max), the Standard Deviation (St. Dev.), the Skwewness and the Kurtosis of the Realized Variances. All the variables are expressed in annualized percentage terms. Sample period: 28 January 2002-16 April 2018.

that the former has a higher average level and shows more variability than the latter.

3.4.2 Estimation Results

The estimation period spans 1 January 2001 to 28 March 2017, consisting of 4055 daily observations. It is worth noticing that, to initialize the MIDAS filter, the first 264 observations are used as starting values, so the sample

TABLE 3.2: Descriptive Statistics of Realized Covariances

	Mean	Median	Min	Max	St. Dev.	Skewness	Kurtosis
AAPL-CVX	1.791	0.688	-1.422	172.481	5.881	14.555	304.422
AAPL-DIS	1.810	0.706	-1.837	182.188	5.183	16.706	451.242
AAPL-GS	2.206	0.784	-1.668	253.184	7.371	15.754	399.398
AAPL-HD	1.787	0.669	-0.576	145.15	5.462	14.466	303.456
AAPL-IBM	1.597	0.644	-0.379	131.539	4.857	14.162	285.238
AAPL-INTC	2.347	0.959	-1.83	131.68	5.573	11.077	191.672
AAPL-MMM	1.294	0.492	-3.041	203.052	4.85	23.918	842.612
AAPL-XOM	0.738	0.146	-4.997	208.197	4.927	25.732	900.322
CVX-DIS	1.754	0.609	-0.99	176.828	5.214	15.092	380.842
CVX-GS	2.119	0.691	-4.285	347.785	8.172	22.644	827.684
CVX-HD	1.652	0.527	-4.414	218.791	5.909	17.902	520.271
CVX-IBM	1.425	0.496	-1.075	171.799	4.898	16.611	438.429
CVX-INTC	1.772	0.674	-1.608	189.507	5.385	16.477	442.973
CVX-MMM	1.333	0.459	-0.695	161.303	4.681	16.879	437.353
CVX-XOM	1.401	0.475	-5.24	337.428	7.188	30.39	1260.382
DIS-GS	2.130	0.74	-0.868	223.831	6.264	15.665	432.812
DIS-HD	1.954	0.69	-0.401	155.005	5.253	12.363	252.92
DIS-IBM	1.549	0.567	-0.43	114.048	4.323	11.794	209.567
DIS-INTC	2.002	0.758	-1.231	119.725	4.672	9.99	167.714
DIS-MMM	1.427	0.517	-0.561	138.212	4.391	16.085	398.29
DIS-XOM	0.988	0.267	-4.639	223.758	5.067	26.505	1001.234
GS-HD	2.358	0.76	-1.741	266.692	7.446	15.952	440.501
GS-IBM	1.862	0.68	-0.958	185.475	5.713	14.408	336.65
GS-INTC	2.321	0.89	-1.236	220.608	6.454	14.760	377.155
GS-MMM	1.723	0.618	-0.684	204.155	5.447	17.587	524.408
GS-XOM	1.216	0.317	-10.883	346.097	7.168	31.353	1366.129
HD-IBM	1.675	0.629	-0.316	133.454	4.739	12.227	230.726
HD-INTC	2.142	0.852	-0.576	167.867	5.309	12.986	295.879
HD-MMM	1.581	0.585	0.206	141.314	4.83	15.520	369.287
HD-XOM	0.994	-0.261	-4.186	241.542	5.288	28.148	1127.165
IBM-INTC	1.868	0.829	-1.113	115.706	4.405	11.117	195.989
IBM-MMM	1.332	0.539	-0.618	139.394	4.214	17.069	437.235
IBM-XOM	0.9	0.265	-3.361	172.769	4.265	22.866	752.043
INTC-MMM	1.657	0.729	-0.933	180.064	4.69	19.683	616.484
INTC-XOM	1.118	0.377	-2.706	204.705	4.948	24.09	819.357
MM-XOM	1.016	0.336	-3.029	197.19	5.088	27.73	971.92

Notes: The table reports the Mean, the Median, the Minimum (Min), the Maximum (Max), the Standard Deviation (St. Dev.), the Skwewness and the Kurtosis of the Realized Covariances. All the variables are expressed in annualized percentage terms. Sample period: 28 January 2002-16 April 2018.

TABLE 3.3: Model Specifications

Model	Functional Form
ReBEKK	$S_t = M(1-\alpha-\beta) + \alpha C_{t-1} + \beta S_{t-1}$
MMAReBEKK	$S_t = M_t^{1/2} S_t^* (M_t^{1/2})'$, with $M_t^{1/2} = chol(M_t)$ $S_t^* = (I_n - \alpha - \beta) + \alpha C_{t-1}^* + \beta S_{t-1}^*$, with $C_t^* = (M_t^{1/2})^{-1} C_t (M_t^{1/2})'^{-1}$ $M_t = \bar{\Lambda} + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) C_{t-k}^{(m)}$, with $C_{t-k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} C_\tau$
ReBEKK-MIDAS	$S_t = M_t(1-\alpha-\beta) + \alpha C_{t-1} + \beta S_{t-1}$ $M_t = \bar{\Lambda} + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) C_{t-k}^{(m)}$, with $C_{t-k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} C_\tau$
HEReBEKK-MIDAS-P_t	$S_t = M_t(1-\bar{\alpha}_t - \bar{\beta}_t) + A_t \odot C_{t-1} + B_t \odot S_{t-1}$ $A_t = \alpha \frac{exp^{\odot}[\phi_A(P_{t-1})]}{exp(\phi_A)}$, with $P_t = \{diag(C_t)\}^{-1/2} C_t \{diag(C_t)\}^{-1/2}$ $M_t = \bar{\Lambda} + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) C_{t-k}^{(m)}$, with $C_{t-k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} C_\tau$
HEReBEKK-MIDAS-R_t	$S_t = M_t(1-\bar{\alpha}_t - \bar{\beta}_t) + A_t \odot C_{t-1} + B_t \odot S_{t-1}$ $A_t = \alpha \frac{exp^{\odot}[\phi_A(R_{t-1})]}{exp(\phi_A)}$, with $R_t = \{diag(S_t)\}^{-1/2} S_t \{diag(S_t)\}^{-1/2}$ $M_t = \bar{\Lambda} + \theta \sum_{k=1}^{K} \varphi_k(\omega_1, \omega_2) C_{t-k}^{(m)}$, with $C_{t-k}^{(m)} = \sum_{\tau=t-mk}^{t-m(k-1)-1} C_\tau$

Notes: The table reports the functional form for the ReBEKK, MMAReBEKK, ReBEKK-MIDAS, HEReBEKK-MIDAS-P_t, and HEReBEKK-MIDAS-R_t specifications.

reduces to 3791 observations. Indeed, we use 12 lagged values of monthly Realized Covariance matrices, where the latter is the sum of 22 daily Realized Covariance matrices. Then, for comparing the models estimated, we delete the first 264 observations for the scalar ReBEKK, which does not need any additional observation. The estimates of the coefficients are in line with those of the previous studies: as we can see from tab. (3.4), the scalar ReBEKK shows a very high degree of persistence, that is, the sum of α and β is close to one. For what concerns component models, the θ value, representing the impact of past monthly Realized Covariance on the long-run component, is significant at one percent level, thus justifying the presence of a time-varying average level. The value of ω_2, governing the weighting scheme of monthly Realized Covariances, is very high, then the weights decay very quickly, that is, the most recent values are the most weighted. As for the univariate models in chapter 2, it is useful to calculate the Marginal effect (ME) of the monthly Realized Covariance on the long-run component:

$$ME = \theta \cdot \varphi_k(\omega_2) \cdot \Delta c_{ij,t-k}^{(m)} \tag{3.15}$$

Then, for the ReBEKK-MIDAS, with $\theta = 0.04$ and $\varphi_1 = 0.542$, a unit increase in currently monthly Realized Covariance will increase next month long-run component by 0.02. Furthermore, a unit increase in the previous month of the monthly RC, with $\varphi_2 = 0.263$, would rise two months ahead long-run component by 0.01, with a cumulative effect of two months unit increase of about 0.03. Moreover, the covariance of component models has a persistence lower than the scalar ReBEKK, that is a model assuming a constant long-run component. For what concerns the Hadamard exponential extension of the ReBEKK-MIDAS, the parameter ϕ is significant at 1% in both specifications considering the lagged Realized or Conditional Correlations, as driving the elements in the matrix A_t. So the impact coefficient of the lagged Realized Covariance is time-varying and asset-pair specific. In fig. (3.2) we can analyze the path of the time-varying coefficient $a_{ij,t}$: when the correlation among the two assets increases, also future conditional variance increases, *ceteris paribus*. Moreover, when the forcing variable is the Conditional Correlation matrix, rather than the Realized one, the series is smoother, as expected. Finally, the adjoint of asset-pair and time-varying coefficient seems to lower the persistence of the model than the other ones.

TABLE 3.4: Estimation results of five CAW based models for Realized Covariance matrices of 9 assets belonging to DIJA

	ReBEKK	MMAReBEKK	ReBEKK-MIDAS	HEReBEKK-MIDAS-P_t	HEReBEKK-MIDAS-R_t
N_p	47	49	49	50	50
α	0.347***	0.351***	0.362***	0.365***	0.393***
	(0.018)	(0.021)	(0.02)	(0.022)	(0.024)
β	0.644***	0.535***	0.545***	0.519***	0.458***
	(0.018)	(0.036)	(0.03)	(0.033)	(0.033)
θ		0.037***	0.04***	0.023***	0.024***
		(0.002)	(0.003)	(0.002)	(0.002)
ω_2		7.102***	9.327***	9.335***	10.809***
		(1.465)	(1.63)	(1.65)	(2.16)
ϕ				0.301***	0.304***
				(0.028)	(0.03)

Notes: For the specification of the models see sec. (3.2). Dependent variable: Annualized Realized kernel Covariance matrix. Low frequency variable: monthly Realized kernel Covariance. Number of lagged monthly Realized kernel Covariances: k=12. Sample period: 28 January 2002-28 March 2017. Daily observations: 3791. N_p = Number of estimated parameters. pvalue: $*** < 0.01$, $** < 0.05$, $* < 0.1$. In brackets, standart errors based on sandwich matrix.

3.4.3 In sample comparison

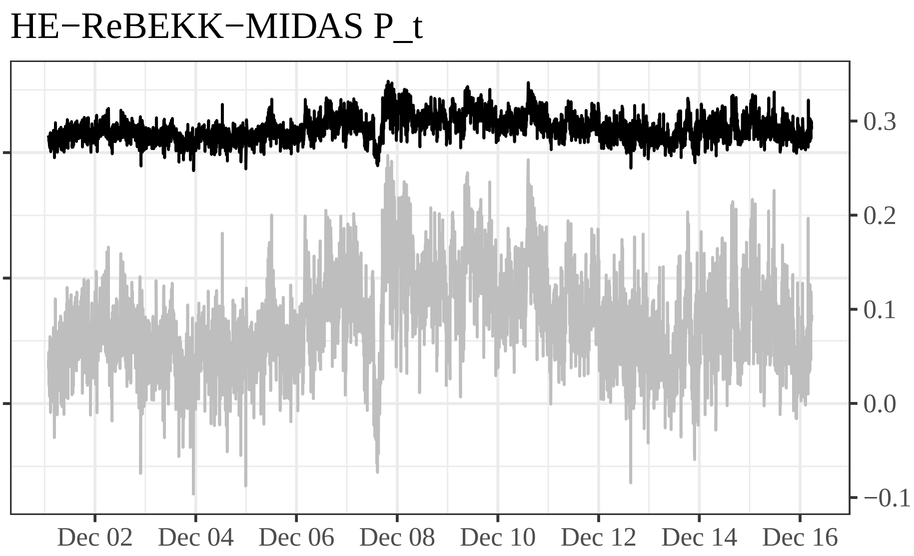
TABLE 3.5: In sample performance

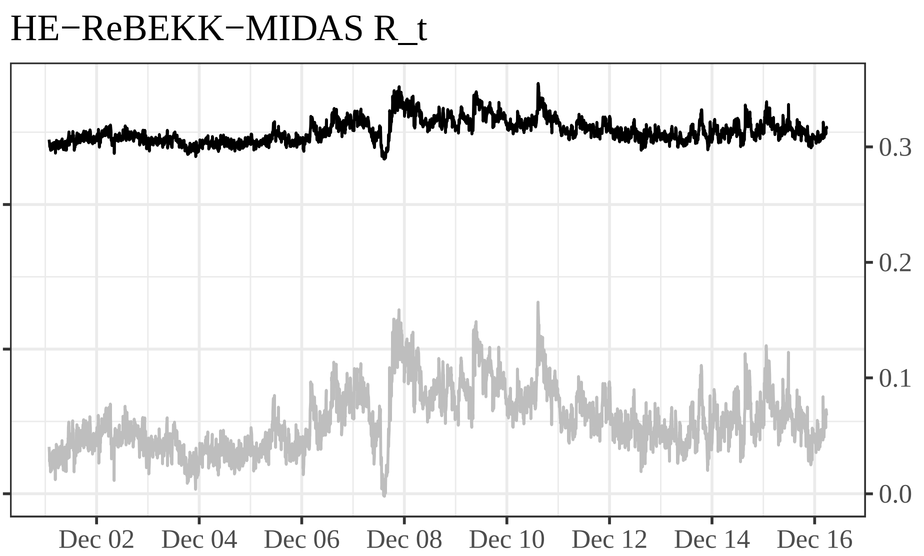
	ReBEKK	MMAReBEKK	ReBEKK-MIDAS	HEReBEKK-MIDAS-P_t	HEReBEKK-MIDAS-R_t
Loglik	-34067.35	-34048.49	-34033.76	-34018.77	**-34003.05**
AIC	17.998	17.989	17.981	17.973	**17.965**
BIC	18.075	18.069	18.062	18.056	**18.048**
Qlik	17.973	17.963	17.955	17.947	**17.939**
Fn	2571.516	3578.789	**2509.439**	**2502.224**	**2505.867**

Notes: Loglik: maximized value of the log-likelihood. AIC: Akaike Information Criterion. BIC: Bayesian Information Criterion. QLIK: Quasi-Likelihood function. FN: squared Frobenius Norm. In bold the best value for each function. For the (QLIK) and the (FN) also the models entering the MCS, $\alpha = 0.25$. Covariance Proxy: annualized Realized kernel Covariance matrix. Sample period: 28 January 2002-28 March 2017.

In terms of in-sample performance, table (3.5) shows us that the HEReBEKK-MIDAS-R_t has the best value relative to the information criteria, and the loss functions considered here, that is the Quasi Likelihood (QLIKE) and the

FIGURE 3.2: Estimated $a_{ij,t}$ for the covariance between Apple Inc. and Chevron Corporation. Sample Period: 28 January 2002-28 March 2017.



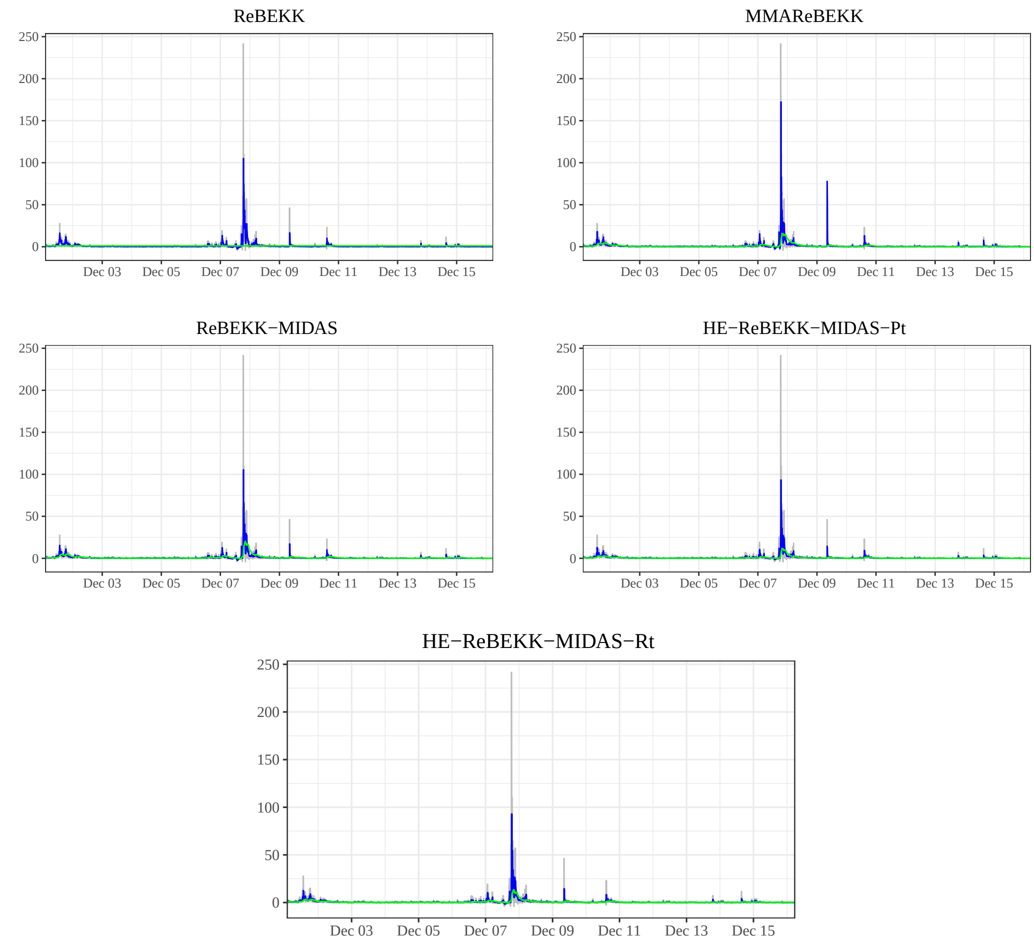


Notes: At the top: $a_{ij,t}$ (black line), Realized Correlation (black line). At the bottom: $a_{ij,t}$ (black line), Conditional Correlation (gray line).

squared Frobenius norm (FN). Both functions are robust in the sense of Patton and Sheppard (2009).

$$\begin{aligned} QLIKE &= \sum_{t=1}^{T} tr\left(S_t^{-1}C_t\right) + \ln|C_t| \\ FN &= \sum_{t=1}^{T} tr\left[(S_t - C_t)'\,(S_t - C_t)\right] \end{aligned} \tag{3.16}$$

FIGURE 3.3: Estimated Conditional Covariance among HD and XOM. Sample Period: 28 January 2002-March 2017.



Notes: Realized Covariance (gray line), Conditional Covariance (blue line), Long run component (green line).

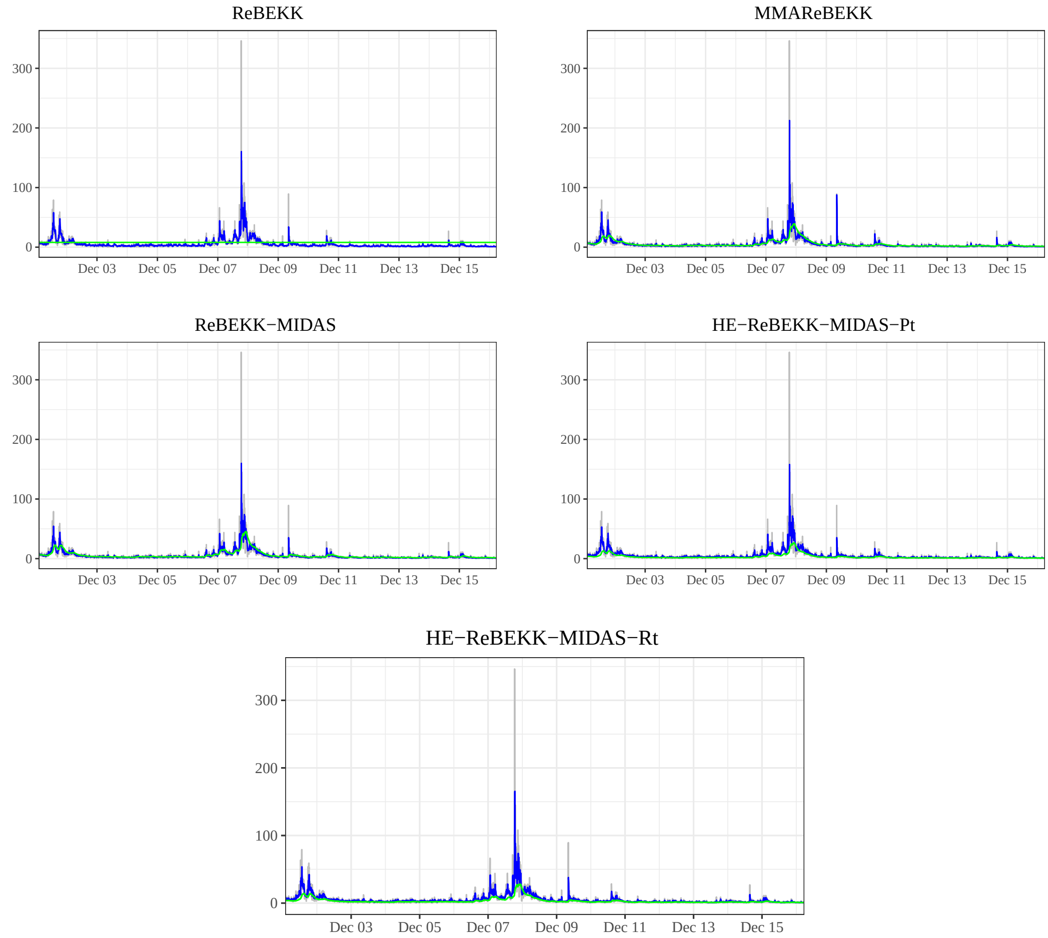
Then, we evaluate the in-sample performance of the models estimated through the model confidence set (MCS) procedure[19]. We consider the quasi-likelihood loss function and the squared Frobenius norm due to their robustness property[20]. Tab. (3.5) shows us that for the QLIKE loss function, only the HEReBEKK-MIDAS-R_t enters the set of superior models, while, for the squared Frobenius norm, all the models, except the MMAReBEKK enter the MCS.

Graphically, (see fig. 3.3, 3.4, and 3.5), all the models considered provide an adequate estimate of the conditional covariances and variances. For what concerns models with changing average level, we can appreciate a persistent increase of the long-run covariance during the second half of the 2008. In the

[19] To implement the MCS procedure, we use the R package MCS of Bernardi and Catania (2018).

[20] See ch. 2 for the definition of a robust loss function.

FIGURE 3.4: Estimated Conditional Variance of HD. Sample Period: 28 January 2002-28 March 2017.



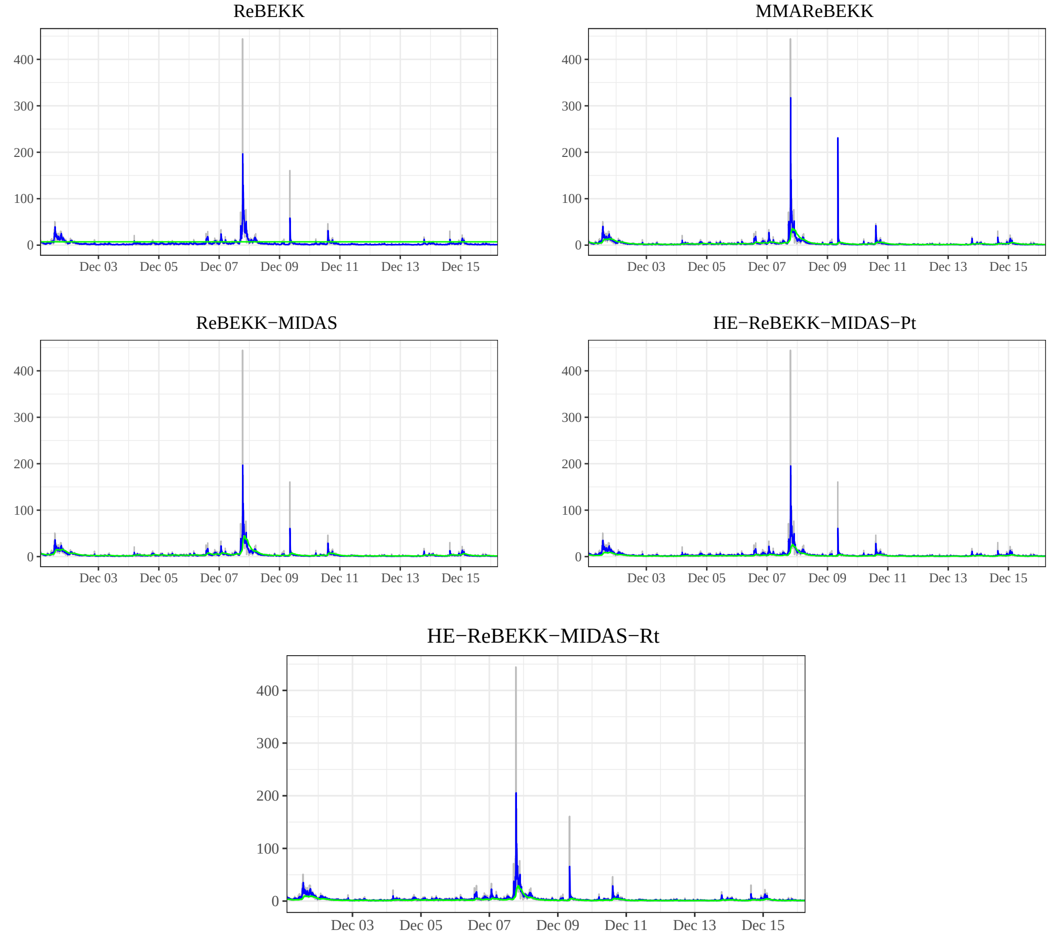
Notes: Realized Variance (gray line), Conditional Variance (blue line), Long run component (green line).

same period, in the graph (3.6) we can notice the high degree of the correlation among the assets. The latter is obtained from variances and covariances through the following ex-post transformation:

$$R_t = diag(H_t)^{-1/2} H_t diag(H_t)^{-1/2} \tag{3.17}$$

where $diag(H_t)$ is a diagonal matrix whose elements are the diagonal elements of the conditional covariance matrix, H_t. Notice that in the graph (3.6), it seems that the Hadamard exponential models do not adequately represent the time-varying average level: this is probably due to the approximation used for the intercept term, that is $(1 - \bar{\alpha}_t - \beta)M_t$ in place of $M_t \odot (J_n - A_t - bJ_n)$. According to Hafner and Franses (2009), this causes a bias in the unconditional covariance matrix, so it cannot be interpreted as the time-varying average level. This is the price to pay for a more flexible model.

FIGURE 3.5: Estimated Conditional Variance of HD. Sample Period: 28 January 2002-28 March 2017.



Notes: Realized Variance (gray line), Conditional Variance (blue line), Long run component (green line).

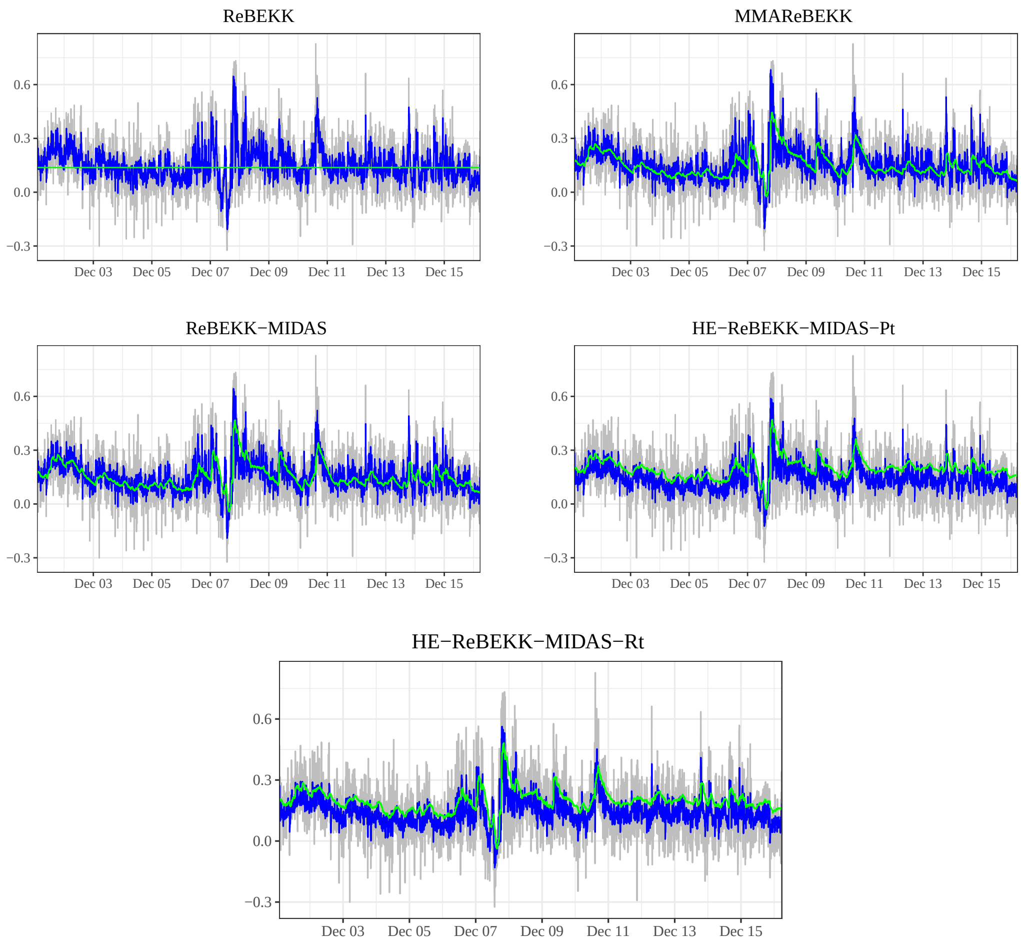
3.4.4 Out of sample analysis

An out-of-sample exercise is conducted for the period between 29 March 2017 and 16 April 2018. For this purpose, we generate 264 one-step-ahead forecasts of the covariance matrix, based on the parameter estimates obtained in Section (3.4.2).

Overall, as we can see from fig. (3.7), all the models seem to provide a good forecast of the covariance, also during turbulent periods.

Then, we compare the predictive performance of the estimated models, as for the in-sample comparison, through the MCS procedure. As we can see from the tab. (3.6), the out-of-sample analysis confirms the best performance of the HEReBEKK-MIDAS-R_t: it enters the set of superior models for both

FIGURE 3.6: Estimated Conditional Correlation among HD and XOM. Sample Period: 28, January 2002-28 March 2017.



Notes: Realized Covariance (gray line), Conditional Covariance (blue line), Long run component (green line).

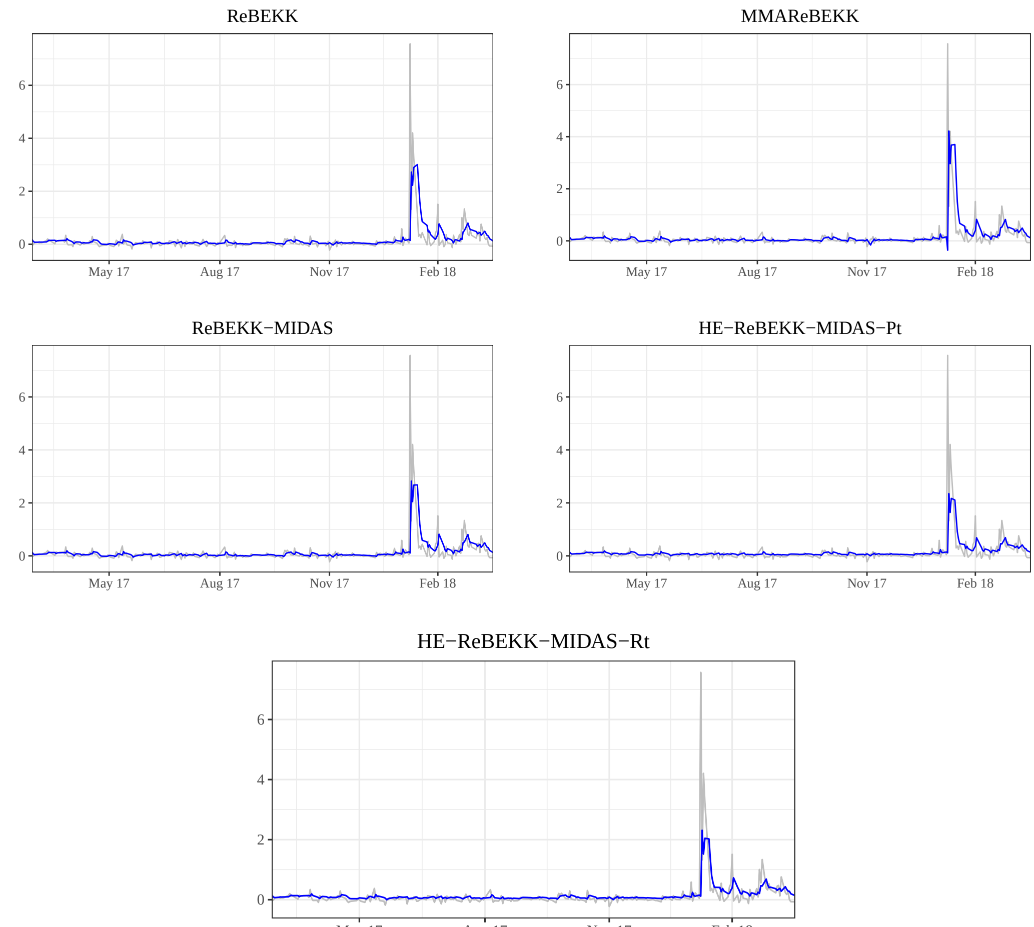
TABLE 3.6: Out of sample exercise

	ReBEKK	MMAReBEKK	ReBEKK-MIDAS	HEReBEKK-MIDAS-P_t	HEReBEKK-MIDAS-R_t
Qlik	12.123	14.883	12.05	12.055	**12.037**
Fn	**186.038**	**190.969**	**175.827**	**175.073**	**174.771**

Notes: QLIK: Quasi-Likelihood function. FN: squared Frobenius Norm. In bold, the models entering the MCS, $\alpha = 0.25$. Covariance Proxy: annualized Realized Covariance matrix. Out-of-Sample period: 29 March 2017-16 April, 2018.

loss functions, and more specifically it is the unique model for the QLIKE function. This is a not so obvious result: it is a recurrent feature in empirical applications that the most sophisticated models get a better in-sample fitting, while the simplest models tend to have a better out-of-sample performance (Hansen, 2010).

FIGURE 3.7: Forecasted Covariance among HD and XOM. Out-of-sample Period: 29 March 2017-16 April 2018.

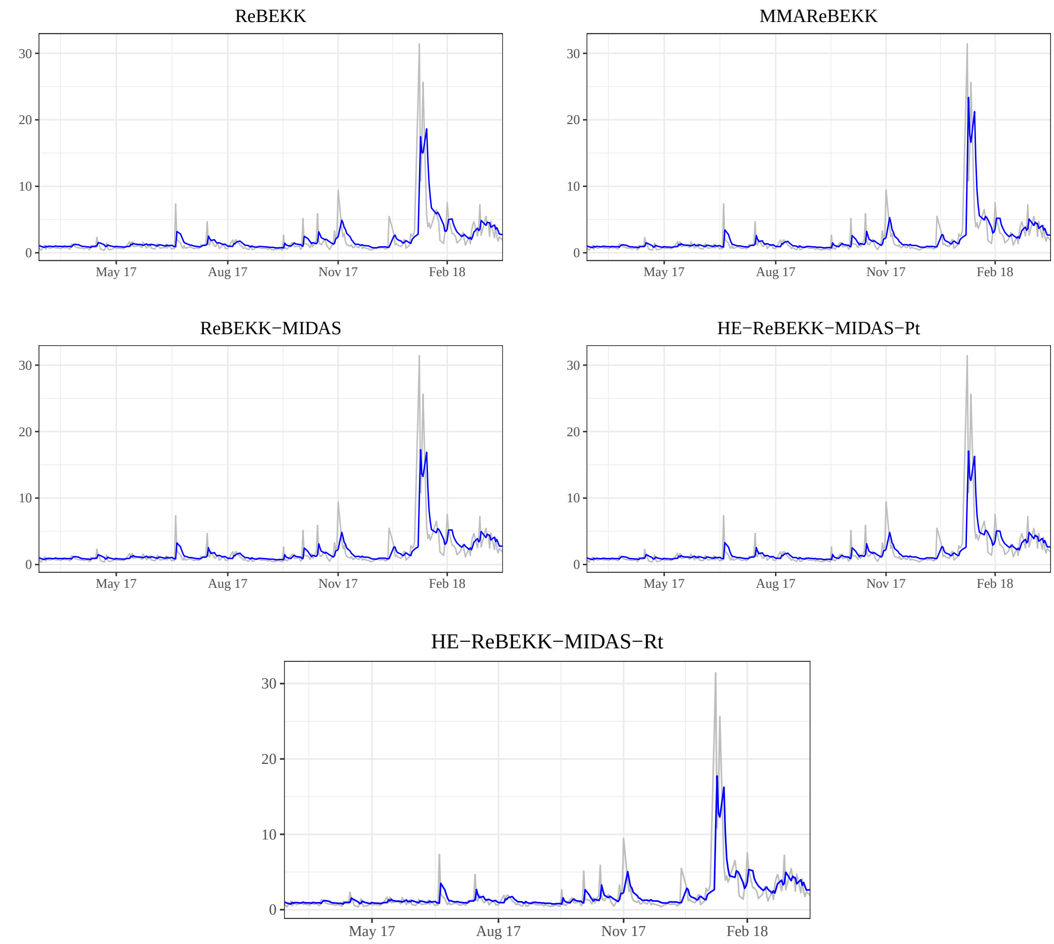


Notes: Realized Covariance (gray line), Forecasted Covariance (blue line).

3.5 Concluding Remarks

In this chapter, we propose a new multivariate component volatility model for Realized Covariance matrices, called ReBEKK-MIDAS, that, differently from other component models usually employed in literature, is insensible to the order of the assets and is easier to estimate from a computational point of view. More in detail, the long-run component of the new model is specified as a time-varying intercept, which is a function of past monthly Realized Covariance matrices through the use of the MIDAS filter. Furthermore, also an extension of the basic model is provided, by including the Hadamard exponential extensions in the spirit of Bauwens and Otranto (2020a), call it Hadamard Exponential Realzed BEKK MIDAS (HEReBEKK-MIDAS). This specification admits asset-pair specific and time-varying impact coefficients,

FIGURE 3.8: Forecasted Variance of HD. Out-of-sample Period: 29 March 2017-16 April 2018.



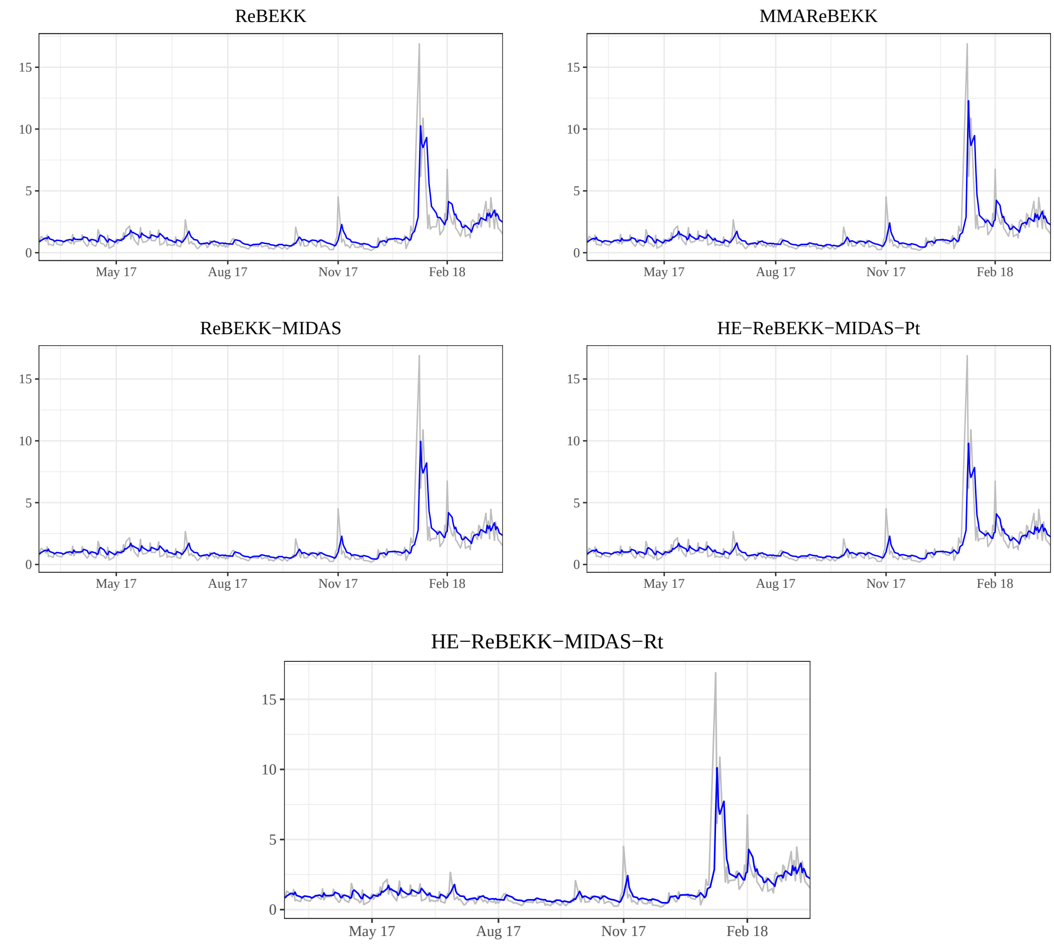
Notes: Realized Variance (gray line), Forecasted Variance (blue line).

only with one parameter more than the basic model, thus avoiding the proliferation of parameters.

The proposed model, more specifically that parameterized through the Hadamard Exponential function, has the best in-sample performance according to the information criteria. Moreover, the in-sample performance is also evaluated by applying the Model Confidence Set procedure for two loss functions, the Quasi Likelihood, and the Squared Frobenius Norm. The models that use the Hadamard exponential function, always enter the set of superior models, thus confirming their better in-sample performance. From an out-of-sample point of view, the models are compared with the same loss functions employed for the in-sample analysis. Results confirm the better performance of the models that allow time-varying and asset-pair specific parameters also for the forecasting exercise.

As future research, it could be interesting to consider other drivers of the

FIGURE 3.9: Forecasted Variance of XOM. Out-of-sample Period: 29 March 2017-16 April 2018.



Notes: Realized Variance (gray line), Forecasted Variance (blue line).

long-run component, such as macroeconomic variables, to analyze the relationship, in a multivariate setting, between economics and financial volatility. Another issue is related to the number of assets: indeed, for Realized Covariance matrices of large dimensions, we face the course of dimensionality problem. More specifically, the constant parameter matrix of the long-run component is a growing function of the number of assets considered. So, if we eliminate the constant parameter matrix from the likelihood function, for example by adopting the Iterative Moment-based Profiling (IMP) algorithm of Bauwens, Braione, and Storti (2017), the model could be estimated for high-dimensional covariance matrices. Furthermore, we can compare the covariance forecasts of the models employed in the empirical analysis, through some economic loss functions such as the Global Minimum Variance (GMV) Portfolio of Engle and Colacito (2006). Indeed, a model with superior covariance forecasts should provide a portfolio with a lower variance (Engle &

Kelly, 2012).

Appendix A

The fourth moment of a GARCH(1,1)

Let us consider a GARCH(1,1) process:

$$\begin{aligned} \epsilon_t &= \eta_t \sqrt{h_t} \qquad \eta_t \sim N(0,1)\ \forall t \\ h_t &= \omega + \alpha \epsilon_{t-1}^2 + \beta h_{t-1}^2 \end{aligned} \tag{A.1}$$

The conditional variance, h_t, in its long-run solution, can be specified as follows:

$$h_t = \sigma^2(1-\alpha-\beta) + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}^2$$

So, the fourth moment of ϵ_t is equal to:

$$E(\epsilon_t^4) = E(h_t^2 \eta_t^4) = 3E(h_t^2) \tag{A.2}$$

Then, let us consider the square of h_t:

$$\begin{aligned} h_t^2 &= \sigma^4(1-\alpha-\beta)^2 + \alpha^2\epsilon_{t-1}^4 + \beta^2 h_{t-1}^2 + 2\sigma^2(1-\alpha-\beta)\alpha\epsilon_{t-1}^2 + \\ &+ 2\sigma^2(1-\alpha-\beta)\beta h_{t-1} + 2\alpha\beta\epsilon_{t-1}^2 h_{t-1} \\ h_t^2 &= \sigma^4(1-\alpha-\beta)(1-\alpha-\beta) + \alpha^2\eta_{t-1}^4 h_{t-1}^2 + \beta^2 h_{t-1}^2 + \\ &+ 2\sigma^2(1-\alpha-\beta)\alpha\eta_{t-1}^2 h_{t-1} + 2\sigma^2(1-\alpha-\beta)\beta h_{t-1} + 2\alpha\beta\eta_{t-1}^2 h_{t-1}^2 \end{aligned} \tag{A.3}$$

Now, let us calculate the expected value of eq. (A.3):

$$\begin{aligned} &E(h_t^2) = \sigma^4(1-\alpha-\beta)(1-\alpha-\beta) + 3\alpha^2 E(h_{t-1}^2) + \beta^2 E(h_{t-1}^2) + 2\sigma^4(1-\alpha-\beta)\alpha + \\ &+ 2\sigma^4(1-\alpha-\beta)\beta + 2\alpha\beta E(h_{t-1}^2) \\ &E(h_t^2)(1-3\alpha^2-\beta^2-2\alpha\beta) = \sigma^4(1-\alpha-\beta)(1+\alpha+\beta) \\ &E(h_t^2) = \frac{\sigma^4(1-\alpha-\beta)(1+\alpha+\beta)}{1-(3\alpha^2+\beta^2+2\alpha\beta)} \end{aligned} \tag{A.4}$$

Finally, by substituting (A.4) above in eq. (A.2) one obtains:

$$E(\epsilon_t^4) = 3\frac{\sigma^4(1-\alpha-\beta)(1+\alpha+\beta)}{1-(3\alpha^2+\beta^2+2\alpha\beta)} \tag{A.5}$$

Notice that the fourth moment exists if $2\alpha^2 + (\alpha+\beta)^2 < 1$.

Appendix B

Unconditional Variance of a MEM process

The unconditional variance of a MEM (1,1) process can be expressed as follows:

$$Var\,(x_t) = E\,[Var\,(x_t|\mathcal{I}_{t-1})] + Var\,[E(x_t|\mathcal{I}_{t-1})] \tag{B.1}$$

Let us consider the alternative parameterization of the conditional mean equation[1]:

$$\mu_t = \mu(1-\alpha-\beta) + \alpha x_{t-1} + \beta\mu_{t-1} \tag{B.2}$$

If we square (B.2) we obtain:

$$\begin{aligned}\mu_t^2 &= \mu^2(1-\alpha-\beta)(1-\alpha-\beta) + \alpha^2\mu_{t-1}^2\epsilon_{t-1}^2 + \beta^2\mu_{t-1}^2 + 2\mu(1-\alpha-\beta)\alpha\mu_{t-1}\epsilon_{t-1} + \\ &+ 2\mu(1-\alpha-\beta)\beta\mu_{t-1} + 2\alpha\beta\mu_{t-1}^2\epsilon_{t-1}\end{aligned} \tag{B.3}$$

Then, let us take the expectation of (B.3):

$$\begin{aligned}E(\mu_t^2) &= \mu^2(1-\alpha-\beta)(1-\alpha-\beta) + \alpha^2(1+1/a)E(\mu_{t-1}^2) + \beta^2 E(\mu_{t-1}^2) + \\ &+ 2\mu^2(1-\alpha-\beta)\alpha + 2\mu^2(1-\alpha-\beta)\beta + 2\alpha\beta E(\mu_{t-1}^2)\end{aligned}$$

Remember that if the process is covariance stationary $E(\mu_t^2) = E(\mu_{t-1}^2) = \ldots = E(\mu_{t-\tau}^2)\ \forall\ \tau$. Then, by rearranging the equation above we get:

$$E(\mu_t^2) = \mu^2 \frac{(1-\alpha-\beta)(1+\alpha+\beta)}{1-[(1+1/a)\alpha^2 + 2\alpha\beta + \beta^2]} \tag{B.4}$$

[1] For the sake of simplicity, I do not consider the asymmetric term, but the results can be easily extended to this case.

Notice that the second moment of μ_t exists if $(1/a)\alpha^2 + (\alpha+\beta)^2 < 1$. The conditional variance of the process is expressed as:

$$Var\,(x_t|\mathcal{I}_{t-1}) = \frac{\mu_t^2}{a} \tag{B.5}$$

Then, let us consider the expectation of (B.5), that corresponds to the first term of (B.1):

$$E\left[Var\,(x_t|\mathcal{I}_{t-1})\right] = \frac{E(\mu_t^2)}{a} = \frac{1}{a}\mu^2\frac{(1-\alpha-\beta)(1+\alpha+\beta)}{1-[(1+1/a)\alpha^2+2\alpha\beta+\beta^2]} \tag{B.6}$$

Now, we can specify the second term of (B.1) as follows:

$$Var\left[E(x_t|\mathcal{I}_{t-1})\right] = Var(\mu_{t-1}) = Var(\mu_t) = E(\mu_t^2) - \mu^2 \tag{B.7}$$

Let us substitute (B.4) into (B.7)

$$Var\left[E(x_t|\mathcal{I}_{t-1})\right] = \mu^2\frac{(1-\alpha-\beta)(1+\alpha+\beta)}{1-[(1+1/a)\alpha^2+2\alpha\beta+\beta^2]} - \mu^2 \tag{B.8}$$

Then, by substituting (B.6) and (B.8) into (B.1) we obtain:

$$\begin{aligned} Var(x_t) &= (1+1/a)\mu^2\frac{(1-\alpha-\beta)(1+\alpha+\beta)}{1-[(1+1/a)\alpha^2+2\alpha\beta+\beta^2]} - \mu^2 = \\ &= (1+1/a)\mu^2\frac{1-(\alpha+\beta)^2}{1-[(\alpha+\beta)^2+(1/a)\alpha^2]} - \mu^2 \end{aligned}$$

Finally, by rearranging the above equation one obtains:

$$Var(x_t) = \mu^2\frac{(1/a)(1-2\alpha\beta-\beta^2)}{1-[(\alpha+\beta)^2+(1/a)\alpha^2]}$$

Appendix C

Regime inference

C.1 Hamilton filter

The optimal inference on regimes depends on the information set we use, indeed if it is based on the observed data up to time $t-1$, we have the predicted probability, $P(s_t = j|\mathcal{I}_{t-1})$, with $\mathcal{I}_{t-1} = \{x_{t-1}, x_{t-2}, ..., x_1\}$, up time t we have the filtered probability, $P(s_t = j|\mathcal{I}_t)$, and through the full sample observations, T, we have the smoothed probability, $P(s_t = j|\mathcal{I}_T)$. The Hamilton filter is an iterative algorithm that computes the predicted probabilities and filtered probabilities, and as a by-product gives us the likelihood function to be maximized numerically, due to its nonlinear structure. Let us consider a MS-MEM(1,1) process:

$$\begin{aligned} x_t &= \mu_{t,s_t}\epsilon_t \\ \epsilon_t|s_t &\sim Gamma\left(a_{s_t}, \frac{1}{a_{s_t}}\right) \ \forall t \\ \mu_{t,s_t} &= \omega_{s_t} + \alpha x_{t-1} + \beta\mu_{t-1,s_{t-1}} + \gamma\mathbb{1}_{(r_{t-1}<0)}x_{t-1} \end{aligned} \tag{C.1}$$

The joint density of x_t and s_t, s_{t-1} conditional on past information, that is $f(x_t, s_t, s_{t-1}|\mathcal{I}_{t-1})$, is given by the product of the conditional and the marginal density :

$$f(x_t, s_t, s_{t-1}|\mathcal{I}_{t-1}) = f(x_t|\mathcal{I}_{t-1}, s_t, s_{t-1})P(s_t, s_{t-1}|\mathcal{I}_{t-1}) \tag{C.2}$$

where $f(x_t|\mathcal{I}_t, s_t, s_{t-1}) = \left[\dfrac{a_{s_t}{}^{a_{s_t}} + \mu_{t,s_t,s_{t-1}}^{-a_{s_t}} + x_t^{a_{s_t}-1}\exp\left(-a_{s_t}\frac{x_t}{\mu_{t,s_t,s_{t-1}}}\right)}{\Gamma(a_{s_t})}\right]$.

The marginal density of x_t, that is $f(x_t|\mathcal{I}_{t-1})$, is given by the sum of $f(x_t, s_t, s_{t-1}|\mathcal{I}_{t-1})$

over all the possible values of s_t and s_{t-1} :

$$
\begin{aligned}
f(x_t|\mathcal{I}_{t-1}) = & \\
&= \sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} f(x_t, s_t, s_{t-1}|\mathcal{I}_{t-1}) \\
&= \sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} f(x_t|\mathcal{I}_{t-1}, s_t, s_{t-1}) P(s_t, s_{t-1}|\mathcal{I}_{t-1})
\end{aligned}
\tag{C.3}
$$

then the loglikelihood function, LL, is equal to:

$$
LL = \sum_{1}^{T} ln\left[f(x_t|\mathcal{I}_{t-1})\right] = \sum_{1}^{T} ln\left[\sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} f(x_t|\mathcal{I}_{t-1}, s_t, s_{t-1}) P(s_t, s_{t-1}|\mathcal{I}_{t-1})\right]
\tag{C.4}
$$

All we need to iterate the algorithm are the predicted probabilities, $P(s_t = j, s_{t-1} = i|\mathcal{I}_{t-1})$. Notice that the latter are equal to the product of the transition probabilities, $P(s_t|s_{t-1})$[1], and the filtered probabilities, $P(s_{t-1}|\mathcal{I}_{t-1})$:

$$
P(s_t, s_{t-1}|\mathcal{I}_{t-1}) = P(s_t|s_{t-1}) P(s_{t-1}|\mathcal{I}_{t-1})
\tag{C.5}
$$

Then the input of the algorithm, at each point in time, t, is $P(s_{t-1}|\mathcal{I}_{t-1})$, that we obtain as output at $t-1$, in fact:

$$
\begin{aligned}
P(s_t, s_{t-1}|\mathcal{I}_t) = & \\
&= P(s_t, s_{t-1}|\mathcal{I}_{t-1}, x_t) \\
&= \frac{f(x_t, s_t, s_{t-1}|\mathcal{I}_{t-1})}{f(x_t|\mathcal{I}_{t-1})} \\
&= \frac{f(x_t|\mathcal{I}_{t-1}, s_t, s_{t-1}) P(s_t, s_{t-1}|\mathcal{I}_{t-1})}{\sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} f(x_t|\mathcal{I}_{t-1}, s_t, s_{t-1}) P(s_t, s_{t-1}|\mathcal{I}_{t-1})}
\end{aligned}
\tag{C.6}
$$

with $P(s_t|\mathcal{I}_t) = \sum_{i=1}^{N} P(s_t = j, s_{t-1} = i|\mathcal{I}_{t-1})$. Then at the next iteration $(t+1)$ we use $P(s_t|\mathcal{I}_t)$ as input. So, we can iterate for $t = 1, .., T$. Notice that at $t = 1$ we need $P(s_0|\mathcal{I}_0)$. For this purpose the ergodic probabilities, π, are usually used as starting value, under the hypothesis of stationarity and ergodicity. They are the long run forecast, that is the unconditional probability of each state and they are calculated as follows[2]:

$$
\pi = (A'A)^{-1} A' e_{N+1}
\tag{C.7}
$$

[1]Remember that, the probability that s_t is equal to a certain value depends only on the value of s_{t-1}, in a first order Markov chain.

[2]See, Hamilton (1994), ch. 22.

with $A = \begin{bmatrix} I_N - P \\ 1' \end{bmatrix}$, e_{N+1} is the $(N+1)$th column of the order $(N+1)$ identity matrix, 1 is an order N column vector of ones, while P is the transition probability matrix.

C.2 Kim's Algorithm

The regime inference could be carried out based on the full sample observations, an information set different from that of filtered and predicted probabilities. This kind of inference, called smoothed probability, $P(s_t = j|\mathcal{I}_T)$, is computed through the algorithm developed by Kim (1994). Let us consider the joint probability of s_t and s_{t+1}:

$$\begin{aligned} P(s_t, s_{t+1}|\mathcal{I}_T) = & \\ & = P(s_t|s_{t+1}, \mathcal{I}_T)P(s_{t+1}|\mathcal{I}_T) \\ & = P(s_t|s_{t+1}, \mathcal{I}_t)P(s_{t+1}|\mathcal{I}_T) \\ & = \frac{P(s_t, s_{t+1}|\mathcal{I}_t)P(s_{t+1}|\mathcal{I}_T)}{P(s_{t+1}|\mathcal{I}_t)} \\ & = \frac{P(s_{t+1}|s_t)P(s_t|\mathcal{I}_t)P(s_{t+1}|\mathcal{I}_T)}{P(s_{t+1}|\mathcal{I}_t)} \end{aligned} \tag{C.8}$$

with $P(s_t|\mathcal{I}_T) = \sum_{s_{t+1}} P(s_t, s_{t+1}|\mathcal{I}_T)$. Notice that at the last iteration of the Hamilton filter, $t = T$, we obtain $P(s_T|\mathcal{I}_T)$ then we can iterate eq. (C.8) for $t = T-1, ..., 1$. In order to consider the equality from the second line to the third one in eq. (C.8), let us consider the first term of the second line:

$$\begin{aligned} P(s_t|s_{t+1}, \mathcal{I}_T) = & \\ & = P(s_t|s_{t+1}, \mathcal{I}_{t+1:T}\mathcal{I}_t) \\ & = \frac{P(s_t, \mathcal{I}_{t+1:T}|s_{t+1}, \mathcal{I}_t)}{P(\mathcal{I}_{t+1:T}|s_{t+1}, \mathcal{I}_t)} \\ & = \frac{P(\mathcal{I}_{t+1:T}|s_t, s_{t+1}, \mathcal{I}_t)P(s_t|s_{t+1}, \mathcal{I}_t)}{P(\mathcal{I}_{t+1:T}|s_{t+1}, \mathcal{I}_t)} \\ & = P(s_t|s_{t+1}, \mathcal{I}_t) \end{aligned}$$

where $\mathcal{I}_{t+1:T} = \{x_{t+1}, ..., x_T\}$. The equality from the fourth to the fifth line holds if $P(\mathcal{I}_{t+1:T}|s_t, s_{t+1}, \mathcal{I}_t) = P(\mathcal{I}_{t+1:T}|s_{t+1}, \mathcal{I}_t)$, that is when the conditional density, $f(x_{t+1}|s_t, s_{t+1}|\mathcal{I}_{t-1})$, depends only on the current regime, s_{t+1}, otherwise it will be only an approximation.

Appendix D

Positive Definiteness of the time-varying intercept A_t

Let us consider the dynamics of the conditional covariance matrix, S_t, of the HEReBEKKMIDAS discussed in sec. 3.2.1:

$$S_t = M_t(1 - \bar{\alpha}_t - \bar{\beta}_t) + A_t \odot C_{t-1} + B_t \odot S_{t-1} \tag{D.1}$$

where A_t:

$$A_t = \alpha \, exp^{\odot}[\phi_A(N_{t-1} - J_n)] = \alpha \frac{exp^{\odot}[\phi_A(N_{t-1})]}{exp(\phi_A)} \tag{D.2}$$

with ϕ_A and α two non-negative scalar coefficients, J_n an order n square matrix of ones, and N_{t-1} the Realized Correlation matrix, P_t, or the Conditional one, R_t. Let us demonstrate that A_t is a Positive Definite (PD) matrix: since ϕ_A is non-negative, $\phi_A N_{t-1}$ is PD[1], then $exp^{\odot} N_{t-1}$ is Positive Semidefinite (PSD) by applying the Lemma 1(b) in Bauwens and Otranto (2020b). In addition, we can notice that the diagonal elements of N_{t-1} are equal to 1, while the off-diagonal are between -1 and 1, then, by applying the lemma1(c) in Bauwens and Otranto (2020b), the matrix $exp^{\odot} \phi_A N_{t-1}$ is also Positive Definite. Finally, multiplying $exp^{\odot} N_{t-1}$ by the non-negative scalar $\frac{\alpha}{exp(\phi_A)}$, the resulting matrix $\alpha \frac{exp^{\odot}[\phi_A(N_{t-1})]}{exp(\phi_A)}$ is demonstrated to be PD. More in details, A_t is a PD matrix with diagonal elements equal to 1, while the off-diagonal ones are between 0 and α.

[1]Remember that N_{t-1} is a correlation matrix, then it is PD by construction.

Bibliography

Amendola, A., Candila, V., Cipollini, F., & Gallo, G. M. (2020). Doubly multiplicative error models with long–and short–run components. *Technical Report*.

Amendola, A., Candila, V., & Gallo, G. M. (2019). On the asymmetric impact of macro–variables on volatility. *Economic Modelling*, *76*, 135–152.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, *39*(4), 885–905.

Asgharian, H., Hou, A. J., & Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. *Journal of Forecasting*, *32*(7), 600–612.

Augustyniak, M., Boudreault, M., & Morales, M. (2018). Maximum likelihood estimation of the Markov-switching GARCH model based on a general collapsing procedure. *Methodology and Computing in Applied Probability*, *20*(1), 165–188.

Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30.

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.

Bauwens, L., Braione, M., & Storti, G. (2016). Forecasting comparison of long term component dynamic models for realized covariance matrices. *Annals of Economics and Statistics*, (123/124), 103–134.

Bauwens, L., Braione, M., & Storti, G. (2017). A dynamic component model for forecasting high-dimensional realized covariance matrices. *Econometrics and Statistics*, *1*, 40–61.

Bauwens, L., Laurent, S., & Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, *21*(1), 79–109.

Bauwens, L., & Otranto, E. (2020a). Modelling realized covariance matrices: A class of Hadamard exponential models. *Technical Report*.

Bauwens, L., & Otranto, E. (2020b). Nonlinearities and regimes in conditional correlations with different dynamics. *Journal of Econometrics*, *217*(2), 496–522.

Bauwens, L., Storti, G., & Violante, F. (2012). Dynamic conditional correlation models for realized covariance matrices. *CORE DP*, *60*.

Bernardi, M., & Catania, L. (2018). The model confidence set package for R. *International Journal of Computational Economics and Econometrics*, *8*(2), 144–158.

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *The Review of Economics and Statistics*, *72*(3), 498–505.

Bollerslev, T., Engle, R. F., & Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, *96*(1), 116–131.

Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, *11*(2), 143–172.

Brownlees, C. T., Cipollini, F., & Gallo, G. M. (2012). Multiplicative error models. In *Handbook of Volatility Models and Their Applications* (pp. 223–247). John Wiley & Sons.

Brownlees, C. T., & Gallo, G. M. (2010). Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, *8*(1), 29–56.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of financial Economics*, *31*(3), 281–318.

Colacito, R., Engle, R. F., & Ghysels, E. (2011). A component model for dynamic correlations. *Journal of Econometrics*, *164*(1), 45–59.

Comte, F., & Lieberman, O. (2003). Asymptotic theory for multivariate GARCH processes. *Journal of Multivariate Analysis*, *84*(1), 61–84.

Conrad, C., & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, *30*(7), 1090–1114.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196.

Ding, Z., & Engle, R. F. (2001). Large scale conditional covariance matrix modeling, estimation and testing. *NYU Working Paper No. FIN-01-029*.

Ding, Z., & Granger, C. W. (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics*, *73*(1), 185–215.

Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*(1), 83–106.

Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics*, *15*(1), 26–34.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.

Engle, R. F. (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. F. (2002b). New frontiers for ARCH models. *Journal of Applied Econometrics*, *17*(5), 425–446.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, *5*(1), 1–50.

Engle, R. F., & Colacito, R. (2006). Testing and valuing dynamic correlations for asset allocation. *Journal of Business & Economic Statistics*, *24*(2), 238–253.

Engle, R. F., & Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, *131*(1-2), 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., & Kelly, B. (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics*, *30*(2), 212–228.

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

Engle, R. F., & Lee, G. (1999). A long-run and short-run component model of stock return volatility. In *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger* (pp. 475–497). Oxford University Press.

Engle, R. F., & Rangel, J. G. (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The Review of Financial Studies*, *21*(3), 1187–1222.

Engle, R. F., & Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, *66*(5), 1127–1162.

Francq, C., & Zakoian, J.-M. (2019). *GARCH models: Structure, statistical inference and financial applications*. John Wiley & Sons.

Gallo, G. M., & Otranto, E. (2015). Forecasting realized volatility with changing average levels. *International Journal of Forecasting*, *31*(3), 620–634.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, *26*(1), 53–90.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, *48*(5), 1779–1801.

Golosnoy, V., Gribisch, B., & Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility. *Journal of Econometrics*, *167*(1), 211–223.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27–62.

Guérin, P., & Marcellino, M. (2013). Markov-switching MIDAS models. *Journal of Business & Economic Statistics*, *31*(1), 45–56.

Haas, M., Mittnik, S., & Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, *2*(4), 493–530.

Hafner, C. M., & Franses, P. H. (2009). A generalized dynamic conditional correlation model: Simulation and application to many assets. *Econometric Reviews*, *28*(6), 612–631.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, *64*(1-2), 307–333.

Hansen, P. R. (2010). A winner's curse for econometric models: On the joint distribution of in-sample fit and out-of-sample fit and its implications for model selection. *Technical Report*.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, *60*(1-2), 1–22.

Kim, C.-J., & Nelson, C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press.

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, *8*(2), 225–234.

Mandelbrot, B. B. (1963). The variation of certain speculative prices. *The Journal of Business*, *36*(3), 394–419.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, *8*(4), 377–399.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1, 1) model. *Econometric Theory*, *6*(3), 318–334.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370.

Officer, R. R. (1973). The variability of the market factor of the New York Stock Exchange. *the Journal of Business*, *46*(3), 434–453.

Otranto, E. (2015). Capturing the spillover effect with multiplicative error models. *Communications in Statistics - Theory and Methods*, *44*(15), 3173–3191.

Pan, Z., Wang, Y., Wu, C., & Yin, L. (2017). Oil price volatility and macroeconomic fundamentals: A regime switching GARCH-MIDAS model. *Journal of Empirical Finance*, *43*, 130–142.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, *160*(1), 246–256.

Patton, A. J., & Sheppard, K. (2009). Evaluating volatility and correlation forecasts. In *Handbook of Financial Time Series* (pp. 801–838). Springer.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, *44*(5), 1115–1153.

Timmermann, A. (2000). Moments of Markov switching models. *Journal of Econometrics*, *96*(1), 75–111.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, *50*(1), 1–25.