Forecasting realized volatility: a review

Andrea Bucci

Università Politecnica delle Marche

December 2017

Forecasting Realized Volatility: a Review

Andrea Bucci*

Abstract

Modeling financial volatility is an important part of empirical finance. This paper provides a literature review of the most relevant volatility models, with a particular focus on forecasting models. We firstly discuss the empirical foundations of different kinds of volatility. The paper, then, analyses the non-parametric measure of volatility, named realized variance, and its empirical applications. A wide range of realized volatility models, both univariate and multivariate, is presented, such as time series models, MIDAS and GARCH-MIDAS models, Realized GARCH, and HEAVY models. We further discuss forecasting evaluation methods specifically suited for volatility models.

JEL classification: C22, C53, G10
Keywords: Realized Volatility, Stochastic Volatility, Volatility Models

*Department of Economics and Social Sciences, Università Politecnica delle Marche, 60121 Ancona, Italy. email: a.bucci@univpm.it

1 Introduction

Financial econometric literature has been focused, over the last two decades, on modelling and forecasting volatility. Since volatility as risk measure is largely used in asset allocation, risk management and option pricing, and since it cannot be a priori determined, the definition of a good proxy of volatility has become extremely relevant in this context. Hence, a wide range of econometric literature has focused on estimating the latent conditional variance.

In the early steps, the bulk of volatility models has been based on the (Generalized) Autoregressive Conditional Heteroskedasticity models, by Engle (1982) and Bollerslev (1986), and stochastic volatility (SV) models.

Given the growing availability of high-frequency data, researchers have moved their attention to alternative non-parametric measures of volatility based on such kind of data. The first attempt of using high-frequency data for measuring volatility was made by Merton (1980), who noted that the conditional variance can be computed as the sum of squared returns sampled at sufficiently high frequency. The recent theoretical findings on the informative content of intra-day data have stimulated an important stream of literature on the properties of non-parametric measures of volatility. In fact, Andersen, Bollerslev, Diebold, and Labys (2000); Andersen, Bollerslev, Diebold, and Ebens (2001) showed that ex-post volatility based on higher frequency data successfully measures underlying return variability. Barndorff-Nielsen and Shephard (2002a,b) provided the theoretical foundation of using realized volatility as a proxy of the latent volatility based on the theory of quadratic variation. In this way, volatility becomes observable and may be modelled directly through traditional time series model.

The aim of this paper is to provide a review of theoretical foundations and empirical applications of realized volatility (RV). Contrary to already published reviews on RV, this paper is mainly focused on practical applications of realized volatility models, with a particular mention of forecasting performance. In order to provide a comprehensive review of RV models, this article also mentions Realized GARCH models, MIDAS model and non-linear models on volatility, which have never been treated in similar papers. Poon and Granger (2003), Andersen, Bollerslev, Christoffersen, and X. (2006), McAleer and Medeiros (2008b) and Bandi and Russell (2006) have also reviewed the RV literature. However, the first two articles do not consider microstructure noise, while the other two works mentioned do focus on theoretical properties and not on empirical applications. We aim at providing a totally comprehensive overview of the relevant models for forecasting RV, in order to inspire possible alternative models overcoming the still existing pitfalls of this particular econometric literature.

This article overviews also the developments on volatility forecasts evaluation and comparison. We analyse self-standing measures, pairwise comparison tests, like the tests proposed by Diebold and Mariano (1995), West (1996) and Giacomini and White (2006), and multiple comparison meth-

ods, like the Model Confidence Set introduced by Hansen, Lunde, and Nason (2011). In most of these direct methods of forecast accuracy, the evaluation of the forecasts relies on the ordering imposed by a statistical loss function. In this article, we discuss the properties of a number of admissible loss functions, both in the univariate and multivariate framework.

The paper is divided as follows: in the first part we provide an overview of the theoretical foundation for the definition of volatility; the second part is a comprehensive review of parametric models, based on a functional form of the expected and instantaneous volatility. The third part concerns non-parametric approaches to model volatility, focusing on realized volatility and its empirical applications, while the last part of the paper discusses forecasting evaluation methods.

1.1 Theoretical foundation

It is known that, in the financial market, negotiations take place at extremely short intervals, and that stock prices may be modelled as continuous processes. We then assume that $p(t)$ is the univariate process of the logarithmic price, defined in a probability space (Ω, I, P), evolving in the continuous time on an interval $[0,T]$, where T is an integer, and that the entire available information is given by $(I_t)_{t\in[0,T]} \subseteq I$.

Assuming the absence of arbitrage and a finite first moment, the price process belongs to the class of special semimartingale[1], as defined in Back (1991) and Shiryaev (1999). The class of semimartingale is particularly relevant in econometrics, since it includes processes like martingale and Lévy process (see also Protter (1992)). The log price process $p(t)$, with finite mean, is a semimartingale if it can be decomposed as the sum of a drift component and a local martingale which can be further decomposed in a realization of a continuous process and a jump component, such that

$$p(t) = p(0) + A(t) + M(t) = p(0) + A^C(t) + \Delta A(t) + M^C(t) + \Delta M(t), \tag{1.1}$$

where $A(0) \equiv M(0) \equiv 0$, $A^C(t)$ and $M^C(t)$ are the realizations of the continuous process, $\Delta A(t)$ and $\Delta M(t)$ are the relative jump components.

If the compound return in the interval $[t-h,t]$, for $0 \le h \le t \le T$ is defined as

$$r(t,h) = p(t) - p(t-h); \tag{1.2}$$

and given that in $[0,t]$ it may also be specified as

$$r(t) \equiv r(t,t) = p(t) - p(0), \tag{1.3}$$

[1] A process X may be defined as a special semimartingale, if it may be written as $X = X_0 + A + M$, where $A_0 = M_0 = 0$, M is a local martingale and A is a predictable finite-variation process.

it follows that

$$\begin{aligned} r(t,h) &= p(t) - p(0) + p(0) - p(t-h) \\ &= r(t) - (p(t-h) - p(0)) \\ &= r(t) - r(t-h). \end{aligned} \tag{1.4}$$

It is further assumed that the asset price follows a finite and almost surely strictly positive process, so that $p(t)$ and $r(t)$ are well defined over $[0,T]$, and that $r(t)$ has only countably jump points over $[0,T]$.

Let the squares of the price and return processes be integrable, the cádlág version of the process is given by

$$r(t-) \equiv \lim_{\tau \to t, \tau < t} r(\tau) \tag{1.5}$$

$$r(t+) \equiv \lim_{\tau \to t, \tau > t} r(\tau) \tag{1.6}$$

$$r(t) = r(t+) \qquad a.s. \tag{1.7}$$

The jumps in the cumulative return process are

$$\Delta r(t) \equiv r(t) - r(t-), \qquad 0 \le t \le T. \tag{1.8}$$

At continuity points $\Delta r(t) = 0$, more generally

$$P(\Delta r(t) \neq 0) = 0 \qquad t \in [0,T]. \tag{1.9}$$

This does not imply that jumps are necessarily rare. There is the possibility of a (countably) infinite number of jumps over any discrete interval - a phenomenon referred to as an *explosion*.

As a consequence of the decomposition of a martingale, the return process is equal to

$$r(t) \equiv p(t) - p(0) = \mu(t) + M(t) = \mu(t) + M^C(t) + M^J(t). \tag{1.10}$$

The instantaneous return is decomposed in a predictable and finite variation process, $\mu(t)$, and a local martingale, $M(t)$, which is further decomposed in a continuous sample path, infinite variation local martingale, $M^C(t)$, and a compensated jump martingale, $M^J(t)$. The instantaneous return is, thus, decomposed into an expected return component and a (martingale) innovation.

1.2 Volatility: definition and theoretical aspects

Volatility is an index of unexpected variability of asset returns in a period. In this section we analyse the different definitions of volatility and the relations among them.

For each semimartingale $X(t)$ and for a couple of semimartingale, $X(t)$ and $Y(t)$, the quadratic variation and covariation of the processes, respectively $[X,X]_t$ and $[X,Y]_t$, for $t \in [0,T]$, can be defined as

$$[X,X]_t = X(t)^2 - 2\int_0^t X(s-)dX(s) \tag{1.11}$$

$$[X,Y]_t = X(t)Y(t) - \int_0^t X(s-)dX(s) - \int_0^t Y(s-)dY(s), \tag{1.12}$$

where the integral of the cádlág processes, $X(s-)$ e $Y(s-)$, is well defined. It follows that the quadratic variation, $[X,X]_t$, is a growing stochastic process.

The quadratic variation of a semimartingale has the following properties:

i. if τ_m is a partition of $[0,T]$, for $0 = \tau_{m,0} \leq \tau_{m,1} \leq \cdots \leq \tau_{m,m} = T$, such that $sup_{j\geq 0}(\tau_{m,j+1} - \tau_{m,j}) \to 0$ for $m \to \infty$, then

$$\lim_{m\to\infty} \left\{\Sigma_{j\geq 1}(X(t\wedge\tau_{m,j}) - X(t\wedge\tau_{m,j-1}))^2\right\} \to [X,X]_t, \tag{1.13}$$

where $t \wedge \tau \equiv \min(t,\tau)$ and the convergence is uniform. The quadratic variation process represents the (cumulative) realized sample-path variability of $X(t)$ over the $[0,t]$ time interval.

ii. if $X(t)$ and $Y(t)$ are square integrable semimartingale, the covariance between X and Y in $[t-h,t]$ is given by

$$Cov\left[X(t),Y(t) \mid I_{t-h}\right] = E([X,Y]_t \mid I_{t-h}) - [X,Y]_{t-h}; \tag{1.14}$$

iii. if the finite variation component in (1.1), A, is continuous, it follows that

$$[X_i,X_j]_t = [M_i,M_j] = [M_i^C,M_j^C] + \sum_{0\leq s\leq t} \Delta M_i(s)\Delta M_j(s). \tag{1.15}$$

Property (iii) shows that quadratic variation in continuous finite variation processes is zero, so that the mean component is irrelevant for the quadratic variation. If it assumed, without loss of generality, that log-price follows a diffusion process

$$dp(t) = \mu(t)dt + \sigma(t)dW(t), \tag{1.16}$$

where W is a Wiener process, $\mu(t)$ is a finite variation predictable process[2] and $\sigma(t)$ is a strictly positive and square integrable process, such that

$$P\left[\int_{t-h}^t \sigma^2(s)ds < \infty\right] = 1, \tag{1.17}$$

then the compound return over the interval $[t-h,t]$ is given by

$$r(t,h) = \mu(t,h) + M(t,h) = \int_{t-h}^t \mu(s)ds + \int_{t-h}^t \sigma(s)dW(s). \tag{1.18}$$

[2]At time t, a process is predictable is the value of the process is known an instant before t. Deterministic trends and cádlág processes are examples of predictable processes.

Accordingly, the quadratic variation can be computed as

$$QV_t = [p,p]_t - [p,p]_{t-h} = \int_{t-h}^{t} \sigma^2(s)ds. \tag{1.19}$$

This quantity, necessary for the definition of the realized variance, is also known as *integrated variance*.

Quadratic variation is crucial to the definition of *notional volatility*, quantified by the realized variance. Notional volatility is a natural *ex-post* expression of return variability, (Andersen, Bollerslev, Diebold, and Labys (2000)). The notional volatility equals the quadratic variation for return series and over the $[t-h,t]$ time interval is equal to

$$v^2(t,h) \equiv [r,r]_t - [r,r]_{t-h} = \int_{t-h}^{t} \sigma^2(s)ds. \tag{1.20}$$

Let I_t denote the available information of r_t until t, in the above setting, the conditional volatility (or *expected volatility*) over $[t-h,t]$ can be defined as

$$\zeta^2(t,h) = Var(r_t \mid I_t) \equiv \mathrm{E}\Big[\{r_t - E(r_t \mid I_t)\}^2 \,\Big|\, I_t\Big] \tag{1.21a}$$

$$= \mathrm{E}\Big[\Big\{\int_{t-h}^{t} \mu(s)ds - E\Big(\int_{t-h}^{t} \mu(s)ds \mid I_t\Big) + \int_{t-h}^{t} \sigma(s)dW(s)\Big\}^2 \,\Big|\, I_t\Big] \tag{1.21b}$$

$$= \mathrm{E}\Big[\Big\{\int_{t-h}^{t} \{\mu(s) - E(\mu(s) \mid I_t)ds\}^2 \,\Big|\, I_t\Big] \tag{1.21c}$$

$$+ \mathrm{E}\Big[\Big\{\int_{t-h}^{t} \sigma(s)dW(s)\Big\}^2 \,\Big|\, I_t\Big] \tag{1.21d}$$

$$+ 2\mathrm{E}\Big[\int_{t-h}^{t} \{\mu(s) - E(\mu(s) \mid I_t)\}ds \int_{t-h}^{t} \sigma(s)dW(s) \,\Big|\, I_t\Big]. \tag{1.21e}$$

By defining $A_h = O_{a.s.}(B_h)$ when A_h/B_h converges almost surely to a finite constant, as $h \to 0$. We have that equation (1.21c) $= O_{a.s.}(h^2)$, equation (1.21d) $= \int_{t-h}^{t} \sigma^2(s)ds = O_{a.s.}(h)$ and that (1.21e) $= O_{a.s.}(h^{3/2})$. Consequently, the conditional variance can be written as

$$Var(r_t \mid I_{t-h}) \simeq \mathrm{E}[v^2(t,h) \mid I_{t-h}] = \mathrm{E}[QV(t,h) \mid I_{t-h}]. \tag{1.22}$$

This implies that the conditional variance is equal to the conditional expected value of the quadratic variation, when $\mu(s) = 0$ or when $\mu(s)$ is measurable with respect to I_{t-h}. This result guarantees that the realized variance is an unbiased estimator of conditional variance.

Notional volatility and expected volatility are latent but can be estimated. The measurement of volatility can pass through parametric models or non-parametric measures. The most diffuse parametric models are the Autoregressive Conditional Heteroscedasticity (ARCH) model and the stochastic volatility models. In ARCH models, the available information, I_{t-h}, depends on the past values of returns and other directly observable variables. In stochastic volatility models the information set, I_{t-h}, incorporates both past values of returns and latent status variables. Non-parametric measures of volatility, instead, quantify notional volatility $v^2(t,h)$ directly. The major advantage of non-parametric measures is that there is no need for a functional form for the stochastic process of local martingale, $M(t)$, and for the finite variation process, $\mu(t)$.

2 Conditional Heteroscedasticity Models

2.1 Univariate GARCH models

In this paper, we analyse parametric models in discrete time, as ARCH models and stochastic volatility models. The current section introduces ARCH models.

The success of ARCH models is due to the wide application of this class of models in finance, specifically in asset allocation problems and risk management. The ability of this type of models to catch stylized facts, as not predictability of returns, presence of heavy tails in asset returns and volatility clustering, rerouted the attention of the researchers on conditional moment of second order. In particular, the class of ARCH models focuses on the variability of the second conditional moment of returns, equal to expected volatility in paragraph 1.2 and defined as

$$\zeta^2(t,h) = \mathrm{E}\Big[\Big(r(t,h) - E[\mu(t,h)] \mid I_{t-h}\Big)^2 \Big| I_{t-h}\Big]. \tag{2.1}$$

In order to explain the variability of the second conditional moment, Engle (1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model, which specifies the error in the linear regression of asset returns y_t, such that

$$y_t = x_t' b + \varepsilon_t.$$

Let the Gauss-Markov assumptions be valid, in particular

$$\mathrm{E}[\varepsilon_t \mid I_{t-1}] = 0,$$

then the innovation, ε_t, at time t may be defined as follows

$$\varepsilon_t = u_t h_t^{1/2} \tag{2.2}$$

where $u_t \sim i.i.d.(0,1)$ is a standard process with zero mean and unitary variance and h_t is the conditional variance of the innovation. It is further assumed that

$$Cov(\varepsilon_t \varepsilon_{t+k}) = 0.$$

While the non-conditional variance of this process is constant, it is possible that the conditional variance, h_t, varies during time:

$$h_t = \mathrm{E}[\varepsilon_t^2 \mid I_{t-1}] = Var(\varepsilon_t \mid I_{t-1}). \tag{2.3}$$

Consequently, the conditional distribution of the error is

$$\varepsilon_t \mid I_{t-1} \sim N(0, h_t).$$

Engle (1982) specifies the conditional variance as a linear function of the past squared returns of ε_t, a model for h_t can be written as

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \tag{2.4}$$

where w is the constant and α_i is the coefficient related to ε^2_{t-i}.

Positiveness of the variance is guaranteed if $\alpha_i \geq 0$, for each lag $i = 1,2,.....,q$, and if $w \geq 0$.

ARCH(q) process can be written according to MA(q) representation for the squares of the innovations, such that

$$h_t = w + A(L)\varepsilon^2_t \tag{2.5}$$

Equation (2.5) is weakly stationary if the roots of the polynomial $1-A(L)$ are external to the unitary circle, where $A(L) = \alpha_1 L + \alpha_2 L^2 + ... + \alpha_q L^q$ is the polynomial of the lag operator. The necessary and sufficient condition for process stationarity is

$$\sum_{i=1}^{q} \alpha_i < 1.$$

The unconditional variance can be expressed as

$$Var(\varepsilon_t) = E(\varepsilon^2_t) = \frac{w}{1 - \sum\limits_{i=1}^{q} \alpha_i} = \frac{w}{1 - A(1)}. \tag{2.6}$$

2.1.1 GARCH model

The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, proposed by Bollerslev (1986), generalizes the ARCH model introducing an autoregressive component in the conditional variance regression. The specification of the conditional variance, under the same assumptions of the ARCH model and given the information set I_{t-1}, can be expressed as

$$h_t = w + \sum_{i=1}^{q} \alpha_i \varepsilon^2_{t-i} + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.7}$$

with $w \geq 0$, $\alpha_i \geq 0$ for $i = 1,2,....,q$ lags and $\beta_j \geq 0$ for $j = 1,2,....,p$ lags. The conditional variance is function of p lags of the conditional variance itself and q lags of ε^2_t, catching the short-term effects related to the evolution of the considered variable and the long-term effects related to the persistence of the volatility. Using the lag operator, equation (2.7) can be written as $h_t = A(L)\varepsilon^2_t + B(L)h_t$ where $A(\cdot)$ and $B(\cdot)$ are the polynomials of the lags of ε^2_t and h_t. GARCH(p,q) process is covariance stationary if the roots of the polynomial $1 - A(L) - B(L)$ fall outside the unitary circle,

$$\sum_{i}^{q} \alpha_i + \sum_{j}^{p} \beta_j < 1.$$

Unconditional variance may be specified as

$$E(\varepsilon^2_t) = \frac{w}{1 - \sum\limits_{i}^{q} \alpha_i - \sum\limits_{j}^{p} \beta_j} = \frac{w}{1 - A(1) - B(1)}.$$

Several extensions have been proposed for the basic GARCH model, see Bollerslev (2009) for a large literature review. The most relevant are the following:

- GARCH-M (GARCH in mean) model proposed by Engle, Lilien, and Robins (1987), where a function of the conditional variance is introduced in the regression function of the asset return, y_t, such that

$$\begin{cases} y_t = X_t b + \delta g(h_t) + \varepsilon_t \\ \varepsilon \mid I_{t-1} \sim N(0, h_t) \end{cases} \tag{2.8}$$

  where $g(h_t)$ is a continuous and differentiable conditional variance function;

- EGARCH (Exponential GARCH) from Nelson (1990), that aims to catch asymmetric effects of a single shock on volatility

$$\log(h_t) = w + \sum_{j}^{p} \beta_j \log(h_{t-j}) + \sum_{i}^{q} \alpha_i g(u_t), \tag{2.9}$$

  where $g(u_t) = \left\{ \phi u_{t-i} + \gamma \left[\mid u_{t-i} \mid -E \mid u_{t-i} \mid \right] \right\}$ is $i.i.d.$ with zero mean. If $\phi \neq 0$, the model takes into account the asymmetric behaviour of the volatility to the shocks;

- NGARCH (Nonlinear GARCH), introduced by Higgins and Bera (1992), also known as Power GARCH. The conditional deviation, δ, is modelled as a function of lagged conditional deviations and lagged absolute innovations at the power δ, then

$$(\sqrt{h_t})^{\delta} = w + \sum_{i=1}^{q} \alpha_t \mid \varepsilon_{t-i} \mid^{\delta} + \sum_{i=1}^{p} \beta_i (\sqrt{h_{t-i}})^{\delta} \tag{2.10}$$

  when $\delta = 2$, the model returns to a classic GARCH.

- TS-GARCH (Taylor-Schwert GARCH), or Absolute Value GARCH, introduced by Taylor (1986) and Schwert (1989), is a particular case of the Power GARCH where the influence of high values in the traditional GARCH(p,q) model is limited:

$$(\sqrt{h_t}) = w + \sum_{i=1}^{q} \alpha_i \mid \varepsilon_{t-i} \mid + \sum_{i=1}^{p} \beta_i (\sqrt{h_{t-i}}); \tag{2.11}$$

- GJR-GARCH from Glosten, Jagannathan, and Runkle (1993), this model allows different responses from the conditional variance to past innovations' sign. The conditional variance is modelled as

$$h_t = w + \sum_{j=1}^{q} \{\alpha_j + \delta_j I(\varepsilon_{t-j} > 0)\} \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.12}$$

  where $I(\cdot)$ is an index function;

- IGARCH (Integrated GARCH), the model, starting from the equation with the lag operator, $h_t = A(L)\varepsilon_t^2 + B(L)h_t$, is obtained when the autoregressive polynomial admits a unitary root;

- FIGARCH (Fractionally Integrated GARCH) introduced by Baillie, Bollerslev, and Mikkelsen (1996), it considers the class of processes between unitary root and stationary processes. The lag operator is defined as

$$1 - A(L) - B(L) = f(L)(1-L)^d.$$

FIGARCH is the more general version of an IGARCH, obtained when $d = 1$, and a GARCH, obtained with $d = 0$.

- SWARCH (regime SWitching ARCH), proposed by Cai (1994) and Hamilton and Susmel (1994) in two different papers. The conditional variance is modelled as

$$\frac{h_t}{\gamma_{s_t}} = w + \sum_{i=1}^{q} \alpha_i \frac{\varepsilon_{t-i}}{\gamma_{s_{t-i}}} \tag{2.13}$$

where γ_{s_t} and $\gamma_{s_{t-i}}$ are the scale parameters.

- Smooth Transition GARCH (STGARCH), proposed by Hagerud (1997) and Gonzalo-Rivera (1998), is a non linear version of the GJR-GARCH model, where the change of regime is driven by a transition function. The Smooth Transition GARCH is defined as

$$h_t = w_1 + \sum_{j=1}^{q} \alpha_{ij}\varepsilon_{t-j}^2 + (w_2 + \sum_{j=1}^{q} \alpha_{2j}\varepsilon_{t-j}^2)G(\gamma, c; \varepsilon_{t-j}) + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.14}$$

where the transition function is denoted by

$$G(\gamma, c; \varepsilon_{t-j}) = (1 + e^{\gamma \prod_{k=1}^{K} (\varepsilon_{t-j} - c_k)})^{-1}, \quad \gamma > 0$$

when $K = 1$ the transition function is a logistic function.

- Threshold GARCH (TGARCH), proposed by Zakoian (1994); the conditional standard deviation is modelled as

$$h_t^{1/2} = w + \sum_{j=1}^{q} (\alpha_j^+ \varepsilon_{t-j}^+ - \alpha_j^- \varepsilon_{t-j}^-) + \sum_{j=1}^{q} \beta_j \tag{2.15}$$

where $\varepsilon_{t-j}^+ = \max(\varepsilon_{t-j}, 0)$ and $\varepsilon_{t-j}^- = \min(\varepsilon_{t-j}, 0)$. If $c = 0$, TGARCH model is linear in parameters.

2.1.2 GARCH models estimation

GARCH models are usually estimated through maximum likelihood, under the following assumptions:

- $u_t(\theta) = \varepsilon_t(\theta)/h_t(\theta)^{1/2}$;
- $u_t(\theta) \sim i.i.d.(0,1)$;
- $u_t(\theta) \sim f(u_t(\theta); \eta)$;

where θ is the vector of unknown parameters, η is the difference between ε_t^2 and its conditional mean h_t. Since the t-th return y_t cannot be considered independent from the other realizations, the likelihood function cannot be equal to the product of marginal distribution. Likelihood function is built as the multiplication of the conditional distributions. Let $\{y_T, y_{T-1} ... y_1\}$ be a sample realization of the GARCH model and $\psi' = (\theta', \eta')$ the $(m + k) \times 1$ vector of parameters to be estimated, the joint probability distribution can be defined as sequential factorization

$$L_T(y, \eta) = f(y_T, y_{T-1}, ...y) = \prod_{t=1}^{T} f(\varepsilon_t(\theta), \eta \mid I_{t-1}) \cdot f(\varepsilon_1) \tag{2.16}$$

where $f(\varepsilon_t(\theta),\eta \mid I_{t-1})$ is the conditional distribution of the innovation. If it is assumed ε_1 as degenerated, $f(\varepsilon_1)$ does not depend on unknown parameters, in this way the join distribution becomes

$$L_T(y,\eta) = f(\varepsilon_t(\theta),\eta \mid I_{t-1}). \tag{2.17}$$

From the distribution of the standard innovation, u_t, it follows that

$$L_T(y,\eta) = \prod_{t=1}^{T} f(u_t(\theta),\eta) h_t(\theta)^{-1/2} \tag{2.18}$$

where $h_t(\theta)^{-1/2} = \mid \partial u_t/\partial \varepsilon_t \mid$ is the determinant of the Jacobian obtained with the transformation of the innovation ε_t to the standard u_t. Thus, the log-likelihood can be written as

$$\ell_T = \log\left[L_T(y_T \ldots y_1\right] = \sum_{t=1}^{T} \log\Big\{\left[f(u_t(\theta),\eta)\right] - \frac{1}{2}\log\left[h_t(\theta)\right]\Big\}. \tag{2.19}$$

The ML estimator, $\hat{\psi}$, is the solution of the equation

$$S_t(y_T \ldots y_1) = \sum_{t=1}^{T} s_t(y_t) = 0 \tag{2.20}$$

where

$$s_t = \frac{\partial \ell_T(y_t,\eta)}{\partial \psi} \tag{2.21}$$

is the score of the t-th observation that can be obtained using a numerical optimisation.

This estimation method is valid under the assumption of Gaussian distributed innovations, $u_t \sim N(0,1)$, such that

$$\ell_T(y,\theta) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log\left[h_t(\theta)\right] - \frac{1}{2}\sum_{t=1}^{T} u_t(\theta)^2.$$

Where the score is

$$s_t = \frac{\partial \mu_t(\theta)}{\partial \theta}\cdot\frac{u_t(\theta)}{h_t} + \frac{1}{2}(h_t(\theta))^{-1}\frac{\partial h_t(\theta)}{\partial \theta}\Big[\frac{u_t^2(\theta)}{h_t(\theta)} - 1\Big].$$

Several alternative distributions have been used in order to take into account of the leptokurtosis in the asset returns distribution, as the Student's t-distribution, the non-central t-distribution and the Generalized Error Distribution (GED).

Since normal distribution of innovations has been empirically denied, in order to avoid the specification of their distribution, the quasi-maximum likelihood (QML) estimator can be used. Under few regularity conditions, the QML estimator is asymptotically normal with

$$\sqrt{T}(\hat{\theta}_n - \theta_0^*) \xrightarrow{d} N(0, A^{-1}BA^{-1})$$

where the matrices A and B are equal to

$$A = -\frac{1}{T}E_0\left[\frac{\partial^2 \ell_T(\theta)}{\partial\theta\partial\theta'}\right]$$
$$B = \frac{1}{2}E_0\left[\frac{\partial \ell_T(\theta)}{\partial\theta}\frac{\partial \ell_T(\theta)}{\partial\theta'}\right].$$

When $u_t \sim N(0,1)$, the matrices A and B coincide.

2.2 Multivariate GARCH models

Normally, problems in economics need the specification and the estimation of a volatility measure in the multivariate framework. Hence, the GARCH literature has been extended in the multivariate case, as further analysed in this section.

Let y_t be a vector of n components and ε_t a vector of n innovations with zero mean, given the information set I_{t-1}, it is assumed that

$$\varepsilon_t = H_t^{1/2} u_t \tag{2.22}$$

where H_t is the $n \times n$ conditional covariance matrix and u_t is a vector such that $E(u_t u_t') = I_n$.

2.2.1 VECH

The VECH model, introduced by Bollerslev, Engle, and Wooldridge (1988), generalises in the multivariate the GARCH model

$$vech(H_t) = c + \sum_{i=1}^{q} A_i vech(\varepsilon_{t-i}\varepsilon_{t-i}') + \sum_{j=1}^{p} B_j vech(H_{t-j}) \tag{2.23}$$

where $vech(\cdot)$ is the mathematical operator that transforms a symmetric matrix in a vector considering only the lower triangular part of the matrix. Since c is a $n(n+1)/2 \times 1$ vector and A_i and B_j are $n(n+1)/2 \times n(n+1)/2$ matrices, the total number of parameters to be estimated is $(p+q)[n(n+1)/2]^2 + n(n+1)/2$. When the number of the asset is particularly high, there can be a numerical problem with the estimation. Moreover, it is not possible to ensure a semi-definite positive covariance matrix without restrictions on the parameters.

Bollerslev, Engle, and Wooldridge (1988) introduced a restricted version of the VECH model, assuming that A_i and B_j are diagonal matrices. As showed in Bollerslev, Engle, and Nelson (1994), H_t can be ensured positive definite for each t. The diagonal GARCH(p,q) model reduces the number of parameters to be estimated to $(p+q+1)n(n+1)/2$. The major limitation of this model is that it does not allow interactions between different conditional covariances.

2.2.2 BEKK

The Baba-Engle-Kraft-Kroner (BEKK) model, formalised by Engle and Kroner (1995), ensures positive definite conditional covariance matrix thanks to the quadratic form of the equation. The model is written as

$$H_t = CC' + \sum_{i=1}^{q}\sum_{k=1}^{K} A_{ki}' \varepsilon_{t-i}\varepsilon_{t-i}' A_{ki} + \sum_{j=1}^{p}\sum_{k=1}^{K} B_{kj}' H_{t-j} B_{kj} \tag{2.24}$$

where A_{ki}, B_{kj} are $n \times n$ non-negative symmetric matrices and C is a $n \times n$ low triangular matrix, K are the called the degree of generality of the model. BEKK model is covariance stationary if and only if the eigenvalues of the matrix

$$\sum_{i=1}^{q}\sum_{k=1}^{K} (A_{ki}' \otimes A_{ki}') + \sum_{j=1}^{p}\sum_{k=1}^{K} (B_{kj}' \otimes B_{kj}')$$

are not greater than 1 in modulus, where $\otimes$ is the Kronecker product. When $K > 1$, there is a identification problem since several parametrizations are possible for the same matrix (cfr. Silvennoinen and Teräsvirta (2008)).

2.2.3 Factorial and orthogonal GARCH models

This class of models aims to minimize the number of the parameters to be estimated. Based on Capital Asset Pricing Model (CAPM) from Sharpe (1964), these models imply that few common determinants drive asset volatility. Engle, Ng, and Rothschild (1990) introduced the first factorial GARCH model, based on the Arbitrage Price Theory from Ross (1976), supposing that H_t matrix is generated by a k number of factors, f_t, such that

$$H_t = \Omega + \sum_{k=1}^{K} w_k w_k' f_{k,t}, \tag{2.25}$$

where Ω is a $n \times n$ semi-definite positive matrix, $f_{k,t}$ are the factors and w_k, for $k = 1, ...K$, are the linearly independent vectors of weighs. The model assumes that the factors are correlated, making hard to understand the effects of the single factors and individuate the factors to be used.

The use of principal component analysis tries to overcome these limitations by introducing the invertible decomposition W

$$y_t = W z_t. \tag{2.26}$$

The original observations are a linear combination of the unobservable factors z_t. The models from this specification are named Orthogonal (O-) GARCH, Alexander and Chibumba (1996), and Generalized Orthogonal (GO-) GARCH, Van der Weide (2002) and are based on the hypothesis that the orthogonal matrix W is constant and invertible, and that the factors are conditionally heteroscedastic and that follow a GARCH process.

2.2.4 CCC and DCC

The last class of multivariate specification of conditional heteroscedasticity models is based on the decomposition of the conditional covariance matrices. Bollerslev (1990) proposes a class of constant conditional correlation (CCC) models, where the conditional covariance matrix is decomposed as

$$H_t = D_t^{1/2} R D_t^{1/2} \tag{2.27}$$

where D_t is a $n \times n$ diagonal matrix with the conditional variance, h_{iit}, on the diagonal and R is the conditional correlations matrix where the single element is $\rho_{ij} = h_{ij}(h_{ii}h_{jj})^{-1/2}$.

When the correlations are constant, the elements of H_t, $h_{ijt} = \rho_{ij}(h_{iit}h_{jjt})^{1/2}$, are time varying only for the effect of the variations of single conditional variances. H_t is positive definite for each t since R is positive definite and constant and the variances on the diagonal of D_t are positive for construction.

It is further assumed that the conditional variances on the diagonal of D_t follow a univariate GARCH process, such that

$$\begin{cases} h_{iit} = w_i + \sum_{r=1}^{q} \alpha_{ir} \varepsilon_{it-r}^2 + \sum_{r=1}^{p} \beta_{ir} h_{iit-r} \\ h_{ijt} = \rho_{ij} (h_{iit} h_{jjt})^{1/2} \end{cases} \tag{2.28}$$

This approach guarantees a positive definite matrix H_t and reduces the number of parameters to be estimated to $n(1+p+q)+n(n+1)/2$, but is based on the unrealistic restriction of constant conditional correlations. Engle and Sheppard (2001) and Engle (2002) proposed the Dynamic Conditional Correlation (DCC) to introduce a dynamic component in the correlations. The model is written as

$$H_t = D_t^{1/2} R_t D_t^{1/2}, \tag{2.29}$$

where R_t is time varying for the effects of $\rho_{ijt} = h_{ijt}(h_{iit}h_{jjt})^{-1/2}$. In the first step, the estimates of the conditional variances follow a GARCH(p,q) process

$$h_{iit} = \alpha_{0i} + \sum_{s=1}^{Q_i} \alpha_s y_{it-s}^2 + \sum_{s=1}^{P_i} \beta_s h_{iit-s}, \tag{2.30}$$

where Q_i and P_i are the numbers of the GARCH lags. In the second step, dynamics correlations are estimated as follows

$$Q_t = (1 - \sum_{r=1}^{q} a_r - \sum_{r=1}^{p} b_r)\bar{Q} + \sum_{r=1}^{q} a_r u_{t-r} u'_{t-r} + \sum_{r=1}^{p} b_r Q_{t-r} \tag{2.31}$$

$$R_t = \widetilde{Q}_t^{-1} Q_t \widetilde{Q}_t^{-1} \tag{2.32}$$

where $\widetilde{Q}_t$ is a matrix with the square roots of the element of Q_t on its diagonal, $\bar{Q} = E(u_t u'_t)$ is the unconditional correlations matrix of y_t and $u_{it} = y_{it}/h_{it}^{1/2}$ are the standardized residuals.

The major advantage of this model is the reduction of the number of parameters to $p+q+n+\sum_{i=1}^{n}(P_i+Q_i)$ and the inclusion of a time varying correlation matrix.

2.2.5 Multivariate GARCH estimation

The estimation of the conditional covariance matrices is based on the maximum likelihood function. Let $\{y_t : t = 1,2,...\}$ be a sequence of $n \times 1$ vectors of random variables y_t, it is supposed that the first two conditional moments are:

$$E(y_t \mid I_{t-1}) = \mu_t(\theta)$$
$$Var(y_t \mid I_{t-1}) = H_t(\theta)$$

where $\theta \in \Theta$ and Θ are a subset of R^p. The log-likelihood is given by:

$$\ell_T = \sum_{t=1}^{T} \ell_t(\theta)$$

where T is the number of the observations. Thus

$$\begin{aligned}\log \ell_t(\theta) &= -\frac{1}{2}\log\det H_t(\theta) - \frac{1}{2}\Big[y_t - \mu_t(\theta)\Big]' H_t(\theta)\Big[y_t - \mu_t(\theta)\Big] \\ &= -\frac{1}{2}\log\det H_t(\theta) - \frac{1}{2}\varepsilon_t(\theta)' H_t^{-1}\varepsilon_t(\theta).\end{aligned}$$

First order conditions for maximization are

$$\frac{\partial \ell_T(\hat{\theta}_T)}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial \log \ell_t(\hat{\theta}_t)}{\partial \theta} = 0.$$

If the conditional moments $\mu_t(\theta)$ and $H_t(\theta)$ are differentiable respect to θ and if $H_t(\theta)$ is a non-singular matrix with probability one, for each $\theta \in \Theta$, then the gradient of the t-th observation is

$$s_t(\theta)' = \frac{\partial \mu_t(\theta)}{\partial \theta} H_t^{-1}(\theta)\varepsilon_t(\theta) + \frac{1}{2}\frac{\partial H_t(\theta)'}{\partial \theta}\Big[H_t^{-1}(\theta) \otimes H_t^{-1}(\theta)\Big] vec\Big[\varepsilon_t(\theta)\varepsilon_t(\theta)' - H_t(\theta)\Big].$$

Under few regularity conditions on conditional variance and on the stationarity of the gradient, it can be proved that the quasi-maximum likelihood estimator exists asymptotically and it is asymptotically normal (Bollerslev and Wooldridge (1992)).

3 Stochastic volatility models

Stochastic volatility models represent a parametric alternative to GARCH models for the estimation of volatility. For this class of models, the informative set I_{t-h} is not directly observable respect to time. As a result of this, the latent volatility is driven by an underlying stochastic process. In this section, a brief review of the principal stochastic volatility models is presented, for a more detailed analysis see Taylor (1994), Shephard (1996) and Ghysels, Harvey, and Renault (1996).

It is supposed that volatility follows a stochastic process, v_t, and that the asset price follows a process like

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t \tag{3.1}$$

where W_t is a Brownian motion with zero mean and unitary variance, μ is the drift component of price S_t. The equation underlying the stochastic volatility is defined as

$$dv_t = \alpha_{s,t} dt + \beta_{s,t} dB_t \tag{3.2}$$

where $\alpha_{s,t}$ and $\beta_{s,t}$ are functions of v_t. The most significant difference with GARCH models is that, conditionally to information set I_{t-1}, volatility v_t is unknown and unobservable (cfr. Bauwens, Hafner, and Laurent (2012)).

Heston (1993) was the first to propose a model of volatility dependent on the price dynamics, such that

$$dv_t = k(\theta - v_t)dt + \sigma\sqrt{v_t} dB_t \tag{3.3}$$

where θ is the mean long-term volatility, k is the rate at which volatility reverts toward its long-term mean, σ is the volatility of the volatility process and dB_t is a zero mean Gaussian process, dW_t and dB_t are correlated with the constant correlation value ρ.

Several extensions of Heston (1993) model have been proposed:

- CEV (constant elasticity of variance) model, in this model the relation between volatility and price is given by

$$dS_t = \mu S_t dt + \sigma S_t^{\beta/2} dW_t \tag{3.4}$$

  where dW is a Wiener process, σ is a positive constant and β is known as the CEV parameter. β influences the direction and the size of the impact of the price on the volatility. Since this model does not present a separated process for volatility, it is called local volatility model;

- Chen (1996) model, the dynamics that drive the interest rates, v_t, are derived from the following system

$$dr_t = (\theta_t - \alpha_t)dt + \sqrt{r_t}\sigma_t dW_t \tag{3.5}$$

$$d\alpha_t = (\zeta_t - \alpha_t)dt + \sqrt{\alpha_t}\sigma_t dW_t \tag{3.6}$$

$$d\sigma_t = (\beta_t - \sigma_t)dt + \sqrt{\sigma_t}\eta_t dW_t; \tag{3.7}$$

- SABR (Stochastic Alpha, Beta, Rho) model, aims to reproduce the dynamics of volatility on the derivatives market. The equations that define the model are:

$$dS_t = \sigma_t S_t^{\beta} dW_t \tag{3.8}$$

$$d\sigma_t = \alpha\sigma_t dz_t, \tag{3.9}$$

  where W_t and z_t are two correlated Wiener processes.

3.0.1 Multivariate stochastic volatility models (MSV)

Stochastic volatility models are also extended in the multivariate framework. Considering a vector of logarithmic prices $S = (S_1, ...S_n)$ of n assets with $y = (y_1, ..., y_n)$ returns vectors, the model for S can be defined as

$$dS_t = H_t^{1/2} dW_t \tag{3.10}$$

$$df[vech(H_t)] = a[vech(H_t)]dt + b[vech(H_t)]dB_t \tag{3.11}$$

where W_t and B_t are two vectors of Brownian motions, H_t is the instantaneous covariance matrix and f, a and b are known functions.

It follows that the generic MSV model in discrete time is given by

$$y_t = H_t^{1/2}\varepsilon_t \tag{3.12}$$

$$\varepsilon_t \sim N(0, I_n) \tag{3.13}$$

$$u_{t-1} \sim N(0, \Sigma_u) \tag{3.14}$$

$$f[vech(H_t)] = a[vech(H_{t-1})] + f[vech(H_{t-1})] + b[vech(H_{t-1})]u_{t-1}, \tag{3.15}$$

where $y_t = S_t - S_{t-1}$.

This model does not guarantee a positive definite H_t; several works aims to overcome this limitation.

Harvey, Ruiz, and Shephard (1994) introduces a new model, where H_t is ensured positive definite and is defined as

$$\varepsilon_t = H_t^{1/2} z_t \qquad z_t \sim N(0, \Sigma_t) \tag{3.16}$$

$$H_t = diag(exp(h_{1t}), ..., exp(h_{Nt})) \tag{3.17}$$

$$h_{t+1} = w + \beta \odot h_t + u_t \qquad u_t \sim N(0, \Sigma_u) \tag{3.18}$$

where $h_t = (h_{1t}, ..., h_{Nt})$ is the vector of the volatility at time t, Σ_t is the correlation matrix, $\odot$ is the Hadamard product operator, w and β are vectors of parameters. The model is too restrictive, because it has constant correlations in a similar way to the CCC model from Bollerslev (1990).

Harvey's model has been extended in several ways, to taking into account of time varying correlations, leverage effects, heavy tails distribution of innovations, for further details see Asai, McAleer, and Yu (2006) and Andersen (2009).

4 Realized variance

Conditional heteroscedasticity and stochastic volatility model represent the most common approaches to measure volatility. However, these models heavily depend on the specification of the underlying process of volatility and necessitate of strong restrictions on parameters to be estimated.

Lately, the attention of the research on volatility measure has moved to high frequency data. Firstly, Merton (1980) showed that volatility can be defined as the sum of the squared returns at high frequency level. Recently, Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001) pointed out that summing up the squared intra-daily returns it is obtained an observable measure of daily volatility named realized variance (RV). *Ex post* volatility, excluding measurement errors and jumps, becomes "observable" and can be directly modelled.

4.1 Realized variance construction

The realized variance theory is based on the idea that the realized measure is the best approximation of the unobservable volatility when the returns are sampled at sufficiently high frequencies.

Supposing that the log-price of an asset, $p(t)$, follows a diffusion process such that

$$dp(t) = \mu(t)dt + \sigma(t)dW(t) \qquad \text{con t=1, 2, 3 ...} \tag{4.1}$$

that describes the trajectories of a *semimartingale* in continuous time over the interval $[0, T]$, with $0 \leq s \leq t \leq T$, where μ_t is the drift component, σ_t is the instantaneous volatility of the process or standard deviation, strictly positive and square integrable (i.e. $E(\int_0^t \sigma_s^2 ds) < \infty$), and W_t is a standard Brownian motion.

Let the continuously compounded return between $t-h$ and t, with $0 < h \leq t$, be

$$r_t = p(t) - p(t-h) = \int_{t-h}^{t} \mu_s ds + \int_{t-h}^{t} \sigma_s dW_s. \tag{4.2}$$

The quadratic variation of the return, natural measure of the variability of the diffusion path of a martingale according to stochastic integration theory, is defined as

$$[p]_t = QV_t = \int_{t-h}^{t} \sigma^2(s) ds. \tag{4.3}$$

Equation (4.3) shows that the drift innovations do not affect the variation of the diffusion path of returns. In this context, since the quadratic variation is totally induced by the innovations of a local martingale, it coincides with the integrated variance[3], index of *ex-post* cumulated variability of returns and given by

$$IV_t = \int_{t-h}^{t} \sigma^2(s) ds = QV_t. \tag{4.4}$$

From the property of the quadratic variation (from the equation (1.13) in section 1.2), it follows that, in absence of microstructure errors[4] and of measurement errors (Andersen, Bollerslev, Diebold, and Labys (2000) and Barndorff-Nielsen and Shephard (2002a)), returns quadratic variation can be approximated as

$$[p(t)] = \operatorname*{plim}_{n \to \infty} \sum_{j=1}^{n} \left[p(s_j) - p(s_{j-1}) \right]^2, \tag{4.5}$$

for each partition sequence $0 = s_0 < s_1 < ... < s_n = t$ with $| s_j - s_{j-1} | \to 0$, when the number of the partitions $n \to \infty$.

Since the intra-daily return is defined as

$$r_{t,i} = p_{t,i} - p_{t,i-1} \qquad \forall i = 1, ..., n$$

and the daily return as

$$r_t = \sum_{i=1}^{n} r_{t,i},$$

[3] Quadratic variation and integrated variance do not coincide in a more general process (e.g. diffusion model with jumps).

[4] A microstructure error emerges for the presence of non-synchronized exchanges, for the absence of trades, for some properties of the trading mechanism (Black (1976) and Amihud and Mendelson (1987)) and for the presence of discrete prices (Harris (1990) and Harris (1991)).

semimartingale theory ensures that the realized variance, defined as the sum of the intra-daily squared returns, converges in probability to the quadratic variation and, consequently, to the integrated variance of the day t, for $n \to \infty$. Let the realized variance be

$$RV_t = \sum_{i=1}^{n} r_{t,i}^2, \tag{4.6}$$

it follows that

$$RV_t \xrightarrow{p} [r,r]_t - [r,r]_{t-h} \equiv QV_t \tag{4.7}$$

$$RV_t \xrightarrow{p} IV_t. \tag{4.8}$$

Realized volatility is a consistent estimator of notional volatility, such that

$$RV_t \xrightarrow{p} v^2(t,h). \tag{4.9}$$

This implies that the expected realized volatility is a consistent estimator of expected notional volatility

$$\mathrm{E}\big[RV_t \,\big|\, I_{t-h}\big] \xrightarrow{p} \mathrm{E}\big[v^2(t,h) \,\big|\, I_{t-h}\big]. \tag{4.10}$$

From equation (1.22), it follows that, if the return process is square integrable and if $\mu(t) \equiv 0$, the realized volatility is an unbiased estimator of the conditional variance of returns:

$$\mathrm{E}\big[RV_t \,\big|\, I_{t-h}\big] = \mathrm{E}\big[QV_t \,\big|\, I_{t-h}\big] = Var\big[r(t,h) \,|\, I_{t-h}\big]. \tag{4.11}$$

This equivalence merges the realized volatility and the conditional variance from ARCH models. In particular, it is possible to build a model of the time series for the realized variance which approximates the conditional variance of returns. If $\mu(t) \not\equiv 0$, the convergence of RV to QV does not automatically imply the convergence in mean of the same objects.

First examples of realized variance can be found in Merton (1980), Poterba and Summers (1986), Schwert (1989), Richardson and Stock (1989), Schwert (1990), Taylor and Xu (1997) e Christensen and Prabhala (1998). The use of a realized measure, however, spread only after the formalization of the measure by Andersen and Bollerslev (1998) that proved, with Andersen, Bollerslev, Diebold, and Labys (2003), that the realized variance is a consistent estimator of the daily volatility only if the distance between intra-daily observations approaches zero or if the sampling frequency approaches infinity.

4.2 Realized variance distribution

The asymptotic distribution of the realized variance has been analysed in two different papers, Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002a).
From their results, it emerges that the realized variance is distributed as

$$n^{1/2} \cdot \frac{1}{\sqrt{2IQ_t}} (RV_t - IV_t) \xrightarrow{d} N(0,1) \tag{4.12}$$

where n is the intra-daily sampling frequency and IQ_t is defined as the *integrated quarticity*:

$$IQ_t = \int_{t-h}^{t} \sigma^4(s)ds.$$

In order to make inference, a consistent estimator of IQ_t is necessary. Barndorff-Nielsen and Shephard (2002a) showed that a consistent estimator of IQ_t is given by the *Realized Quarticity*, RQ, defined as

$$RQ_t = \frac{1}{3}\sum_{i=1}^{n} r_i^4, \tag{4.13}$$

such that

$$\frac{\delta^{-1/2}(RV_t - IV_t)}{\sqrt{\frac{2}{3}RQ_t}} \xrightarrow{d} N(0,1). \tag{4.14}$$

Barndorff-Nielsen and Shephard (2002a) proved that the best approximation of the realized distribution requires a log-linearisation, such that

$$\frac{\delta^{-1/2}(\log RV_t - \log IV_t)}{\sqrt{\frac{2}{3}\frac{RQ_t}{RV_t^2}}} \xrightarrow{d} N(0,1). \tag{4.15}$$

4.3 Caveats with realized variance

The previous results show that the highest sampling frequency should be usually preferred to measure realized variance. However, the logarithm of prices does not spread in continuous time in practice, but it is usually observed at discrete and not regular intervals. The sampling methods influence the discretization of the price. We analyse different sampling methods:

- *Calendar time sampling*, transactions are selected by regularly spaced calendar time, such as every 5 minutes or every hour. De Pooter, Martens, and Van Dijk (2008) attempt to find the optimal sampling frequency, showing that an optimal sampling frequency for the measurement of the daily realized variance is comprised between 30 and 65 minutes.
- *Transaction time sampling*, data are sampled each n transactions.
- *Business time sampling*, the price process is sampled at equidistantly spaced points in business time, such that $IV_{t,i} = \frac{IV_t}{n}$.
- *Tick time sampling*, prices are sampled tick-by-tick.

Several works dealt with the choice of the best sampling method. Among others, Oomen (2005a, 2005b) concluded that the *transaction time sampling* method is the best choice, despite the large diffusion of the *calendar time sampling* method.
In addition to the choice of the sampling methodology, the process for determining the realized variance must face the presence of the microstructure error. There exists, in fact, a *trade-off* between the microstructure error and the accuracy of the estimation that might lead the realized variance to be a non-robust estimator of the daily integrated variance.

The effect of the microstructure error is enhanced when *calendar time sampling* methodology is used. Andersen *et al.* (2000, 2001, 2003) suggested that the issue related to microstructure error could be solved through subsampling. A practical solution is to implement *sparse sampling*, which implies to sample not too frequently, e.g. every 5 or 30 minutes.

Following this procedure, Bandi and Russell (2005), Bandi and Russell (2007) and Zhang, Mykland, and Aït−Sahalia (2005) proposed an approach to determine the optimal subsampling frequency that relies on the minimization of the mean square error (MSE):

$$MSE\left(n_t^{(sparse)}\right) = 2n_t^{(sparse)}E(\varepsilon_{t,i}^2) + 4n_t^{(sparse)}E(\varepsilon_{t,i}^4)$$
$$+\left[8RV_t^{(sparse)}E(\varepsilon_{t,i}^2) - 2V(\varepsilon_{t,i}^2)\right] + \frac{2}{n_t^{(sparse)}}IQ_t^{(sparse)}$$

where n_t is the number of subsamples, RV_t is the realized variance, IQ_t is the integrated quarticity, ε_t is the microstructure error with zero mean. Thus, the optimal sampling frequency is approximated as

$$n_t^* = \left\{\frac{IQ_t}{4\left[E(\varepsilon_{t,i}^2)\right]^2}\right\}^{1/3}.$$

Where $E(\varepsilon_{t,i}^2)$ can be consistently estimated by $\frac{1}{2n_t}RV_t$ and the integrated quarticity can be estimated through realized quarticity.

Nevertheless, Zhang (2006) showed that on one side the use of subsampling reduces the magnitude of the microstructure error, specifically by $2nE(\varepsilon_{t,i}^2)$, on the other side it increases the variance for the larger subsampling interval.

Alternative integrated variance estimators have been proposed in order to address this *trade-off* and to consider the microstructure error. Zhang, Mykland, and Aït−Sahalia (2005) proposed the *two-time scales Estimator (TTSE)*, given by

$$RV_t^{(TTSE)} = \frac{1}{K}\sum_{k=1}^{K} RV_t^{(k)} - \frac{\overline{n}_t}{n_t}RV_t.$$

Where the set of daily observations is divided in K non-overlapping subset, where k=1,...,K, n_t is the number of observations of the entire grid, $RV_t^{(k)}$ is the realized variance of the subset k, RV_t is the realized variance of the day t and $\overline{n}_t$ is defined as

$$\overline{n}_t = \frac{1}{K}\sum_{k=1}^{K} n_t^{(k)} = \frac{n_t - K - 1}{K}.$$

Aït−Sahalia, Mykland, and Zhang (2006) re-interprets this estimator according to the following specification:

$$RV_t^{(TTSE,adj)} = \left(1 - \frac{\overline{n}_t}{n_t}\right)^{-1} RV_t^{(TTSE)}.$$

The most common estimator in presence of microstructure error is the *kernel* estimator. Zhou (1996) firstly proposed this estimator for high-frequency data, proving that the microstructure error is

time-dependent and has time-varying properties. From this proposed estimator, Hansen and Lunde (2004) and Hansen and Lunde (2006) developed the following estimator

$$RV_t = RV_t + 2\sum_{h=1}^{H}\frac{n_t}{n_t - h}\hat{\gamma}_h, \tag{4.16}$$

with

$$\hat{\gamma}_h = \frac{n_t}{n_t - h}\sum_{j=1}^{n_t - h} r_{t,j} r_{t,j+h}.$$

Nonetheless, the proposed estimators by Zhou (1996), with $H = 1$ for Equation (4.16), and from Hansen and Lunde (2004) are not consistent.

Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008) tried to overcome this drawback through the *flat-top kernel-based* estimator

$$RV_t^{(BHLS)} = RV_t + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)\left(\hat{\gamma}_h + \hat{\gamma}_{-h}\right) \tag{4.17}$$

where $k(x)$ for $x \in [0,1]$ is a non-stochastic weight function such that $k(0) = 1$ and $k(1) = 0$.

In the early phases of RV models, an alternative way to mitigate the effects of the microstructure noise was to pre-filter the intraday returns. For example, Bollen and Inder (2002) relied on an autoregressive (AR) filter, while Ebens (1999), Maheu and McCurdy (2002) and Andersen, Bollerslev, Diebold, and Ebens (2001) used a moving average filter.

4.4 Realized Covariance

The realized variance approach may be extended in the multivariate framework. Considering n financial activities, p_t is a $n \times 1$ vector, μ and W are vectors of n-dimensional processes, the logarithmic price process may be defined as:

$$dp(t) = \mu(t) + \Omega(t)dW(t) \qquad\qquad t = 1, 2, \ldots$$

where $\Omega(t)$ is a $n \times n$ matrix defined as instantaneous co-volatility such that $\Sigma_t = \Omega(t)\Omega(t)'$ is the matrix of instantaneous covariances, assuming that $\Omega(t)$ is orthogonal to $W(t)$.

The quadratic variation of $p(t)$ is equal to

$$[p,p]_t = \int_{t-h}^{t} \Sigma_s ds = \int_{t-h}^{t} \Omega(s)\Omega(s)' ds.$$

In this context, the quadratic variation coincides with the integrated variance that may be defined as

$$ICov_t = \int_{t-h}^{t} \Sigma(s) ds. \tag{4.18}$$

Andersen, Bollerslev, Diebold, and Labys (2003) proved that the realized covariance

$$RCov_t = \sum_{i=1}^{m} r_{i,t} r_{i,t}' \tag{4.19}$$

is a consistent estimator of the integrated covariance, where m is the number of intra-daily partitions.

In the multivariate framework, since the negotiations are not simultaneous among the assets, it may emerge the "Epps effect" (Epps, 1979), which shows that the correlation between the assets tends to be under-estimated for the effect of non-synchronicity. Several studies have tried to overcome this effect, see Bandi and Russell (2005) for further details.

5 Modelling and Forecasting Realized Covariance

Since volatility forecasting models are largely employed in empirical application, such as portfolio optimisation, option pricing and risk hedging, a large part of the literature has been trying to best approximate volatility dynamics. Realized variance has allowed researchers to focus on the specification of the forecasting models, giving rise to three strands of literature: time series model of realized variance, Mixed Data Sampling and Realized GARCH models.

5.1 Time series models

Firstly, Andersen, Bollerslev, Diebold, and Labys (2003), Oomen (2001) aimed at analysing the persistence of the logarithm of the variance through a fractionally integrated ARMA model. The model introduced by Andersen, Bollerslev, Diebold, and Labys (2003) can be specified as follows:

$$\Phi(L)(1-L)^d(y_t-\mu)=\Theta(L)\varepsilon_t \tag{5.1}$$

where y_t is the logarithm of the realized volatility, L is the lag operator, $\Phi(L)=1-\Phi_1 L.....\Phi_p L^p$, $\Theta(L)=1+\Theta_1 L+.....+\Theta_q L^q$ and $(1-L)^d$ is the fractional difference operator defined as

$$(1-L)^d=\sum_{k=0}^{\infty}\frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(d+1)} \tag{5.2}$$

where $\Gamma(\cdot)$ is the gamma function. The d parameter can assume any values between 0 and 1, when $d=0$ the Equation (5.1) defines a $I(0)$ model, while with $d=1$ the Equation (5.1) defines a firstly integrated model. The multivariate version of this model is proposed by Halbleib-Chiriac and Voev (2011) and will be analysed in this section.

As an alternative of ARFIMA models, long-term memory can be captured also through a Heterogeneous Autoregressive (HAR) model, proposed by Corsi (2009). In the HAR model, the daily realized volatility is function of the lagged daily, weekly and monthly realized volatility

$$RV_{t+1g}^{(g)}=c^{(g)}+\beta^{(g)}RV_t^{(g)}+\beta^{(w)}RV_t^{(w)}+\beta^{(m)}RV_t^{(m)}+\varepsilon_{t+1,g} \tag{5.3}$$

where g, w and m represent the daily, weekly (5 days) and monthly (20 days) frequencies. The regressors $RV_t^{(w)}$, $RV_t^{(m)}$ are the average of the past values of RV_t scaled for the size of the frequency,

e.g. $RV_t^{(m)} = \frac{1}{20}\sum_{i=0}^{19} RV_{t-i}$. The relatively simple structure and the possibility to estimate the model through a OLS estimation have spread the use of this specification. Corsi (2009) showed that this model generates more accurate out-of-sample forecasts than short-term memory models.

Due to the non-parametric nature of the realized measure, the implementation of multivariate volatility models has become quite simple. A first attempt to model multivariate volatility is provided by Halbleib-Chiriac and Voev (2011), which suggested a VARFIMA*(p,d,q)* of the Cholesky factors of the realized covariance matrix, in order to capture the highly persistent behaviour of the volatility and to guarantee a semi-positive definite forecast matrix.

Let Y_t be a $n \times n$ realized covariance matrix at time t, where n is the number of asset, the Cholesky decomposition of Y_t is given by the lower triangular matrix, P_t, such that $P_t P_t' = Y_t$. Since the matrix Y_t is symmetric and positive definite, the elements of the matrix P_t are all real if the number of intra-daily observations is larger than n. Assuming that $X_t = vech(P_t)$ is the $m = n(n+1)/2$ Cholesky factors vector, obtained by stacking the components of the matrix P_t, the authors propose the following VARFIMA model

$$\Phi(L)D(L)[X_t - BZ_t] = \Theta(L)\varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma_t) \tag{5.4}$$

where Z_t is the $k \times 1$ vector of exogenous variables, B is the coefficients matrix of dimension $m \times k$, $\Phi(L) = I_n - \Phi_1 L - \Phi_2 L^2 ..... \Phi_p L^p$ and $\Theta(L) = I_n - \Theta_1 L - \Theta_2 L^2 - ..... - \Theta_q L^q$ are matrix lag polynomials where Φ_i, for $i = 1, ..., p$ and Θ_j, for $j = 1.....q$, are the AR- and MA- coefficient matrices and $D(L) = diag\{(1-L)^{d1}, ..., (1-L)^{dm}\}$, where $d_1 .....d_m$ are the degrees of fractional integration of each element of the vector X_t, Σ_t is the covariance matrix of ε_t. The authors assume that the roots of $\Phi(L)$ and $\Theta(L)$ lie outside the unit circle and X_t stationary when $d_i < 0,5$, as shown in Sowell (1992).

In their article, the model is estimated in final equations form[5] to limit the number of parameters to be estimated and guarantee a unique representation. The final model is estimated via quasi-maximum likelihood and its (1,d,1) specification has the following form

$$(1 - \Phi L)D(L)[X_t - c] = (1 - \Theta L)\varepsilon_t \qquad \varepsilon_t \sim N(0, \Sigma) \tag{5.5}$$

where c is a $m \times 1$ vector. The number of parameters significantly reduces from $qn^2 + (k+1)n + p$ to $2n + 2$, when $D(L) = diag\{(1-L)^{d1}, ..., (1-L)^{dn}\}$, and to $n + 3$, when $D(L) = (1-L)^d I_n$.

A recent study of Baruník and Čech (2016) suggested to model the Cholesky factors through a generalized HAR (GHAR). The proposed model is a multivariate extension of the HAR model also analysed in Halbleib-Chiriac and Voev (2011). The authors propose a system of unrelated HAR equations for all the elements of the vector of Cholesky factors, X_t, relying on a system of unrelated regressions (Zellner (1962)). Their results highlight the strong performance of this model in a portfolio optimisation problem when compared with a VARFIMA model on the Cholesky factors, the Riskmetrics model and a DCC-GARCH model.

[5] See Theil and Boot (1962). The VARFIMA(p,q) is said to be in final equations form, i.e. $\Phi(L)Y_t = \Theta(L)\varepsilon_t$, if $\Theta_0 = I_n$ and $\Phi(L) = 1 - \Phi_1 L - ..... - \Phi_p L^p$ is a scalar operator with $\Phi_p \neq 0$.

An alternative approach to guarantee a semi-positive definite realized covariance matrix, introduced by Bauer and Vorkink (2011), implies the logarithmic transformation of the matrix. This parametrization produces the log-volatilities, defined as $a_t = vech(A_t)$, where A_t is equal to

$$A_t = B_t \log(G_t) B_t' \tag{5.6}$$

where B_t and G_t are the matrices resulting from the spectral decomposition $Y_t = B_t G_t B_t'$ (see also Appendix A).

The dynamics of the log-volatilities, a_t, are modelled through a VAR(1):

$$a_t = \gamma_0 + \gamma_1 a_{t-1} + \varepsilon_t \tag{5.7}$$

where γ_0 is a vector of intercepts of dimension $n \times 1$, γ_1 is a $n \times n$ matrix of coefficients and ε_t is the vector of residuals.

Once obtained the fitted values of $\hat{a_t} \equiv E_t(a_t) \equiv \hat{\gamma_0} + \hat{\gamma_1} a_{t-1}$, the realized covariance matrix can be reconstructed. The inverse of the vech operator allows to define the matrix $\hat{A}_t$ and, finally, the exponential function returns the estimated covariance matrix, $\hat{V}_t$:

$$\hat{V}_t = \exp(\hat{A}_t). \tag{5.8}$$

The matrix $\hat{V}_t$ is positive definite by definition.

A further innovation of Bauer and Vorkink's paper is the use of exogenous variables as volatility determinants. The authors rely on a set of regressors that includes lagged dependent variables and macroeconomic/financial variables that have been proven to improve volatility forecasting accuracy. The model has the following form:

$$a_t = \gamma_0 + \gamma_1 a_{t-1} + \gamma_2 a_{t-2} + ... + \gamma_k a_{t-M} + \gamma_X X_{t-1} + \varepsilon_t. \tag{5.9}$$

Since the number of parameters to be estimated is relatively high, the authors suggest three methods to reduce it. The first method implies the use of the previously described Heterogeneous Autoregressive model. Consider the logarithmic transformation of the *bi-power* covariance matrix[6] and the vector of the elements, $a^{BP}(d)_t$, a model of the log-volatilities can be expressed as

$$a_t = \gamma_0 + \gamma_1 a^{BP}(1)_{t-1} + \gamma_5 a^{BP}(5)_{t-1} + \gamma_{20} a^{BP}(20)_{t-1} + \gamma_X X_{t-1} + \varepsilon_t \tag{5.10}$$

where $a^{BP}(1)_t$, $a^{BP}(5)_t$ and $a^{BP}(20)_t$ are the matrix-logarithms of daily, weekly and monthly multivariate bi-power covariation, respectively. In this way, the number of parameters is reduced by $(M-3)p^2$.

[6]As proposed in Barndorff-Nielsen and Shephard (2004), when a jump component is introduced in the diffusion process of the logarithmic price, p_t, a robust estimation of the IV can be computed as follows

$$BV = \sum_{i=1}^{n-1} | r_i | \cdot | r_{(i+1)} |$$

where n is the number of intra-daily partitions.

The previous approach still implies the estimation of a large number of parameters. A second-dimension reduction technique supposes that the $a_t^{BP}(d)_t$ series are driven by a small number of factors. The authors test this hypothesis by estimating the principal components of $a^{BP}(d)_t$, where $a^{BP}(d,i)$ is the i^{th} principal component of the covariation matrix.

The third methodology proposed involves the use of a latent factor approach. Assuming that the set of explanatory variables

$$Z_t = (a^{BP}(1,1)_t, ..., a^{BP}(5,1)_t, ..., a^{BP}(20,1)_t, ..., X_t) \tag{5.11}$$

is related to the unknown volatility factors, the k-th volatility factor, $v_{k,t}$, can be specified as

$$v_{k,t} = \theta_k Z_{t-1}. \tag{5.12}$$

Thus, the volatility factor is a linear combination of the set of N variables Z_t, $\theta_k = \{\theta_{k,(1)}, ..., \theta_{k,(N)}\}$ are coefficients that combine the explanatory variables. The log-volatilities are function of the volatility factors, such that

$$a_t^i = \gamma_0^i + \beta^i \theta Z_{t-1} + \varepsilon_t^i, \tag{5.13}$$

with $i = 1, ..., p$, where γ_0^i is the i-th element of the vector γ_0, β^i is the $1 \times K$ vector of loadings of log-space volatility and θ is a $k \times N$ matrix containing the coefficients on the Z_{t-1} variables for the k factors. Aggregating for the p log-volatilities, we have

$$a_t = \gamma_0 + \beta\theta Z_{t-1} + \varepsilon_t. \tag{5.14}$$

This approach has the major edge to significantly reduce the number of parameters and to allow to combine lagged volatility with exogenous explanatory variables.

Recently, Gourieroux, Jasiak, and Sufana (2009) proposed a model of multivariate volatility dynamics, called Wishart Autoregressive (WAR) model, based on the Wishart distribution of the covariance matrix. Let $X_{k,t}$, with $k = 1, ..., K$, be a vector of n Gaussian independent VAR(1) processes, then:

$$X_{k,t} = MX_{k,t-1} + \varepsilon_{k,t} \qquad \varepsilon_{k,t} \overset{\text{i.i.d}}{\sim} N(0, \Sigma). \tag{5.15}$$

The process, defined as

$$Y_t = \sum_{k=1}^{K} X_{k,t} X_{k,t}' \qquad Y_t \sim W_n(K, M, \Sigma) \tag{5.16}$$

is a Wishart Autoregressive process or order 1, where K denotes the degrees of freedom. Merging the two equations, Y_t can be written as

$$Y_t = MY_{t-1}M' + K\Sigma + \eta_t \tag{5.17}$$

where η_t is the heteroscedastic error term with zero mean and M is a matrix of dimension $n \times n$. In order that Y_t has a Wishart distribution and that Y_t is positive definite, it is necessary that K is

larger than n even if this does not usually happen in practice (Halbleib-Chiriac (2007)). Gourieroux, Jasiak, and Sufana (2009) suggested to estimate K through a method of moments (MM) estimator:

$$\hat{k} = \frac{2\left[\alpha'\hat{\Sigma}^*(\infty)\alpha\right]^2}{\hat{V}\left[\alpha' Y_t \alpha\right]}$$

where α is a $n \times 1$ vector defined as allocated portfolio and $\hat{\Sigma}^*(\infty)$ is given by $\hat{\Sigma}^*(\infty) = \hat{M}\hat{\Sigma}^*(\infty)\hat{M}' + \hat{\Sigma}^*$, $\hat{M}$ and $\hat{\Sigma}^*$ are the estimates of the MM model, $\hat{V}(\alpha' Y_t \alpha)$ is the sample variance of the portfolio volatility.

The Wishart model has been further extended by Bonato (2009), Bonato, Caporin, and Ranaldo (2009), Jin and Maheu (2012) and Halbleib-Chiriac and Voev (2011). The major pitfall of this model is that the parametric assumptions are particularly restrictive and that the only estimation technique available, a Bayesian Markov Chain Monte Carlo (MCMC), is highly computational expensive.

More recently, a part of the literature on volatility models have tried to understand, via time series model, the usefulness of exogenous variables in forecasting logarithmic realized volatility, see Paye (2012) and Christiansen, Schmeling, and Schrimpf (2012). Relying on a log-volatility model, the authors aim at identifying the exogenous determinants of volatility. Both the papers interestingly find that pure macroeconomic variables do not drive the dynamics of volatility, while financial variables strongly impact on the process of volatility.

During the last few years, a growing stream of literature is implementing non-linear models of realized volatility. In fact, it is well known that linear models do not consider some stylized facts, as asymmetric responses of returns. This remains valid also for realized volatility models. Then, Martens, De Pooter, and Van Dijk (2004) proposed to use a long memory model with asymmetries and structural breaks for realized volatility. Lately, McAleer and Medeiros (2008a) extended their paper with a smooth transition model, introducing tests for the presence of structural breaks and non-linearity tests.

5.2 MIDAS and GARCH-MIDAS models

Realized measures and high-frequency data are also used in a different framework, called Mixed Data Sampling (MIDAS) approach. The seminal papers of the MIDAS approach are Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels, Santa-Clara, and Valkanov (2006), which rely on data sampled at different frequencies in order to efficiently forecast volatility.

Let V_{t+1} be a volatility measure, as realized volatility, the MIDAS regression at time $t+h$ can be written as:

$$\hat{V}_{t+h} = \mu + \varphi \sum_{j=0}^{jmax} b(j,\theta) X_{t-j} + \varepsilon_t \tag{5.18}$$

where $jmax$ is the maximum lag considered, X_{t-j} is a set of explanatory variables and $b(j,\theta)$ is a weight function of lagged regressors. The regressors are sampled at higher frequencies than the dependent variable. The parameters of the MIDAS model are estimated through maximum likelihood.

Several papers have implemented the MIDAS approach for an empirical analysis. Becker, Clements, and O'Neill (2010) combined the use of mixed frequencies sampling with Cholesky decomposition of the realized covariance matrix. The first two steps of the procedure are the same as those showed in Halbleib-Chiriac and Voev (2011), whereas the realized covariance matrix is modelled according to the approach presented in Hansen and Lunde (2005). The authors treat the close to open period as a separate return period, so that the total 24-hour realized covariance matrix for day t, V_t, is computed as

$$V_t = r_{co,t} r'_{co,t} + \sum_{i=1}^{n} r_{i,t} r'_{i,t} \tag{5.19}$$

where $r_{co,t}$ is the vector of returns from closure of day $t-1$ to the opening of day t. The matrix for m days, $V_t^{(m)}$, is defined as the sum of daily covariance matrices, V_t. The Cholesky decomposition is given by $V_t^{(m)} = C_t^{(m)} C_t^{(m)\prime}$, while the vector of elements of the lower triangular C_t is defined as $P_t^{(m)} = vech(C_t^{(m)})$.

The $n(n+1)/2$ elements of $P_t^{(m)}$ are modelled through the MIDAS approach, which implies a weighted average of their past values:

$$P_{i,t+m}^{(m)} = \beta_{i0} + \beta_{i1} \sum_{k=1}^{K} B(k,1,\theta_i) P_{i,t-k+1} + v_t \tag{5.20}$$

where B is a weighting function, in this case a beta function, such that

$$B(k,i,\theta_i) = \frac{f(\frac{k}{K},1,\theta_i)}{\sum_{k=1}^{K} f(\frac{k}{K},1,\theta_i)} \tag{5.21}$$

$$f(z,a,b) = \frac{z^{a-1}(1-z)^{b-1}\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \tag{5.22}$$

where K is the maximum number of lags and β_{i0}, β_{i1} and θ_i are the parameters to be estimated. Including explanatory variables in the model, it becomes

$$P_{i,t+m}^{(m)} = \beta_{i0} + \beta_{i1} B(k,1,\theta_i) P_{i,t-k+1} + \beta_{ix} B(k,1,\theta_{ix}) X_{t-k+1} + v_t. \tag{5.23}$$

In order to forecast $P_{i,t+m}^{(m)}$, an estimation of the Cholesky-MIDAS model is necessary through a non-linear least squares (NLS) regression. Once obtained the estimation estimated parameters, m step ahead forecasts $P_{i,t+m}^{(m)}$ may be produced and, consequently, the forecast covariance matrix.

An interesting analysis of different specifications of the weighting function, B, is provided by Ghysels, Rubia, and Valkanov (2009). The authors compare the forecasts obtained with a MIDAS model with those from a GARCH model and an AR model on the realized variance. The out-of-sample forecasts are given by the following equation

$$\tilde{V}_{t+1}^{k} = \mu_k + \varphi_k \sum_{j=0}^{j^{max}} b_k(j,\theta) r_{t-j}^2 + \varepsilon_{k,t} \tag{5.24}$$

where $\tilde{V}_{t+1}^{k}$ is a measure of volatility, like the realized variance, such that $\tilde{V}_{t+1}^{k} = RV_{t+1}^{k} = \sum_{j=1}^{k} r_{t+j}^2$, $b_k(j,\theta)$ is a weight function. μ_k, ϕ_k and θ must be estimated via quasi-maximum likelihood. The

authors show that the MIDAS approach provides more accurate forecasts than the competing models, in terms of mean square error (MSE) and accordingly to West (1996) and Giacomini and White (2006) tests.

The MIDAS approach is usually combined with a GARCH model, as proposed in the Spline-GARCH model of Engle and Rangel (2008). The short-term component is modelled through a GARCH process moving around a long-term trend, the long-term component is modelled with a *Spline* function[7]. Recently, Engle, Ghysels, and Sohn (2013) introduced the GARCH-MIDAS model, in order to understand the effect of macroeconomic and financial variables on return volatility.

In the Spline-GARCH, proposed by Engle and Rangel (2008), the returns follow a process defined as

$$r_{i,t} - E[r_{i,t} \mid I_{i-1,t}] = \sqrt{g_{i,t}\tau_t} Z_{i,t} \tag{5.25}$$

where $r_{i,t}$ are the daily logarithmic returns, $I_{i-1,t}$ is the available information at day i, $Z_{i,t} \overset{iid}{\sim} (0,1)$ are the innovations, $g_{i,t}$ is a GARCH process and τ_t is an exponential spline function. The volatility can be denoted by two components, a short-term component for analysing the daily fluctuations, $g_{i,t}$, and the long-term component, τ_t. In the GARCH-MIDAS approach, the spline function is replaced with a MIDAS equation.

Engle, Ghysels, and Sohn (2013) combined a GARCH-MIDAS model with the use of exogenous macroeconomic variables, focusing on inflation rate and industrial production growth. Starting from Equation (5.25), the equation of returns at day i and month[8] t has the following form

$$r_{i,t} = \mu + \sqrt{\tau_t g_{i,t}} Z_{i,t} \qquad \forall i = 1, ..., N_t \tag{5.26}$$

where N_t is the number of days included in t and μ is the conditional average of $r_{i,t}$. $g_{i,t}$ follows a GARCH process:

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t}. \tag{5.27}$$

Following the literature on realized variance determinants (Schwert (1989)), the authors model τ_t as a function of monthly realized variance, RV_t. According to the MIDAS scheme, the long-run component can be computed as follows

$$\tau_t = m + \theta \sum_{k=1}^{K} \varphi_k(w_1, w_2) RV_{t-k} \tag{5.28}$$

where m is a constant, θ measures the impact of the lags of r_t^2 and $RV_t = \sum_{i=1}^{N_t} r_{i,t}^2$. The weight function,

[7] A spline function has the purpose to interpolate a set of points in an interval, through a set of polynomials combined together. See also Wold (1976) for further details on spline functions.

[8] Also lower frequencies are allowed, such as quarterly.

$\varphi_k(w_1, w_2)$, is specified in a twofold way:

$$\begin{cases} \frac{(k/K)^{w_1-1}(1-k/K)^{w_2-1}}{\sum_{j=1}^{K}(j/K)^{w_1-1}(1-j/K)^{w_2-1}} & \text{Beta Function} \\ w^k/(\sum_{j=1}^{K} w^j) & \text{Exponential Weights} \end{cases} \tag{5.29}$$

The model can be implemented with lower frequency, such as monthly, quarterly and yearly, and the number of lags of the MIDAS component may vary considerably. The estimation is carried out through quasi-maximum likelihood.

Finally, this specification allows to embed macroeconomic variables in the model. Engle, Ghysels, and Sohn (2013) analysed a model with one and two filters. The first approach implies the use of lagged macroeconomic variables as regressors in the long-term component, such that

$$\log \tau_t = m_l + \theta_l \sum_{k=1}^{K_l} \varphi_k(w_{1,l}, w_{2,l}) X_{l,t-k}^{mv} \tag{5.30}$$

where $X_{l,t-k}^{mv}$ represents the level of a macroeconomic variable, such as inflation rate or industrial production growth, m_l is a constant and θ_l measures the impact of the lagged exogenous variable on the logarithm of the long-term component.

The "two-sided filter" model relies on past and future observations of the macroeconomic variable to model the long-term component. The univariate specification is equal to:

$$\log \tau_t = m_2 + \sum_{k=-K_l^{(l)}}^{K_f^{(l)}} \varphi_k(w_1, w_2) \theta_l^{(k)} X_{l,t+k}^{mv} \tag{5.31}$$

where the impacts of the macroeconomic variables are free to vary, then

$$\theta_l^{(k)} = \begin{cases} \theta_l^f & k \geq 0 \\ \theta_l^b & k < 0. \end{cases}$$

In their framework, the GARCH-MIDAS model provides highly accurate forecasts, in particular in the sub-sample such as the Great Depression or the period following the Second World War.

Conrad and Loch (2014) extended the model of Engle, Ghysels, and Sohn (2013), including two exogenous variables in the long-term component, the model can be written as

$$\log(\tau_t) = m + \theta^X \sum_{k=1}^{K} \varphi_k(w_1^X, w_2^X) X_{t-k} + \theta^Y \sum_{k=1}^{K} \varphi_k(w_1^Y, w_2^Y) Y_{t-k} \tag{5.32}$$

where Y_{t-k} is a second explanatory macroeconomic variable. Let Y_{t-k} be the realized monthly variance, the annual long-term component is specified as follows

$$\log(\tau_t) = m + \theta^{RV} \sum_{k=1}^{K=12} \varphi_k(w_1^{RV}, w_2^{RV}) RV_{t-k} + \theta^{MV} \sum_{k=1}^{K=12} \varphi_k(w_1^{MV}, w_2^{MV}) X_{t-k}. \tag{5.33}$$

They forecast volatility one step ahead. At the beginning of the period t, the long-term component, τ_t, is pre-determined respect to the informative set I_{t-1}, then the volatility forecast for day i in period t is given by

$$E\Big[g_{i,t}\tau_t Z_{i,t}^2 \mid I_{t-1}\Big] = \tau_t E\Big[g_{i,t} \mid I_{t-1}\Big].$$

Since $E\big[g_{i,t} \mid I_{t-1}\big] = 1 + (\alpha+\beta+\gamma/2)^{i-1}(g_{1,t}-1)$ converges to the unconditional variance of $g_{i,t}$, the forecasts tend to the long-term component, for i sufficiently large. The forecast for the period t is given by

$$E\Big[\sum_{i=1}^{N_t} g_{i,t}\tau_t Z_{i,t}^2 \mid I_{t-1}\Big] = \tau_t\Big(N_t + (g_{1,t}-1)\frac{1-(\alpha+\beta+\gamma/2)^{N_t}}{1-\alpha-\beta-\gamma/2}\Big).$$

When $g_{1,t}$ is equal to its unconditional variance, the forecast for period t is $\tau_t N_t$.

The authors analyse a large set of macroeconomic variables, like the GDP growth, the industrial production and the unemployment rate, the *term spread*[9], the GDP deflator, the CPI and others. The relevance of macroeconomic variables to forecast volatility through a GARCH-MIDAS model is further analysed in Asgharian, Hou, and Javed (2013).

The GARCH-MIDAS has been extended in the multivariate framework. Let consider a vector of n assets, in the DCC specification of Engle (2002), it follows a $r_t \sim N(\mu, H_t)$ process, where the conditional covariance matrix, H_t, can be written as in Equation 2.29. In the bivariate case, the conditional volatilities, for asset i and asset j, defined as $q_{i,t}$ and $q_{j,t}$, follow a univariate GARCH model and are estimated in a separated first stage. The estimation of the conditional covariances represents the second step of the procedure. The conditional covariance is specified as in Equation 2.32, while Q_t is equal to

$$q_{ij,t} = \overline{\rho}_{ij,t}(1-a-b) + a(u_{i,t-1}u_{j,t-1}) + b(q_{ij,t-1}) \tag{5.34}$$

where $u_{i,t}$ and $u_{j,t}$ are the standardized residuals of the univariate model and the conditional correlation is given by

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}}. \tag{5.35}$$

$q_{ij,t}$ is the short-run covariance.

Firstly, Colacito, Engle, and Ghysels (2011) proposed a combination of the DCC model with the MIDAS approach. In the DCC-MIDAS model, the conditional covariance is defined as in 5.34, the long-run correlation is specified according to the MIDAS approach:

$$\overline{\rho}_{ij,t} = \sum_{k=1}^{K_c} \varphi_k(w_k^{ij}) C_{ij,t-k} \tag{5.36}$$

where K_c is the number of the lags of the historical correlations, $C_{ij,t-k}$, specified as

$$C_{ij,t} = \frac{\sum_{k=t-N}^{t} u_{i,k}u_{j,k}}{\sqrt{\sum_{k=t-N^{ij}}^{t} u_{i,k}^2}\sqrt{\sum_{k=t-N^{ij}}^{t} u_{j,k}^2}}. \tag{5.37}$$

[9]The term spread represents the difference between the interest rate of a short-term bond and a long-term bond.

$\overline{\rho}_{ij,t}$ is the slowly moving long-run correlation and u_t is the standardized innovation.

Rewriting the set of correlations in matrix form, the model can be computed as:

$$R_t = (Q_t^*)^{-1/2} Q_t (Q_t^*)^{-1/2} \tag{5.38}$$

$$Q_t^* = diag(Q_t) \tag{5.39}$$

$$Q_t = (1-a-b)\overline{R}_t(\underline{w}_r) + a u_t u_t' + b Q_{t-1} \tag{5.40}$$

where

$$\overline{R}_t(\underline{w}_r) = \sum_{k=1}^{K_c} \phi_k(\underline{w}_r) \odot C_{t-k} \tag{5.41}$$

$$C_t = \begin{pmatrix} v_{1,t} & 0 & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & v_{n,t} \end{pmatrix}^{-\frac{1}{2}} \left(\sum_{k=t-N_c}^{t} u_k u_k' \right) \begin{pmatrix} v_{1,t} & 0 & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots & v_{n,t} \end{pmatrix}^{-\frac{1}{2}} \tag{5.42}$$

$$v_{i,t} = \sum_{k=t-N_c}^{t} u_{i,k}^2 \qquad \forall i = 1,...,n \tag{5.43}$$

where $\phi_k(\underline{w}_r) = \varphi_k(\underline{w}_r)\iota'$. In this specification, a and b are imposed as common parameters across all asset combinations. Without these restrictions, the short-run dynamics can be written as

$$Q_t = G \odot \overline{R}_t(\underline{w}_r) + A \odot u_{t-1} u_{t-1}' + B \odot Q_{t-1} \tag{5.44}$$

where G, A and B are $n \times n$ matrices of parameters.

Assuming a single parameter w_r, the covariance matrix is positive definite under a small set of assumptions. It may be noticed that the matrix Q_t is a weighted average of the three matrices. Since $\overline{R}_t$ is a weighted average of correlation matrices, it is also semi-positive definite and $u_{t-1}u_{t-1}'$ is semi-positive definite by construction. When the initial matrix Q_0 is semi-positive definite, Q_t is semi-positive definite in each point.

When two or more weighting schemes are available, the matrix $\overline{R}_t$ is not semi-positive definite for each MIDAS specification. Further restrictions are necessary to ensure a semi-positive definite sequence of matrices $\left\{\phi_k\right\}_{k=1}^{K}$. To estimate the parameters of the model, the authors employ the two-step procedure of Engle (2002).

Asgharian, Christiansen, and Hou (2014) extended the work of Colacito, Engle, and Ghysels (2011) by including macroeconomic variables and lagged realized correlations in the long-run com-

ponent. The DCC-MIDAS-XC model is defined as follows

$$q_{ij,t} = \overline{\rho}_{ij,t}(1-a-b) + a u_{i,t-1} u_{j,t-1} + b q_{ij,t-1} \tag{5.45}$$

$$\overline{\rho}_{ij,t} = \frac{\exp(2\overline{z}_{ij,\tau}) - 1}{\exp(2\overline{z}_{ij,\tau}) + 1} \tag{5.46}$$

$$\overline{z}_{ij,\tau} = m_{ij} + \theta_{RC} \sum_{k=1}^{K} \varphi_k(w_1, w_2) RC_{ij,t-k} + \theta_X \sum_{k=1}^{K} \varphi_k(w_1, w_2) X^Q_{t-k} \tag{5.47}$$

$$RC_{ij,t} = \frac{\sum_{k=1}^{N_t} u_{i,k} u_{j,k}}{\sqrt{\sum_{k=1}^{N_t} u^2_{i,k}} \sqrt{\sum_{k=1}^{N_t} u^2_{j,k}}} \tag{5.48}$$

where $RC_{ij,t}$ is the realized correlation, measured on a quarterly basis, X^Q_t is a macroeconomic variable measured at the same frequency, while K is the number of the MIDAS component.

The authors include future macroeconomic variables in the model, which is then a DCC-MIDAS-XCF model. If θ_{RC} is equal to zero, the specification is in the following form:

$$\overline{z}_{ij,t} = m + \theta_X \sum_{k=1}^{K_{lag}} \varphi_k(w_1, w_2) X^Q_{t-k} + \theta_X \sum_{k=-K_{lead}}^{0} \varphi_k(w_1, w_2) X^{SPF}_{t-k|t}. \tag{5.49}$$

The future observations, X^{SPF}, are replaced by the expectation data provided by the *Survey of Professional Forecasters*. Since the combination of historical data and forecast data is quite difficult, the authors suggest treating the forecast data as an individual variables, such that

$$\overline{z}_{ij,t} = m + \theta_X \sum_{k=1}^{K_{lag}} \varphi_k(w_1, w_2) X^Q_{t-k} + \theta_{FX} \sum_{k=-K_{lead}}^{0} \varphi_k(w_1, w_2) X^{SPF}_{t-k|t}. \tag{5.50}$$

Following Engle (2002) and Colacito, Engle, and Ghysels (2011), the authors estimate the parameters of the model through a two-step quasi-maximum likelihood estimator, by maximizing the following function:

$$L = -\sum_{t=1}^{T} (T \log(2\phi) + 2\log|D_t| + u_t' D_t^{-2} u_t) - \sum_{t=1}^{T} (\log|R_t| + u_t' R_t^{-1} u_t - u_t' u_t)$$

where D_t is a diagonal matrix with standard deviations of returns on the diagonal and R_t is the conditional correlation matrix of standardized return residuals.

5.3 Realized GARCH and HEAVY model

Recently, an approach that includes a realized measure (like the realized variance, the bi-quadratic variation and the realized kernel) in the GARCH equation has been proposed. Some specifications of this method imply the use of multiple latent volatility processes, such as the Multiplicative Error Model (MEM) of Engle and Gallo (2006) and the HEAVY (High-frEquency-bAsed VolatilitY) model proposed by Shephard and Sheppard (2010). Hansen, Huang, and Shek (2011), instead, proposed a single latent volatility process, called Realized GARCH.

The variable of interest is the conditional variance, $h_t = Var(r_{t-1} \mid I_{t-1})$, where r_t is the series of returns of an asset. In the classical GARCH specification, h_t is only function of h_{t-1} and r^2_{t-1}, while in Hansen, Huang, and Shek (2011), h_t it is also function of some realized measure of volatility, x_{t-1}. A measurement equation completes the model. The Realized GARCH is then specified in the following form:

$$r_t = \sqrt{h_t} z_t \tag{5.51}$$

$$h_t = w + \beta h_{t-1} + \gamma x_{t-1} \tag{5.52}$$

$$x_t = \xi + \varphi h_t + \tau(z_t) + u_t \tag{5.53}$$

where $z_t \sim i.i.d.(0,1)$, $u_t \sim i.i.d.(0,\sigma^2_u)$. In this way, h_t is an autoregressive model of order one, $h_t = \mu + \phi h_{t-1} + w_{t-1}$, where $\mu = w + \gamma\xi$, $\phi = \beta + \varphi\gamma$ and $w_t = \gamma\tau(z_t) + \gamma u_t$. In order to consider an asymmetric response in volatility to return shocks, the authors suggest specifying $\tau(z)$ as follows

$$\tau(z) = \tau_1 z + \tau_2(z^2 - 1).$$

Their model can be easily estimated via quasi-maximum likelihood and can be extended to multiple asset, as proposed in Hansen, Lunde, and Voev (2014).

Instead, the high-frequency-based volatility (HEAVY) of Shephard and Sheppard (2010) relies on two latent processes

$$var(r_t \mid I^{HF}_{t-1}) = h_t = w + \alpha RM_{t-1} + \beta h_{t-1}, \qquad w, \alpha \geq 0, \beta \in [0,1) \tag{5.54}$$

$$E(RM_t \mid I^{HF}_{t-1}) = \mu_t = w_R + \alpha_R RM_{t-1} + \beta_R \mu_{t-1}, \qquad w_R, \alpha_R, \beta_R \geq 0, \alpha_R + \beta_R \in [0,1) \tag{5.55}$$

where I^{HF}_{t-1} is the high-frequency information, RM_{t-1} is the realized measure, w and w_R are the constants and μ_t is the latent conditional mean of the realized measure. This semi-parametric model can be extended to a more complex structure of the dynamics of μ_t and to the use of r^2_{t-1}, as in the traditional GARCH specification of h_{t-1}. The parameters are estimated through quasi-maximum likelihood.

A multivariate extension of the previous model is provided by Noureldin, Shephard, and Sheppard (2011). Let $r_{i,t}$ be the vector of intra-daily returns of dimension $n \times 1$, the realized measure, V_t, is a $n \times n$ matrix. For example, considering the realized covariance RC_t:

$$RC_t = \sum_{i=1}^{n} r_{i,t} r'_{i,t}$$

where n are the intra-daily partitions, the cross product of the daily returns is equal to $P_t = r_t r'_t$, then the HEAVY model can be written as

$$E\left[P_t \mid I^{HF}_{t-1}\right] = E\left[r_t r'_t \mid I^{HF}_{t-1}\right] = H_t$$

$$E\left[V_t \mid I^{HF}_{t-1}\right] = M_t,$$

with $E\left[r_t \mid I_{t-1}^{HF}\right] = 0$, such that H_t is the conditional covariance matrix. Among the possible parametrizations, the authors choose a BEKK specification Engle and Kroner (1995)), such that:

$$H_t = \overline{C}_H \overline{C}_H' + \overline{B}_H H_{t-1} \overline{B}_H' + \overline{A}_H V_{t-1} \overline{A}_H'$$
$$M_t = \overline{C}_M \overline{C}_M' + \overline{B}_M M_{t-1} \overline{B}_M' + \overline{A}_M V_{t-1} \overline{A}_M'$$

where $\overline{B}_H, \overline{A}_H, \overline{B}_M$ and $\overline{A}_M$ are $n \times n$ matrices with n^2 parameters, while $\overline{C}_H$ and $\overline{C}_M$ are lower triangular matrices with $n^* = n(n+1)/2$ parameters. H_t and M_t matrices are positive semi-definite for each t, when H_0 and M_0 are positive semi-definite. The high number of parameters to be estimated may lead to convergence problems in the quasi-maximum likelihood estimator. This drawback may be solved by imposing $\overline{B}_H, \overline{A}_H, \overline{B}_M$ e $\overline{A}_M$ to be scalars or diagonal matrices.

6 Forecasting Evaluation Methods

In this section, we analyse several methods to evaluate volatility forecasts. The evaluation of the forecasts accuracy represents a crucial aspect for the selection of the volatility model and is based on direct and indirect methods. Direct methods rest on a statistical evaluation of the forecasts to understand the ranking of compared models, while indirect methods are based on portfolio allocation or (Conditional) Value-at-Risk forecasting.

6.1 Direct Methods

A major pitfall that characterises the forecasts evaluation concerns the latent nature of the variable of interest, which implies the use of a proxy (see also Patton (2011)). Typically, this issue is solved through an unbiased estimator of volatility, like the squared returns. It is known, however, that the squared returns are a noisy proxy of the latent volatility. For this reason, the attention moved to the use of an unbiased estimator of volatility like the realized variance. The realized variance is a more efficient estimator respect to the squared returns, since $E[(r_t^2 - \sigma_t^2)^2 \mid I_{t-1}] = 2\sigma_t^4$, while the expected value of the realized variance is equal to $E[(RV_t^m - \sigma_t^2)^2 \mid I_{t-1}] = 2\sigma^4/m$, where σ_t^2 is the true unobservable variance and m is the number of intradaily periods. The use of such a proxy allows to implement the classical evaluation methods, like the Mincer and Zarnowitz (1969) regression and the test introduced by Diebold and Mariano (1995), but does not lead to the same ranking results obtained in presence of the observed volatility. In fact, Andersen and Bollerslev (1998) and Andersen, Bollerslev, and Meddahi (2005) showed that the tests are less powerful in presence of a proxy. Moreover, Hansen and Lunde (2006) showed the presence of a distortion in the ranking of several models, when a proxy of volatility is used. Recently, Patton (2011) contributed to this literature proving that the use of the realized variance as proxy of volatility presents less distort tests and ranking respect to other proxies.

When the forecasts from two or more competing models are available, the evaluation of the volatility forecasts is based on the ranking determined by some kind of loss function. Several papers have been focused on the definition of the necessary conditions to obtain a consistent ranking. Patton (2011) determined the necessary and sufficient conditions on the functional form of the loss function in the univariate framework. Laurent, Rombouts, and Violante (2013) extended the analysis of those conditions in the multivariate.

Let be $E[r_t \mid I_{t-1}] = 0$ and $E[r_t^2 \mid I_{t-1}] = \sigma_t^2$ the true unobservable conditional variance of the returns r_t and let be h_t a forecast from a forecasting model, the loss function is given by $L(\sigma_t^2, h_t)$. In the multivariate framework, the variable of interest is the conditional covariance matrix, $\Sigma_t = E[r_t r_t' \mid I_{t-1}]$, and the loss function is given by $L(\Sigma_t, H_t)$.

Since the true latent variance is not observable, the loss function should rely on a proxy of volatility, defined as $\hat{\sigma}_t^2$ in the univariate and $\hat{\Sigma}_t$ in the multivariate. Consequently, the loss functions can be defined as $L(\hat{\sigma}_t^2, h_t)$ and $L(\hat{\Sigma}_t, H_t)$. Patton (2011) determined the properties that a loss function should have in presence of a proxy of volatility. In particular, the loss function is considered "robust" if the ranking between the competing models is the same obtained in presence of the real conditional variance. Thus, the expected loss function, $E[L(\sigma_t^2, h_t)]$, given two competing models k and j should satisfy the following condition:

$$E[L(\sigma_t^2, h_{k,t})] \leq E[L(\sigma_t^2, h_{j,t})] \Leftrightarrow E[L(\hat{\sigma}_t^2, h_{k,t})] \leq E[L(\hat{\sigma}_t^2, h_{j,t})]. \tag{6.1}$$

The condition (6.1) is guaranteed if

$$\frac{\partial^2 L(\sigma_t^2, h_t)}{(\partial \sigma_t^2)^2} \tag{6.2}$$

exists and does not depend from h_t (cfr. Laurent and Violante (2012)).

Patton (2011) provided also necessary and sufficient conditions to define a class of homogeneous[10] loss functions in the univariate dimension, robust to the use of a proxy. If the degree of homogeneity is equal to $b+2$, the class of loss functions is given by

$$L(\hat{\sigma}_t^2, h_t; b) = \begin{cases} \frac{1}{(b+1)(b+2)}(\hat{\sigma}_t^{2b+4} - h_t^{b+2}) - \frac{1}{b+1} h_t^{b+1}(\hat{\sigma}_t^2 - h_t), & b \neq -1, -2 \\ h_t - \hat{\sigma}_t^2 + \hat{\sigma}_t^2 \log \frac{\hat{\sigma}_t^2}{h_t}, & b = -1 \\ \frac{\hat{\sigma}_t^2}{h_t} - \log \frac{\hat{\sigma}_t^2}{h_t} - 1, & b = -2 \end{cases} \tag{6.3}$$

For $b = 0$, the loss function corresponds to the mean squared error loss function (MSE), $L(\hat{\sigma}_t^2, h_t) = (\hat{\sigma}_t^2 - h_t)^2$, while for $b = 2$ the loss function is equal to a quasi-likelihood (QLIKE), $L(\hat{\sigma}_t^2, h_t) = \log h + \frac{\hat{\sigma}_t^2}{h_t}$. Patton and Sheppard (2009) proved that the DM test and the test introduced by West (1996) are more powerful when a QLIKE loss function is used respect to a MSE function, in the univariate dimension.

In the multivariate case, the necessary condition for a robust ranking becomes

$$E[L(\Sigma_t, H_{k,t})] \leq E[L(\Sigma_t, H_{j,t})] \Leftrightarrow E[L(\hat{\Sigma}_t, H_{k,t})] \leq E[L(\hat{\Sigma}_t, H_{j,t})], \tag{6.4}$$

[10] A loss function, L, is homogeneous of order k if $L(a\hat{\sigma}_t^2, ah_t) = a^k L(\hat{\sigma}_t^2, h_t)$, $\forall a > 0$.

and the sufficient condition to ensure (6.4) is equal to

$$\frac{\partial^2 L(\Sigma_t, H_t)}{\partial \sigma_{k,t} \partial \sigma_{j,t}}, \qquad \forall k, j = 1, \ldots, N(N+1)/2 \tag{6.5}$$

where $\sigma_{k,t}$ is the k-th element of $\sigma_t = vech(\Sigma_t)$. A loss function L is robust if and only if it assumes the following form:

$$L(\hat{\Sigma}_t, H_t) = \tilde{C}(H_t) + \tilde{C}(\Sigma_t) + C'(H_t) vech(\hat{\Sigma}_t - H_t), \tag{6.6}$$

where

$$C(H_t) = \begin{bmatrix} \frac{\partial \tilde{C}(H_t)}{\partial h_{1,t}} \\ \vdots \\ \frac{\partial \tilde{C}(H_t)}{\partial h_{K,t}} \end{bmatrix}, \quad C'(H_t) = \begin{bmatrix} \frac{\partial \tilde{C}(H_t)}{\partial h_{1,t} \partial h_{1,t}} & \cdots & \frac{\partial \tilde{C}(H_t)}{\partial h_{1,t} \partial h_{K,t}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \tilde{C}(H_t)}{\partial h_{K,t} \partial h_{1,t}} & \cdots & \frac{\partial \tilde{C}(H_t)}{\partial h_{K,t} \partial h_{K,t}} \end{bmatrix}$$

and where $C(\cdot)$ and $C'(\cdot)$ are the gradient and the Hessian of $\tilde{C}(\cdot)$ with respect to H_t.

Equation (6.6) is re-defined by Laurent, Rombouts, and Violante (2013) on the basis of the forecast errors, such that

$$L(\hat{\Sigma}_t, H_t) = L(\hat{\Sigma}_t - H_t) = vech(\hat{\Sigma}_t - H_t)' \hat{\Lambda} vech(\hat{\Sigma}_t - H_t) \tag{6.7}$$

where $\hat{\Lambda}$ is a positive definite matrix of weights associated to the elements of the forecast error matrix. The loss function defined in Equation (6.7) nests several loss functions, both in the vector space, like the Euclidean distance and the weighted Euclidean distance, and matrix space, like the Frobenius distance and the Stein distance.

As underlined in Laurent, Rombouts, and Violante (2013), while in the univariate case an analytical expression is available for the entire class of consistent loss functions, in the multivariate case this generalization is not feasible due to the infinite combinations between forecasts and forecasts errors that satisfy (6.7). Given (6.7), Laurent, Rombouts, and Violante (2013) proved that specific loss functions can be easily derived.

In this section, we further review several tests for forecasting accuracy that strongly rely on a statistical loss function.

Firstly, a simple method to evaluate the accuracy of volatility forecast is the Mincer-Zarnowitz (MZ) regression, introduced by Mincer and Zarnowitz (1969). This approach is based on the estimation of the coefficients of a linear regression, such that

$$\sigma_t^2 = \alpha + \beta h_t + \varepsilon_t \tag{6.8}$$

where σ_t^2 is the true conditional variance, α is a constant and β is the coefficient of the forecast at time t. A forecast is optimal when $\sigma_t^2 = h_t$, which is equivalent to the null hypothesis $H_0 : \alpha = 0 \cup \beta = 1$. Given that volatility is latent, the regression is only feasible when a reliable proxy is used, then the MZ regression can be expressed as

$$\hat{\sigma}_t^2 = \alpha + \beta h_t + \varepsilon_t. \tag{6.9}$$

The regression (6.9) gives a twofold information on the accuracy of the forecasts by testing the joint hypothesis on the coefficients and through the R^2 that can be seen as an indicator of how much the prediction is correlated with the proxy of volatility.

The MZ regression strongly depends on the accuracy of the volatility proxy that influences the estimation of the coefficients and the accuracy of the R^2. Moreover, Hansen and Lunde (2006) showed that, when a volatility proxy is used, the R^2 cannot always be considered adequate and may lead to a distort ordering.

The use of the MZ can be extended in the multivariate case. A simple approach is to estimate the regression (6.9) for each element of the covariance matrix, such that

$$\hat{\sigma}_{ij,t} = \alpha_{ij} + \beta_{ij} h_{ij,t} + \varepsilon_{ij,t} \tag{6.10}$$

where $\hat{\sigma}_{ij}$ is the element ij of the realized covariance matrix $\hat{\Sigma}$, for $i = 1,\dots,N(N+1)/2$ and $j = 1,\dots,N(N+1)/2$. When the number of series is relatively high, there emerge difficulties with this approach. A feasible alternative is to define the MZ regression as follows

$$vech(\hat{\Sigma}_t) = \alpha + diag(\beta) vech(H_t) + \varepsilon_t \tag{6.11}$$

where α and β are $N(N+1)/2 \times 1$ vectors of parameters, $vech(\cdot)$ is the half-vector operator and $diag(\cdot)$ is the operator that transforms a $k \times 1$ vector in a $k \times k$ matrix with the elements of the vector along the diagonal. The joint test that $\alpha = 0$ and $\beta = 1$ can be computed on (6.11). Patton and Sheppard (2009) underlined the possibility of adversely affected results in finite sample, proposing to impose a constraint on the parameters, such that $\alpha = \alpha_i$ and $\beta = \beta_i$, $\forall i = 1,\dots,N(N+1)/2$.

Some of the most used methods of forecast evaluation can only be applied when different models have been implemented on the same dataset and shall be based on the forecast error. These measures, relying on the forecast error and the relative transformations, include the Mean Squared Error (MSE), the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE). The former measures are the most common in the forecasting evaluation literature, although several authors suggest that MAE is a more reliable measure, since it is less sensitive to outliers of the forecast error.

MAE measures the accuracy of the forecasts through the average of the magnitude of the forecast error and can be computed as

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \mid e_i \mid$$

where e_i is the forecast error, for $i \in \tilde{n}$, where $\tilde{n} = N(N+1)/2$.

MSE and RMSE are robust measures when the competing models are based on the same series. Since they strictly depend on the forecast error, it follows that the smaller the measure, the better

the accuracy. In the univariate case, the measures can be written as

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2}$$

where n is the number of forecast errors. In the multivariate case, the RMSE can be based on the Frobenius norm[11] of the $n \times \widetilde{n}$ matrix of the forecast error, and it is defined as follows

$$RMSE = \frac{\|e\|_2}{\sqrt{(n \cdot \widetilde{n}) - 1}}.$$

In the multivariate framework, a model should also be preferred if it exhibits the lowest RMSE.

For the pairwise comparison of competing models, two tests are usually implemented, the DM and the GW test. The equal predictive ability tests share the null hypothesis of absence of predictive ability. The test introduced by Diebold and Mariano (1995) is considered the first attempt to compare two rival models in terms of forecasting accuracy, jointly with the test proposed by West (1996). The DM test is based on assumptions made on the difference of the forecast error loss functions.

Assuming a robust loss function, like those defined in this section, the DM test is based on the differential of the loss functions for the models k and j, such that

$$d_t = L(\hat{\sigma}_t^2, h_{k,t}) - L(\hat{\sigma}_t^2, h_{j,t}), \tag{6.12}$$

in the univariate case and

$$d_t = L(\hat{\Sigma}_t, H_{k,t}) - L(\hat{\Sigma}_t, H_{j,t}), \tag{6.13}$$

in the multivariate case, where $L(\hat{\Sigma}_t, H_{k,t})$ is a loss function as the Euclidean distance between vectors and the Frobenius distance between matrices.

The null hypothesis of equal predictive ability can be expressed as $H_0 : E[d_t] = 0$; the test assumes the following form

$$DM = \sqrt{T}\frac{\overline{d}}{\sqrt{w}} \xrightarrow{d} N(0,1) \tag{6.14}$$

where

$$\overline{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$$

and

$$w = \lim_{t\to\infty} VAR(\sqrt{T}d)$$

is its asymptotic variance, generally estimated through the sample variance.

[11] The Frobenius norm of a matrix is a Euclidean norm, built on matrix A, of dimension $m \times n$, equal to the square root of the summed squares of the matrix elements, defined as $\|A\|_2 = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2}$.

The bulk of the literature relies on a MSE loss function in the DM test, such that $L(\hat{\sigma}_t^2, h_t) = (\hat{\sigma}_t^2 - h_t)^2$, while in the multivariate framework two loss functions are usually implemented, the Frobenius norm between matrices defined as $L_F(\hat{\Sigma}_t, H_t) = \sum_{ij}(\hat{\sigma}_{ij,t} - h_{ij,t})^2$, and the Euclidean distance between vectors, that can be specified as $L(\hat{\sigma}_t, h_t) = \sum_{k=1}^{N(N+1)/2}(\hat{\sigma}_{k,t} - h_{k,t})^2$.

Giacomini and White (2006) extended the DM test to consider the previous information set in the test, computing a conditional test of superior predictive ability. More specifically, the authors proved that, given the volatility forecasts from two competing models, the null hypothesis may be defined as

$$H_0 : E\left[(Y_{t+1} - f_t(\hat{\beta}_{1t}))^2 - (Y_{t+1} - g_t(\hat{\beta}_{2t}))^2 \mid I_t\right] = 0 \tag{6.15}$$

where $f_t(\hat{\beta}_{1t})$ and $g_t(\hat{\beta}_{2t})$ are two forecasting models for the conditional mean of the variable of interest Y_{t+1}, given a quadratic loss function.

The GW test is the best candidate to evaluate forecasts from a rolling window scheme, while it does not allow the use of a recursive scheme. When the test is based on a rolling window scheme, it allows to remove some of the assumptions made for other pairwise tests like the Diebold and Mariano (1995) and West (1996). In particular, the hypothesis of stationary observations is no longer needed, while the test can be applicable to a wider class of models, including linear and non-linear models, semi-parametric or non-parametric models, nested or not-nested models.

For nested models, Clark and West (2007) introduced an equal predictive ability test based on the mean squared prediction error (MSPE), defined as $MSPE_i = \sum_t^T (\hat{\sigma}_t^2 - h_{i,t})^2$, where $i = 1,2$. Let be $\hat{f}_t = MSPE_1 - MSPE_2 + T^{-1}\sum_t^T (h_{1,t} - h_{2,t})^2)$ and its average $\bar{f} = T^{-1}\sum_t^T \hat{f}_t$, the adjusted test for the squared forecasts difference can be computed as

$$CW = \sqrt{T}\bar{f} / \sqrt{\sigma_{\hat{f}_t}^2} \tag{6.16}$$

where $\sigma_{\hat{f}_t}^2 = T^{-1}\sum_t^T (\hat{f}_t - \bar{f})^2$. The null hypothesis of the test is given by $MSPE_1 = MSPE_2$, while the alternative hypothesis is given by $MSPE_2 < MSPE_1$, determining the CW test as a unilateral test.

Alternatively to pairwise tests, forecasts may be compared for more than two models. At this end, Hansen, Lunde, and Nason (2011) introduced the model confidence set (MCS) to compare all forecasts against each other. For a given confidence level, the MCS defines the set of models containing the best out-of-sample forecasts. The MCS approach consists in a sequential procedure that allows to test the equal predictive ability of the compared models, discard any inferior model and define the set of superior models (SSM). Given a set of M^0 forecasts, the MCS procedure tests whether all models in M^0 have equal forecasting ability. The performance is measured pairwise by the loss functions difference, $d_{k,j,t} = L(\sigma_t, h_{k,t}) - L(\sigma_t, h_{j,t})$, for all k, $j \in M^0$ and $k \neq j$. Assuming that $d_{k,j,t}$ is stationary, the null hypothesis takes the following form:

$$H_0 : E[d_{k,j,t}] = 0, \quad \forall k, j \in M^0. \tag{6.17}$$

A model is discarded if the null is rejected at a given confidence level α. The test is sequentially repeated until the non-rejection of the null. The remaining models define the set of statistically

equivalent models with respect to a given loss function. As for pairwise models, also for MCS may be implemented robust loss functions like MSE, Frobenius norm between matrices and Euclidean distance among vectors.

6.2 Indirect methods

An alternative to statistical evaluation of the forecasts is evaluating predictive ability through indirect methods.

Volatility is implemented in many economic decisions, like mean-variance portfolio optimisation, hedging risk measurement, option pricing and utility maximisation. In general, economic evaluation of volatility and correlation forecasts relies on several assumptions, such as the utility function of the hypothetical investor (in portfolio choice or hedging applications), the density of the standardised returns (as in Value-at-Risk and Expected Shortfall forecasting, density forecasting portfolio choice applications with non-quadratic utility), the derivative pricing model (in option, and other derivative securities, pricing applications). Although "non-robust", these approaches can yield valuable information on competing volatility and correlation forecasts.

6.2.1 Portfolio Optimisation

The use of covariance in portfolio optimisation has its roots in the doctoral thesis of Markowitz. Since then, a wide stream of literature has been analysing the forecast accuracy in a context of portfolio optimisation. Recently, the same literature has been wondering if a realized covariance matrix does provide tangible advantages in the definition of the weights of a portfolio. Firstly, Fleming, Kirby, and Ostdiek (2003) analysed the effects of the use of the realized covariance matrix, highlighting the increased performance in terms of risk and return for the underlying portfolio. A great part of the papers has investigated the realized covariance matrix in a Global Minimum Variance approach, Kyj, Ostdiek, and Ensor (2009), Halbleib-Chiriac and Voev (2011) and Hautsch, Kyj, and Malec (2015), among the others, confirm the usefulness of the non-parametric measure in the process of investment decision.

Markowitz (1952) succeeded to synthesize the choices of portfolio allocation in two quantitative variables, the mean and the standard deviation of the portfolio. According to Markowitz (1952), an investor takes her decisions based on her expected utility function. Following the approach of von Neumann and Morgenstern (1947), the investor aims at maximising her expected utility, reducing the portfolio optimisation problem to the following form

$$\max_{w} \quad E[U(W)]$$

where w are the optimal portfolio weights and $U(W)$ is the utility of the investor function of her wealth, W. Assuming returns normally distributed, a convenient choice is to define $U(\cdot)$ as a *constant*

absolute risk adversion[12] (CARA), given by

$$U(W) = -\exp\{cW\}, \quad c > 0 \tag{6.18}$$

where c is the Arrow-Pratt measure of risk-aversion. The expected utility to be maximized becomes

$$E[U(W)] = -\exp\{-c[\mu_p - \frac{c}{2}\sigma_p^2]\}. \tag{6.19}$$

Maximizing the expected utility when the utility function is exponential implies to find the solution to the following problem

$$\max_w \quad \mu_p - \frac{1}{2}c\sigma_p^2 \tag{6.20}$$

under the constraint

$$w'\iota = 1. \tag{6.21}$$

Equation (6.21) bounds the sum of the weights to one. Equation (6.20) can be computed as a risk minimization problem given an objective return. The optimum problem can be written as

$$\begin{aligned} \min_w \quad & w'\hat{\Sigma}w \\ s.t. \quad & w'\hat{\mu} = \mu_p \\ & w'\iota = 1 \end{aligned} \tag{6.22}$$

where w is the weights vector, $\hat{\Sigma}$ is the estimated covariance matrix. The optimum problem is constrained to the full investment of the capital, $w'\iota = 1$, and to the objective return, $w'\hat{\mu} = \mu_p$, while the vector $\hat{\mu}$ is an estimation of the expected value of the returns of the assets composing the portfolio.

The mean-variance remains the most used approach in portfolio optimisation, for its simple implementation and for the possible extensions. Although appealing, the approach introduced by Markowitz (1952) has been criticized due to the symmetric nature of the risk measure, i.e. the covariance matrix equally responds to positive and negative shocks (see also Hanoch and Levy (1969)), and the poorly realistic assumptions of a CARA utility function and normal distribution of returns.

The mean-variance approach has been further criticized for relying on both the first two conditional moments of the returns. There are well-known problems, however, concerning the prediction of the first conditional moment. For this reason, the Global Minimum Variance (GMV), which focuses on the prediction of the conditional covariances avoiding to define a process for μ_p, has been implemented and has been mostly diffused in allocation problem involving a specific forecasting model for the covariance matrix (e.g. DCC, BEKK, Realized Covariance). Some articles show how portfolios obtained from a mean-variance approach are less stable and perform worse than the competing portfolios obtained from a GMV problem, see also Chan, Karceski, and Lakonishok (1999), Jagannathan and Ma (2003) and Kyj, Ostdiek, and Ensor (2009).

[12] In this kind of function the Arrow-Pratt measure of risk-aversion, expressed as $A(W) = -\frac{U''(W)}{U'(W)}$, is constant.

The GMV weights can be derived from the following optimum problem

$$\begin{aligned} \min_{w} \quad & w'\hat{\Sigma}w \\ s.t. \quad & w'\iota = 1 \end{aligned} \tag{6.23}$$

where $\hat{\Sigma}$ is the forecast of the conditional covariance matrix of portfolio. The underlying weights are equal to

$$w = \frac{\hat{\Sigma}^{-1}\iota}{\iota'\hat{\Sigma}^{-1}\iota}. \tag{6.24}$$

As seen in this section, classical optimisation problems are usually based on the covariance matrix as risk measure. However, the financial operators have been investigating alternative risk measure to account also for empirical evidences not included in covariance matrix. This lead the financial institution J.P. Morgan to introduce, in 1994, a risk measure called Value-at-Risk (VaR) that would have been the most used in the financial institutions in the following two decades, due also to its legal implementation in the European Union (see Basilea II). The Value at Risk is defined as the maximum portfolio loss at a given confidence level, α, in a time interval, formalized as follows

$$P(R_p \leq -VaR_\alpha) = \alpha \tag{6.25}$$

where R_p is the portfolio return. It follows that the greater the VaR, the larger the risk of the portfolio. A risk-adverse investor would prefer to minimize the VaR of the portfolio.

There exist several methods to compute VaR in the financial literature, the most common are the mean-variance approach, the historical simulation method, the use of the Monte Carlo simulations, and the Extreme Value Theory, see Kuester, Mittnik, and Paolella (2006) for a comprehensive review of these methods. Since in this article we focused on the realized variance, we only analyse the method introduced by Giot and Laurent (2004) based on the realized volatility. Let be r_t the returns at time t of a single asset, it is assumed that

$$r_t = \sqrt{h_t} z_t, \qquad z_t \sim F \tag{6.26}$$

where h_t is the conditional variance, z_t is a i.i.d. variable with unitary variance and F is the cumulative function of the returns. The one-step ahead forecast of the VaR is given by

$$\hat{VaR}^{\alpha}_{t+1|t} = -\hat{F}^{-1}\sqrt{h_t}$$

assuming h_t as known conditionally to time t. For example, a GARCH model can be implemented to forecast the conditional variance one-step ahead and, assuming a certain distribution for F, the quantile of the distribution can be defined, and, consequently the VaR. When the VaR is the object of the minimization problem, it becomes

$$\begin{aligned} \min_{w} \quad & VaR_\alpha \\ s.t. \quad & w'\iota = 1. \end{aligned} \tag{6.27}$$

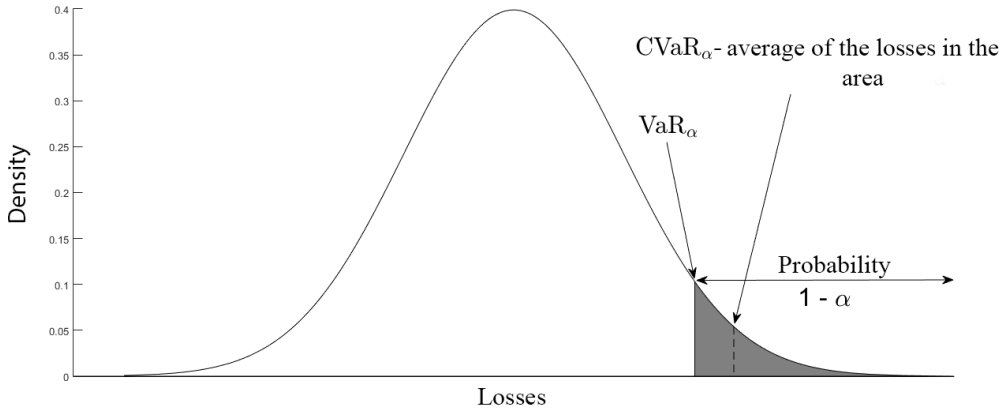
Despite the advantages denoting the Value-at-Risk, it also presents several weaknesses. The most relevant, in terms of portfolio optimisation, is the not-additivity property. In fact, among the properties defined by Artzner, Delbaen, Eber, and Heath (1999), a risk measure should be sub-additive, meaning that a diversification should reduce the overall risk or leave it not alternated. Instead, for two portfolios, X and Y, the VaR obtained from a combination of the two may be greater than the sum of the single VaR, i.e. $VaR(X+Y) > VaR(X) + VaR(Y)$. Moreover, the VaR does not dive any information about the losses exceeding the maximum loss at a given confidence level. For these reasons, a more coherent risk measure like the Conditional VaR (or Expected Shortfall) should be implemented.

The Conditional Value-at-Risk (CVaR) denotes the maximum loss conditional to the fact that this happens in the tail of the distribution at left of the VaR. Let L be a continuous variable representing the losses of a portfolio, the CVaR may be defined as

$$CVaR_\alpha = E\big[L \mid L \geq VaR_\alpha\big]. \tag{6.28}$$

In other terms, the CVaR is equal to the average of the losses exceeding the VaR (see Figure 1). If

Figure 1: Conditional Value at Risk



the returns are normally distributed, then

$$CVaR_\alpha = -\left(\mu - \sigma \frac{\Phi(N^{-1}(\alpha))}{\alpha}\right) \tag{6.29}$$

where $\Phi(N^{-1}(\alpha))$ is the cumulative function of the quantile α. Let be the average of the returns null, the formula can be written as

$$CVaR_\alpha = \sigma \frac{\Phi(N^{-1}(\alpha))}{\alpha}. \tag{6.30}$$

When the returns are normally distributed, for $\alpha \to 0$, then $VaR_\alpha \to CVaR_\alpha$ (cfr. Barr (2013)).

Contrarily to VaR, the CVaR is a coherent risk measure according to the definition of Artzner, Delbaen, Eber, and Heath (1999), since the sub-additivity property allows to optimally capture the effects of a portfolio diversification.

Using the definition of the CVaR in (6.30), the optimum problem is given by

$$\begin{aligned} &\min_{w} \quad CVaR_{\alpha} \\ &s.t. \quad w'\iota = 1. \end{aligned} \tag{6.31}$$

6.2.2 Risk hedging

A possible criterion to evaluate volatility forecasts implies forecasting Value-at-Risk (VaR).

The VaR and the related measure, like the Conditional-Value-at-Risk (CVaR), are an operative standard in financial institutions.

These measures are based on the quantiles of the distribution, specifically on extreme losses. The main advantage of this kind of risk measure is that only few information are necessary to define them. Moreover, they can be determined through econometric models similar to those presented in the previous chapter. The way VaR and CVaR are computed is described in the previous section, in this section we focus on the evaluation of quantiles forecasts.

Ex-post evaluation methods are necessary in presence of VaR and CVaR forecasts. We introduce two tests of VaR/CVaR forecast accuracy: the unconditional test of Kupiec (1995) and the conditional test introduced by Christoffersen (1998). In order to determine the tests, it becomes necessary to define the index variable. Let consider the returns of a financial asset, the index variable is equal to

$$\eta_t = \begin{cases} 1 & se \quad r_t < -VaR \\ 0 & se \quad r_t \geq -VaR, \end{cases} \tag{6.32}$$

where 1 denotes an exception and 0 denotes a return lower than the VaR (or CVaR for the conditional measure). The exceptions are summed and divided for the total number of the out-of-sample VaR estimates to gather an empirical measure.

The unconditional coverage test, introduced by Kupiec (1995), is based on the assumption that the frequency of the exceptions empirically detected, ϕ, is coherent with the theoretical frequency, α. The statistical test can be computed as

$$LR_{uc} = -2\Big[\ln(\alpha^{x}(1-\alpha^{N-x})) - \ln(\phi^{x}(1-\phi)^{N-x})\Big] \sim \chi^{2}(1), \tag{6.33}$$

where x is the number of exceedances, N is the sample size. Thus, a rejection of the null hypothesis implies that the dimension of the empirical VaR is significantly different from the nominal VaR.

The Kupiec (1995) test can be applied only when the exceptions are independent, when the exceptions are not independent and clustered, the analysis should rely on Christoffersen (1998) test. The test introduced by Christoffersen (1998) is the most common method to evaluate the performance of VaR models. Let be the probability of two consecutive exceptions equal to

$$p_{ij} = P(\eta_t = 1 \mid \eta_{t-1} = j),$$

where η is the index function as in the Equation (6.32). Two exceptions are considered independent if they do not occur in two consecutive days. Christoffersen (1998) suggested a likelihood test with time independent observations as null hypothesis and violations following a Markov chain as alternative hypothesis. Assuming that the violations follow a Markov chain with the following transition function

$$\Pi = \begin{pmatrix} \pi_{0,0} & \pi_{1,0} \\ \pi_{0,1} & \pi_{1,1,} \end{pmatrix} \tag{6.34}$$

where 1 represents a violation and 0 denotes a non-violation. Let be n_0, n_1, n_{00}, n_{01}, n_{10}, n_{11} the stages or the transitions of the stochastic Markov process, then

$$\pi_{00} = \frac{n_{00}}{n_{00}+n_{01}}, \quad \pi_{01} = \frac{n_{01}}{n_{00}+n_{01}} \tag{6.35}$$

$$\pi_{10} = \frac{n_{10}}{n_{10}+n_{11}}, \quad \pi_{11} = \frac{n_{11}}{n_{10}+n_{11}}. \tag{6.36}$$

Let be $\pi_0 = n_0/N$, $\pi_1 = n_1/N$, the LR test can be defined as follows

$$LR_{ind} = -2\ln\left[(\pi_0^{n_0}\pi_1^{n_1}) - \ln(\pi_{00}^{n_{00}}\pi_{01}^{n_{01}}\pi_{10}^{n_{10}}\pi_{11}^{n_{11}})\right] \sim \chi^2(1). \tag{6.37}$$

For a confidence level equal to 95%, if $LR_{ind} > LR_{critical} = 3.841$, the null hypothesis is rejected, and the violations are not independent. Thus, the model fails the independence test.

The investors are, however, often not interested in the performance of a self-standing model for the VaR, but in a comparison with others VaR models. The literature provided several ways to rank two or more models, like the *quadratic probability score function* (Lopez (1998)) and the *quadratic score function* (Blanco and Ihle (1999)). Both the functions allow to measure the performance of a model related to another model.

Lopez (1998) introduced the *quadratic probability score* to measure the pairwise comparison of VaR (CVaR) models, specified as follows

$$QPS = \frac{2}{n}\sum_{t=1}^{n}(C_t - p)^2, \tag{6.38}$$

where n is the number of observations, p is the expected probability of a violation (i.e. the actual loss is larger than the estimated VaR/CVaR). C_t is a loss function. Lopez (1998) relies on a loss function defined as

$$C_t = \begin{cases} 1 & se \quad L_t > VaR_t \\ 0 & se \quad L_t \leq VaR_t. \end{cases} \tag{6.39}$$

QPS function assumes a value comprise between 0 and 2. Under general conditions, accurate estimates of VaR (CVaR) generate the least possible score, thus a lower QPS indicates a better performance in terms of violations.

In addition to the QPS function, a common method for the evaluation of VaR forecasts is the already mentioned Root mean square error. Let VaR_t be the estimated VaR (CVaR) and L_t the real

loss, the RMSE can be expressed as

$$RMSE = \sqrt{E[(VaR_t - L_t)]} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(VaR_t - L_t)^2}. \tag{6.40}$$

A lower RMSE denotes a greater predictive accuracy of the model.

7 Conclusion

In this article, we provided an overview of volatility measures, focusing on volatility forecasting and evaluation methods, considering both the univariate and multivariate settings.

We discussed the different types of volatility and the methods to measure them, which can be divided into three specific categories, namely ARCH models, stochastic volatility models and realized volatility models.

The review has been particularly focused on predictive models for the realized variance, introduced in the recent literature of the non-parametric measure of volatility, analysing in detail time series models, MIDAS models and a combination of parametric models, like GARCH, and non-parametric models, named Realized GARCH models.

We paid particular attention to forecast accuracy evaluation methods, considering the drawbacks related to the latent nature of the conditional variance. In fact, in direct methods, it becomes highly relevant the use of a volatility proxy which may lead to distortions in the ordering between forecasts.

Thus, this article provided the conditions to define a robust loss function. Since the distortions introduced by the use of some volatility proxy may be avoided by an appropriate choice of the loss function.

Finally, since researchers are usually interested in the economic evaluation of the forecasts, indirect methods have been analysed to evaluate volatility forecasts, such as portfolio optimisation and risk hedging.

References

AÏT–SAHALIA, Y., P. A. MYKLAND, AND L. ZHANG (2006): "Comment," *Journal of Business and Economic Statistics*, 24, 162–167.

ALEXANDER, C., AND A. M. CHIBUMBA (1996): "Multivariate orthogonal factor GARCH," Discussion Paper in mathematics, University of Sussex.

AMIHUD, Y., AND H. MENDELSON (1987): "Trading mechanisms and stock returns: an empirical investigation," *Journal of Finance*, 42, 533–553.

ANDERSEN, T. G. (2009): "Stochastic volatility," in *Encyclopedia of Complexity and Systems Science*. Springer Verlag.

ANDERSEN, T. G., AND T. BOLLERSLEV (1998): "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

ANDERSEN, T. G., T. BOLLERSLEV, P. CHRISTOFFERSEN, AND D. F. X. (2006): "Volatility and correlation forecasting," in *Handbook of Economic Forecasting*, pp. 778–878. Amsterdam: North-Holland.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND H. EBENS (2001): "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43–76.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2000): "Exchange rate returns standardized by realized volatility are (nearly) Gaussian," *Multinational Finance Journal*, 4, 159–179.

——— (2001): "The distribution of exchange rate volatility," *Journal of American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

ANDERSEN, T. G., T. BOLLERSLEV, AND N. MEDDAHI (2005): "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities," *Econometrica*, 73(1), 279–296.

ARTZNER, P., F. DELBAEN, J.-M. EBER, AND D. HEATH (1999): "Coherent measures of risk," *Mathematical Finance*, 9, 203–228.

ASAI, M., M. MCALEER, AND J. YU (2006): "Multivariate stochastic volatility: A review," *Econometric Reviews*, 25, 145–175.

ASGHARIAN, H., C. CHRISTIANSEN, AND A. J. HOU (2014): "Macro-Finance Determinants of the Long-Run Stock-Bond Correlation: The DCC-MIDAS Specification," Working paper.

ASGHARIAN, H., A. J. HOU, AND F. JAVED (2013): "The Importance of the Macroeconomic Variables in Forecasting Stock Return Variance: A GARCH-MIDAS Approach," *Journal of Forecasting*, 32, 600–612.

BACK, K. (1991): "Asset pricing for general processes," *Journal of Mathematical Economics*, 20, 371–395.

BAILLIE, R. T., T. BOLLERSLEV, AND H. O. MIKKELSEN (1996): "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 74, 3–30.

BANDI, F. M., AND J. R. RUSSELL (2005): "Realized covariation, realized beta and microstructure noise," Unpublished paper, Graduate School of Business, University of Chicago.

——— (2006): "Volatility," in *Handbook of Financial Engineering*. Elsevier.

——— (2007): "Microstructure noise, realized volatility, and optimal sampling," Unpublished paper, Graduate School of Business, University of Chicago.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realized kernels to measure the ex post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

BARNDORFF-NIELSEN, O. E., AND N. SHEPHARD (2002a): "Econometric analysis of realised volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society*, 64, 253–280.

——— (2002b): "Estimating Quadratic Variation Using Realised Variance," *Journal of Applied Econometrics*, 17, 457–477.

——— (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

BARR, D. G. (2013): "Value at Risk," Bank of England, Centre for Central Banking Studies.

BARUNÍK, J., AND F. ČECH (2016): "On the modelling and forecasting multivariate realized volatility: Generalized Heterogeneous Autoregressive (GHAR) model," *Journal of Forecasting*.

BAUER, G. H., AND K. VORKINK (2011): "Forecasting multivariate realized stock market volatility," *Journal of Econometrics*, 160, 93–101.

BAUWENS, L., C. HAFNER, AND S. LAURENT (2012): *Handbook of Volatility Models and Their Applications*. Wiley.

BECKER, R., A. CLEMENTS, AND R. O'NEILL (2010): "A Cholesky-MIDAS model for predicting stock portfolio volatility," Working paper, Centre for Growth and Business Cycle Research Discussion Paper Series.

BLACK, F. (1976): "Noise," *Journal of Finance*, 41, 529–543.

BLANCO, C., AND G. IHLE (1999): "How Good is Your VaR? Using Backtesting to Assess System Performance," *Financial Engineering News*, 11(8), 1–2.

BOLLEN, B. E., AND B. INDER (2002): "Estimating daily volatility in financial markets utilizing intraday data," *Journal of Empirical Finance*, 9, 551–562.

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

——— (1990): "Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model," *The Review of Economics and Statistics*, pp. 498–505.

——— (2009): "Glossary to ARCH (GARCH)," Working paper, Duke University.

BOLLERSLEV, T., R. F. ENGLE, AND D. B. NELSON (1994): "ARCH models," in *Handbook of Econometrics*. Elsevier Science, Amsterdam.

BOLLERSLEV, T., R. F. ENGLE, AND J. M. WOOLDRIDGE (1988): "A capital asset pricing model with time-varying covariances," *The Journal of Political Economy*, pp. 116–131.

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

BONATO, M. (2009): "Estimating the degrees of freedom of the Realized Volatility Wishart Autoregressive model," Working paper.

BONATO, M., M. CAPORIN, AND A. RANALDO (2009): "Forecasting realized (co)variances with a block structure Wishart autoregressive model," Working papers, Swiss National Bank.

CAI, J. (1994): "A Markov Model of Switching-Regime ARCH," *Journal of Business and Economic Statistics*, 12, 309–316.

CHAN, L. K., J. KARCESKI, AND J. LAKONISHOK (1999): "On Portfolio Optimization: Forecasting Covariances and Choosing the Risk Model," *The Review of Financial Studies*, 12, 937–974.

CHEN, L. (1996): "Stochastic Mean and Stochastic Volatility: A Four-Dimensional Term Structure of Interest Rates and Its Application to the Pricing of Derivative Securities," *Financial Markets, Institutions, and Instruments*, 5, 1–88.

CHRISTENSEN, B., AND N. PRABHALA (1998): "The relation between implied and realized volatility," *Journal of Financial Economics*, 37, 125–150.

CHRISTIANSEN, C., M. SCHMELING, AND A. SCHRIMPF (2012): "A comprehensive look at financial volatility prediction by economic variables," *Journal of Applied Econometrics*, 27, 956–977.

CHRISTOFFERSEN, P. F. (1998): "Evaluating interval forecasts," *International Economic Review*, 39, 841–862.

CLARK, T. E., AND K. D. WEST (2007): "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, 138(1), 291–311.

COLACITO, R., R. F. ENGLE, AND E. GHYSELS (2011): "A Component Model for Dynamic Correlations," *Journal of Econometrics*, 164, 45–59.

CONRAD, C., AND K. LOCH (2014): "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, 174–196.

DE POOTER, M., M. MARTENS, AND D. VAN DIJK (2008): "Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use?," *Econometric Reviews*, 27, 199–229.

DIEBOLD, F. X., AND R. S. MARIANO (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13, 253–263.

EBENS, H. (1999): "Realized Stock Index Volatility," Working PaperNo. 420, Department of Economics, Johns Hopkins University, Baltimore.

ENGLE, R. F. (1982): "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

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

ENGLE, R. F., AND G. M. GALLO (2006): "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131, 3–27.

ENGLE, R. F., E. GHYSELS, AND B. SOHN (2013): "Stock Market Volatility and Macroeconomic Fundamentals," *Review of Economics and Statistics*, 95, 776–797.

ENGLE, R. F., AND K. F. KRONER (1995): "Multivariate simultaneous generalized ARCH," *Econometric Theory*, 11, 122–150.

ENGLE, R. F., D. M. LILIEN, AND R. P. ROBINS (1987): "Estimating time varying risk premia in the term structure: The ARCH-M model," *Econometrica*, 55, 391–407.

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

ENGLE, R. F., AND J. G. RANGEL (2008): "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

ENGLE, R. F., AND K. SHEPPARD (2001): "Theoretical and empirical properties of dynamical conditional correlation model multivariate GARCH," UCSD Discussion.

EPPS, T. W. (1979): "Comovements in Stock Prices in the Very Short Run," *Journal of the American Statistical Association*, 74, 291–298.

FLEMING, J., C. KIRBY, AND B. OSTDIEK (2003): "The economic value of volatility timing using realized volatility," *Journal of Financial Economics*, 67, 473–509.

GHYSELS, E., A. HARVEY, AND E. RENAULT (1996): "Stochastic volatility," in *Statistical models in finance*, pp. 119–191. Amsterdam: North-Holland.

GHYSELS, E., A. RUBIA, AND R. VALKANOV (2009): "Multi-Period Forecasts of Volatility: Direct, Iterated, and Mixed-Data Approaches," Working paper.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2004): "The MIDAS touch: Mixed data sampling regression models," Working paper, UNC and UCLA.

——— (2006): "Predicting volatility: getting the most out of return data sampled at different frequencies," *Journal of Econometrics*, 131, 59–95.

GIACOMINI, R., AND H. WHITE (2006): "Tests of conditional predictive ability," *Econometrica*, 74, 1545–1578.

GIOT, P., AND S. LAURENT (2004): "Modelling daily Value-at-Risk using realized volatility and ARCH type models," *Journal of Empirical Finance*, 11, 379–398.

GLOSTEN, L. R., R. JAGANNATHAN, AND D. E. RUNKLE (1993): "On the Relation Between the Expected Value and the Volatility of the Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1801.

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

GOURIEROUX, C. S., J. JASIAK, AND R. SUFANA (2009): "The Wishart Autoregressive process of multivariate stochastic volatility," *Journal of Econometrics*, 150, 167–181.

HAGERUD, G. (1997): "A New Non-Linear GARCH Model," Ph.D. Thesis, Stockholm School of Economics.

HALBLEIB-CHIRIAC, R. (2007): "Nonstationary Wishart Autoregressive Model," Working Paper, CoFE, University of Konstanz.

HALBLEIB-CHIRIAC, R., AND V. VOEV (2011): "Modelling and Forecasting Multivariate Realized Volatility," *Journal of Applied Econometrics*, 26, 922–947.

HAMILTON, J. D., AND R. SUSMEL (1994): "Autoregressive conditional heteroskedasticity and changes in regime," *Journal of Econometrics*, 64, 307–333.

HANOCH, G., AND H. LEVY (1969): "The Efficiency Analysis of Choices Involving Risk," *The Review of Economic Studies*, 36, 335–346.

HANSEN, P. R., Z. HUANG, AND H. H. SHEK (2011): "Realized GARCH: A Joint Model for Returns and Realized Measures of volatility," *Journal of Applied Econometrics*, 27, 877–906.

HANSEN, P. R., AND A. LUNDE (2004): "An unbiased measure of realized variance," Unpublished manuscript, Stanford University.

——— (2005): "A realized variance for the whole day based on intermittent data," *Journal of Financial Econometrics*, 3, 525–554.

——— (2006): "Consistent ranking of volatility models," *Journal of Econometrics*, 131(1-2), 97–121.

HANSEN, P. R., A. LUNDE, AND J. M. NASON (2011): "The Model Confidence Set," *Econometrica*, 79, 435–497.

HANSEN, P. R., A. LUNDE, AND V. VOEV (2014): "Realized Beta GARCH: A Multivariate GARCH Model With Realized Measures of Volatility," *Journal of Applied Econometrics*, 29, 774–799.

HARRIS, L. (1990): "Estimation of stock variance and serial covariance from discrete observations," *Journal of Financial and Quantitative Analysis*, 25, 291–306.

——— (1991): "Stock price clustering and discreteness," *Review of Financial Studies*, 4, 389–415.

HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): "Multivariate stochastic variance models," *Review of Economic Studies*, 61, 247–264.

HAUTSCH, N., L. M. KYJ, AND P. MALEC (2015): "Do High-Frequency Data Improve High Dimensional Portfolio Allocations?," *Journal of Applied Econometrics*, 30(2), 263–290.

HESTON, S. L. (1993): "A closed form solution for options with stochastic volatility with applications to bond and currency options," *The Review of Financial Studies*, 6, 327–343.

HIGGINS, M. L., AND A. K. BERA (1992): "A Class of Nonlinear ARCH Models," *International Economic Review*, 33, 137–158.

JACOD, J., AND P. PROTTER (1998): "Asymptotic error distributions for the Euler method for stochastic differential equations," *Annals of Probability*, 26, 267–307.

JAGANNATHAN, R., AND T. MA (2003): "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance*, 58, 1651–1683.

JIN, X., AND J. M. MAHEU (2012): "Modelling Realized Covariances and Returns," *Journal of Financial Econometrics*, 11, 335–369.

KUESTER, K., S. MITTNIK, AND M. S. PAOLELLA (2006): "Value-at-risk prediction: a comparison of alternative strategies," *Journal of Financial Econometrics*, 4, 53–89.

KUPIEC, P. (1995): "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, 2, 73–84.

KYJ, L., B. OSTDIEK, AND K. ENSOR (2009): "Realized Covariance Estimation in Dynamic Portfolio Optimization," Working Paper.

LAURENT, S., J. V. ROMBOUTS, AND F. VIOLANTE (2013): "On Loss Functions and Ranking Forecasting Performances of Multivariate Volatility Models," *Journal of Econometrics*, 173(1), 1–10.

LAURENT, S., AND F. VIOLANTE (2012): "Volatility forecasts evaluation and comparison," *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1), 1–12.

LOPEZ, J. A. (1998): "Regulatory evaluation of Value-at-Risk models," *Federal Reserve Bank of New York Economic Policy Review*, 4(3), 119–124.

MAHEU, J. M., AND T. H. MCCURDY (2002): "Nonlinear features of realized volatility," *Review of Economics and Statistics*, 84, 668–681.

MARKOWITZ, H. (1952): "Portfolio selection," *Journal of Finance*, 7, 77–91.

MARTENS, M., M. DE POOTER, AND D. J. VAN DIJK (2004): "Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity," Tinbergen Institute Discussion Paper No. 04-067/4.

MCALEER, M., AND M. MEDEIROS (2008a): "A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries," *Journal of Econometrics*, 147, 104–119.

——— (2008b): "Realized Volatility: a Review," *Econometric Reviews*, 27, 10–45.

MERTON, R. C. (1980): "On estimating the expected return on the market: An explanatory investigation," *Journal of Financial Economics*, 8, 323–361.

MINCER, J. A., AND V. ZARNOWITZ (1969): "The Evaluation of Economic Forecasts," in *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, NBER Chapters, pp. 3–46. National Bureau of Economic Research, Inc.

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

NOURELDIN, D., N. SHEPHARD, AND K. SHEPPARD (2011): "Multivariate high-frequency-based volatility (HEAVY) models," *Journal of Applied Econometrics*, 27, 907–933.

OOMEN, R. C. (2001): "Using High Frequency Stock Market Index Data to Calculate, Model & Forecast Realized Return Variance," Working paper.

——— (2005a): "Properties of bias-corrected realized variance under alternative sampling schemes," *Journal of Financial Econometrics*, 3, 555–577.

——— (2005b): "Properties of realized variance under alternative sampling schemes," *Journal of Business and Economic Statistics*, 24, 219–237.

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

PATTON, A. J., AND K. SHEPPARD (2009): *Evaluating Volatility and Correlation Forecasts*Springer Berlin Heidelberg.

PAYE, B. S. (2012): "Dèjà vol: Predictive regressions for aggregate stock market volatility using macroeconomic variables," *Journal of Financial Economics*, 106, 527–546.

POON, S.-H., AND C. W. GRANGER (2003): "Forecasting Volatility in Financial Markets: A Review," *Journal of Economic Literature*, 41(2), 478–539.

POTERBA, J. M., AND L. H. SUMMERS (1986): "The persistence of volatility and stock market fluctuations," *The American Economic Review*, 76, 1142–1151.

PROTTER, P. E. (1992): *Stochastic Integration and Differential Equations: A New Approach*. New York: Springer-Verlag.

RICHARDSON, M., AND J. H. STOCK (1989): "Drawing inference from statistics based on multiyear asset returns," *Journal of Financial Economics*, 25, 323–348.

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

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

——— (1990): "Indexes of U.S. stock prices from 1802 to 1987," *Journal of Business*, 63, 399–426.

SHARPE, W. F. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–442.

SHEPHARD, N. (1996): "Statistical aspects of ARCH and stochastic volatility," in *Time Series Models in Econometrics, Finance and Other Fields*.

SHEPHARD, N., AND K. SHEPPARD (2010): "Realising the future: forecasting with high frequency based volatility (HEAVY) models," *Journal of Applied Econometrics*, 25, 197–231.

SHIRYAEV, A. N. (1999): *Essentials of Stochastic Finance: Facts, Models, Theory*. World Scientific.

SILVENNOINEN, A., AND T. TERÄSVIRTA (2008): "Multivariate GARCH Models," Working paper.

SOWELL, F. (1992): "Maximum likelihood estimation of fractionally integrated time series models," *Journal of Econometrics*, 53, 165–188.

TAYLOR, S. J. (1986): *Modeling Financial Time Series*. John Wiley and Sons.

——— (1994): "Modeling Stochastic Volatility: A Review and Comparative Study," *Mathematical Finance*, 4, 183–204.

TAYLOR, S. J., AND X. XU (1997): "The incremental volatility information in one million foreign exchange quotations," *Journal of Empirical Finance*, 4, 317–340.

THEIL, H., AND J. C. G. BOOT (1962): "The final form of econometric equation systems," *Review of International Statistical Institute*, 30, 136–152.

VAN DER WEIDE, R. (2002): "GO-GARCH: a multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17, 549–564.

VON NEUMANN, J., AND O. MORGENSTERN (1947): *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey, 2^a ed.

WEST, K. (1996): "Asymptotic inference about predictive ability," *Econometrica*, 64, 1067–1084.

WOLD, S. (1976): "Spline Functions in Data Analysis," *Technometrics*, 16(1), 1–11.

ZAKOIAN, J. M. (1994): "Threshold heteroskedastic," *Journal of Economics Dynamics and Control*, 18, 931–955.

ZELLNER, A. (1962): "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias," *Journal of the American Statistical Association*, 57(298), 348–368.

ZHANG, L. (2006): "Estimating Covariation: Epps Effect, Microstructure Noise," Working paper.

ZHANG, L., P. A. MYKLAND, AND Y. AÏT−SAHALIA (2005): "A tale of two time scales: determining integrated volatility with noisy high frequency data," *Journal of the American Statistical Association*, 100, 1394–1411.

ZHOU, B. (1996): "High frequency data and volatility in foreign-exchange rates," *Journal of Business and Economic Statistics*, 14, 45–52.

A

Parametrizations for Variance-Covariance Matrices

A general variance-covariance matrix, Σ, of dimension $n \times n$, defined as

$$\Sigma = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

is symmetric and semi-positive definite for each x, such that

$$x^T \Sigma x \geq 0.$$

This appendix relies only on a definite positive matrix, leaving aside the particular cases due to linear combinations of the underlying casual variables. Let θ be the set of parameters to determine Σ, the covariance matrix can be written as

$$\Sigma = L'L \tag{A.1}$$

where $L = L(\theta)$ is an $n \times n$ matrix of full rank obtained from a vector of unconstrained parameters of dimension $n(n+1)/2$. Any Σ defined as in (A.1) is positive definite.

Different choices of L lead to different parametrizations of Σ. We will consider here two classes of L: one based on the Cholesky factorization of Σ and another based on the spectral decomposition of Σ. The following variance-covariance will be used throughout this section to illustrate the use of the various parametrizations.

$$\Sigma = \begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} \tag{A.2}$$

A.1 Cholesky Parametrization

Since Σ is positive definite, it may be factored as $\Sigma = L'L$, where L is an upper triangular matrix.

Starting from a symmetric and positive definite matrix, Σ:

$$\Sigma = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

its Cholesky factorization is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{12} & l_{22} & 0 \\ l_{13} & l_{23} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ 0 & l_{22} & l_{23} \\ 0 & 0 & l_{33} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11}l_{12} & l_{11}l_{13} \\ l_{12}l_{11} & l_{12}^2 + l_{22}^2 & l_{12}l_{13} + l_{22}l_{13} \\ l_{13}l_{11} & l_{13}l_{12} + l_{23}l_{12} & l_{13}^2 + l_{23}^2 + l_{33}^2 \end{bmatrix}$$

L is obtained from the following formula:

$$l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} l_{jk}^2}$$

$$l_{ij} = \frac{1}{l_{jj}}\left(a_{ij} - \sum_{k=1}^{j-1} l_{ik}l_{jk}\right), \quad \text{per } i > j.$$

Knowing that $l_{11} = \sqrt{a_{11}}$, L can be generated in twofold way: in the first method, the matrix is calculated row by row starting from the top-left edge of the matrix, according to the Cholesky-Banachiewicz algorithm; the second method foresees to build the lower triangular column by column, according to the Cholesky-Crout algorithm.

For example, the Cholesky factorization of Σ, defined in (A.2), is

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{12} & l_{22} & 0 \\ l_{13} & l_{23} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ 0 & l_{22} & l_{23} \\ 0 & 0 & l_{33} \end{bmatrix}$$

The first row of L is given by:

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & l_{22} & 0 \\ -1 & l_{23} & l_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & l_{22} & l_{23} \\ 0 & 0 & l_{33} \end{bmatrix}$$

The second row of L is equal to

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} l_{22} & 0 \\ l_{23} & l_{33} \end{bmatrix} \begin{bmatrix} l_{22} & l_{23} \\ 0 & l_{33} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & l_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & l_{33} \end{bmatrix}$$

The third column is given by

$$10 - 1 = l_{33}^2, \quad l_{33} = 3$$

Then, the final decomposition is defined as follows

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

One problem with the Cholesky parametrization is that the Cholesky factor is not unique. In fact, if L is a Cholesky factor of Σ, then so is any matrix obtained by multiplying a subset of the rows of L by -1. This has implications on parameter identification, since up to 2^n different θ may represent the same Σ. Numerical problems can arise in the optimisation of an objective function when different optimal solutions are close together in the parameter space.

Another problem with Cholesky parametrization is the lack of a straightforward relationship between θ and the elements of Σ. This makes it hard to interpret the estimates of θ and to obtain confidence intervals for the variances and covariances in Σ based on confidence intervals for the elements of θ. One exception is $\mid L_{11} \mid = \sqrt{\mid \Sigma_{11} \mid}$, so confidence intervals on $\mid \Sigma_{11} \mid$ can be obtained from confidence intervals on $\left[L_{11}\right]$, where $\left[A\right]_{ij}$ denotes the ij-th element of the matrix Σ. By appropriately permuting the columns and rows of Σ, the confidence intervals can be derived for all the variance terms based on confidence intervals for the elements of L.

The main advantage of this parametrization, apart from the fact that it ensures positive definiteness of the estimate of Σ, is that it is computationally simple and stable.

A.2 Matrix Logarithm Parametrization

This parametrization is based on the spectral decomposition of the covariance matrix, Σ. Because Σ is positive definite, it has n positive eigenvalues λ. Let U denote the orthogonal matrix of orthonormal eigenvectors of Σ and $\Lambda = diag(\lambda)$, it can be written

$$\Sigma = U\Lambda U'.$$

By setting

$$L = \Lambda^{1/2}U'$$

in $\Sigma = L'L$, where $\Lambda^{1/2}$ represents the diagonal matrix with $\left[\Lambda^{1/2}\right]_{ii} = \sqrt{\left[\Lambda\right]_{ii}}$, a factorization of Σ can be derived from the spectral decomposition. The matrix logarithm of Σ is defined as $\log(\Sigma) = U\log(\Lambda)U'$, where $\log(\Lambda) = diag\left[\log(\Lambda)\right]$. Σ and $\log(\Sigma)$ share the same eigenvectors. The matrix $\log(\Sigma)$ can assume any value in the space of $n \times n$ symmetric matrices.

The matrix logarithm parametrization defines a one-to-one mapping of the θ elements of $\log(\Sigma)$, therefore Σ does not have the identification problem of the Cholesky decomposition. Similarly to the Cholesky decomposition, the vector θ in the matrix logarithm parametrization does not have a straightforward interpretation in terms of the original variance and covariances in Σ. In order to define $\log(\Sigma)$, the Schur-Fréchet algorithm may be used.

Let Σ be equal to (A.2). the matrix logarithm parametrization is given by

$$\log(\Sigma) = \begin{bmatrix} 2.8371 & 0.9245 & -0.3995 \\ 0.9245 & 2.4708 & 0.1956 \\ -0.3995 & 0.1956 & 2.3054 \end{bmatrix}$$