

This is a postprint version of the following published document:

Nieto, M. R. & Ruiz, E. (2016). Frontiers in VaR forecasting and backtesting. *International Journal of Forecasting*, 32(2), pp. 475–501.

DOI: [10.1016/j.ijforecast.2015.08.003](https://doi.org/10.1016/j.ijforecast.2015.08.003)

© 2015 International Institute of Forecasters. Published by Elsevier B.V.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

# Frontiers in VaR forecasting and backtesting

Maria Rosa Nieto\* and Esther Ruiz†

July 2015

## Abstract

The interest on forecasting Value at Risk (VaR) has been growing over the last two decades due to the practical relevance of this risk measure for financial and insurance institutions. Furthermore, VaR forecasts are often used as a battleground when alternative models are fitted to represent the dynamic evolution of time series of financial returns. There is a vast amount of alternative methods for constructing and evaluating VaR forecasts. In this paper, we survey the new benchmarks proposed in the recent literature.

**Keywords:** Backtesting, extreme value theory, GARCH, quantile, risk.

## 1 Introduction

"The advantage of knowing about risks is that we can change our behaviour to avoid them.", Engle (2003).

Value at risk (VaR) measures the potential loss in value of a risky portfolio over a defined period of time for a given probability. Forecasting VaR attracts a great deal of attention in the Financial Econometrics literature due to its relevance for financial and insurance institutions. Some adverse results along history have forced the agencies that regulate financial activity to look for a quantitative way to define the risk associated with a position in the market; see Granger (2002) for alternative definitions and measures of risk. The Basel accords explicitly recognize the role of VaR that financial institutions must implement and report in order to monitor their financial risk and to determine the amount of capital subject to regulatory control. Consequently, VaR is now established as the most popular risk measure designed for controlling and managing market risk. Although VaR has been mainly analyzed as a measure of risk associated with financial institutions, the recent Solvency 2 regulation also establishes it as the risk measure to be considered by insurance companies operating in the European Union; see Dowd and Blake (2006) and Sandström (2011) for a description of applications of VaR in the insurance

---

\*School of Business and Economics, Universidad de Anahuac México Norte.

†Corresponding author, Department of Statistics, Universidad Carlos III de Madrid, e-mail: ortega@est-econ.uc3m.es.

‡Financial support from Project ECO2012-32401 by the Spanish Government is gratefully acknowledged by the second author. We are also grateful to the Editor Rob Hyndman for his support and to three anonymous reviewers for their detailed and constructive comments.

sector. Also, the recent deregulation has also heightened the importance of risk management in electricity markets; see, for example, Chan and Gray (2006). All in all, forecasting VaR is crucial for many different sectors.

From a methodological point of view, VaR is a quantile of the density of returns and forecasting quantiles raises several issues of interest. Furthermore, forecasting VaR is also important as it is routinely implemented as an empirical check for alternative models to forecast conditional means and variances; see, for example, Asai and McAleer (2008), Martens *et al.* (2009), Wilhelmsson (2009), Grigoletto and Lisi (2009) and Brownlees and Gallo (2010) among many others.

In this paper, we survey recent methodological and empirical developments in VaR forecasting and testing, updating previous surveys in the literature; see Kuester *et al.* (2006), Christoffersen (2009), Gouriéroux and Jasiak (2010a) and Chen and Lu (2012) for previous surveys and Embrechts *et al.* (2000), Dowd (2007), McNeil *et al.* (2005), Jorion (2006), Danielsson (2011) and Christoffersen (2012) for comprehensive textbooks. Given that the number of recent contributions related with VaR forecasting and testing is extremely large, in an attempt to focus this survey, we only describe univariate models, leaving apart the interesting discussion on multivariate VaR forecasts. Furthermore, we only consider the VaR on the left tail of the distribution of returns as it concentrates most of the interest in the literature. Finally, note that although the Basel accords require daily forecasts of the VaR for returns over a holding period of 10 days, they allow these forecasts to be obtained from returns over shorter holding periods by using the square-root-of-time-rule. Consequently, we focus on daily one-step-ahead VaR forecasts corresponding to returns over a holding period of one day<sup>1</sup>. Moving to daily one-step-ahead forecasts of the VaR corresponding to returns over a holding period of 10 days, as required by the Basel accords, raises interesting forecasting issues.

The rest of the paper is organized as follows. Section 2 describes the VaR and establishes notation. Section 3 is devoted to the description of alternative procedures for point VaR forecasting while Section 4 deals with the construction of VaR forecast intervals. Section 5 describes backtesting procedures. Section 6 describes empirical implementations of VaR forecasting. Finally, Section 7 concludes the paper.

## 2 VaR as a risk measure

"VaR is defined as a *worst-case scenario on a typical day*", McAleer (2009).

The VaR is defined as the  $100\alpha\%$  quantile of the distribution of returns, such that, at time  $t$ , there is an  $100\alpha\%$  probability that the return of a portfolio over a one-day holding period,  $R_t$ , will fall below it. By regulatory convention, the VaR is positive and, consequently, it is given by

$$\text{VaR}_t^\alpha = -\sup [r \mid P [R_t \leq r] \leq \alpha]. \quad (1)$$

The probability in (1) is usually defined with respect to the distribution of returns conditional on the information available at time  $t - 1$ .<sup>2</sup>

---

<sup>1</sup>We focus on the horizon allowed by the Basel accords. However, in other sectors, the required VaR horizon could be different.

<sup>2</sup>There are some authors who define the VaR with respect to the marginal distribution of returns; see, for

The Basel accords describe an standard approach to obtain VaR forecasts which is known to render VaR estimates larger than necessary, leading to excessively high capital requirements (CR); see, for example, Pérignon *et al.* (2008), McAleer (2009) and Pérignon and Smith (2010b). Alternatively, the accords allow financial institutions to use internal models to forecast their VaR. From the perspective of financial institutions, using the standard approach is undesirable given that regulatory capital involves an opportunity cost. Hence, they have incentives to use their own VaR forecasts; see Pérignon *et al.* (2008) who present empirical evidence about the use of internal models being widespread among large financial institutions. Although internal VaR models are subjected to supervisory approval based on qualitative and quantitative standards, financial institutions enjoy a large degree of freedom in devising the precise nature of their models. This flexibility does not, however, imply that they are tempted to pursue the lowest possible VaR forecasts. The relation between the VaR and the CR is non-monotonic, as it takes into account not only the magnitude of the VaR but also the number of past VaR violations (i.e. actual losses exceeding the VaR). Specifically, the regulatory capital required to be held on day  $t + 1$  is determined as follows

$$CR_{t+1} = \max \{ VaR_t^\alpha, (3 + k) \overline{VaR}_t^\alpha \}, \quad (2)$$

where  $\overline{VaR}_t^\alpha = \frac{1}{60} \sum_{j=0}^{59} VaR_{t-j}^\alpha$ . The penalty  $k$  ranges between 0 and 1 and its value is determined by the number of VaR violations during the last 250 business days. The expression for CR in (2) suggests that lower capital charges could be achieved by lower VaR forecasts. This, however, need not be the case as lower VaR forecasts are possibly violated more often, thus increasing CR through the effect of the penalty factor  $k$ . Apart from direct costs due to the larger amount of capital that needs to be put aside, a large number of violations may also damage the institution's reputation; see McAleer and da Veiga (2008) for a discussion. Furthermore, underestimation of the own level of risk may lead to insufficient amount of CR to cover potential losses, thus increasing the risk of bankruptcy. Finally, note that the Basel accords establish that when ten or more VaR violations occur during a period of 250 business days, the financial institution may be forced to adopt the standard approach. On the other hand, the exaggeration of the own level of risk implies an excessive amount of CR, directly affecting the profitability of the bank. Another at least as undesirable consequence is that such banks appear more risky than they actually are, thus generating reputational concerns about their risk management systems. This affects the perception of investors and can induce underinvestment in VaR-overstating and VaR-understating banks. Indeed, Jorion (2002) shows that VaR disclosures are informative about the future variability in trading revenues, thus corroborating the idea that analysts/investors may be using VaR forecasts to support investment decisions. As a consequence, there are incentives for financial institutions to pursue VaR forecasts as accurate as possible.

The obvious benefit of VaR is that it is easily and intuitively understood by non-specialists; see Embrechts *et al.* (2000), Krause (2003) and Jorion (2006). However, the VaR has been criticized as a risk measure for not being coherent; see Acerbi and Tasche (2002) for the definition of coherent risk measures. In particular, the VaR is not subadditive given that the VaR of a diversified portfolio could be greater than the sum of the VaRs of the individual portfolios; see Krause (2003) for an interesting example. However, Danielsson *et al.* (2013) show that, for

---

example, Lien *et al.* (2014).

continuous random variables, either the VaR is coherent and satisfies subadditivity or the first moment of returns does not exist. Furthermore, they also explore the potential for violations of the VaR subadditivity and conclude that, for most practical applications, the VaR is subadditive. Consequently, for a very large class of distributions and practical situations, one does not have to worry about subadditivity violations of the VaR and there is no reason to choose alternative risk measures, solely for reasons of coherence.


A second limitation of VaR is that it does not measure losses exceeding itself. It is possible to have two distributions with the same VaR, but the losses that exceed VaR could be totally different. However, Danielsson *et al.* (2006) show that for heavy tailed distributions, as those observed in financial returns, the choice of downside risk measures does not seem to matter much as all of them order risk in a similar manner.

### 3 Forecasting VaR

"The statistical estimation of rare events belongs to the world of *extremistan*", Embrechts (2009).

This section describes and illustrates some of the most popular VaR forecasting procedures proposed in the literature. We survey first procedures based on directly forecasting the  $\alpha$  quantile of the distribution of returns (one-step procedures). Second, we describe two-step procedures based on estimating first the conditional mean and variance and then the conditional quantile<sup>3</sup>. All procedures are summarized in Table 1.

#### 3.1 Data and design of empirical study

The procedures described along this paper are illustrated by implementing them to forecast and backtest the 1% VaR of a series of S&P500 returns corresponding to a holding period of 1 day, observed daily from July 25, 2005 to May 19, 2014.<sup>4</sup> This observation period is split into an in-sample period with  $T$  observations reserved for estimation and an out-of-sample period with  $H$  observations reserved for backtesting. In order to choose  $T$ , Kuester *et al.* (2006) point out that estimation of the models usually used to forecast VaR should be based on large periods of time. However, Halbleib and Pohlmeier (2012) show that there is a trade-off between using large samples to estimate stable parameters and using recent samples to estimate parameters which easily adapt to the market conditions. On the other hand, when choosing  $H$ , Engle and Manganelli (2004), Escanciano and Olmo (2010) and Escanciano and Pei (2012) show that the backtesting procedures **to** work adequately when the ratio between  $H$  and  $T$  is small; Engle 

<sup>3</sup>Note that this classification is not the usual one in the literature according to which VaR forecast procedures are classified into parametric, semiparametric and non-parametric. However, we prefer to describe, first, those procedures that estimate directly the quantile of the distribution of returns without setting any particular specification for the conditional mean and variance of returns (one-step) and, second, those methods that estimate the conditional moments and then the quantile of standardized returns.

<sup>4</sup>In this survey, all procedures are illustrated for  $\alpha = 0.01$  as this is the quantile required by the Basel accords. However, the recent stress tests implemented to financial institutions, require VaR forecasts for levels smaller than 1%. For example, Chan and Gray (2006) and Chen *et al.* (2008) consider forecasting the 0.5% VaR while Danielsson *et al.* (2013) also consider 0.1%, 0.05% and 0.03% VaR forecasts.

and Manganelli (2004), Escanciano and Olmo (2010) and Escanciano and Pei (2012) consider  $H/T = 0.1$  and Boucher *et al.* (2014) choose  $H/T = 250/1040$ . Consequently, to illustrate the effects of  $T$  and  $H/T$  on the estimation and backtesting results, we consider two alternative splits of the observation period. First, the in-sample period goes up to May 27, 2010 with  $T = 1220$  and  $H = 1000$ . Second, the in-sample period covers observations up to May 21, 2013 with  $T = 1970$  and  $H = 250$ . All VaR forecast procedures considered are implemented using a rolling window scheme so that the size of the sample used for estimation is maintained fixed. Then, the VaR is forecast one-step-ahead during the out-of-sample period.

Figure 1 plots the S&P500 returns for the full observation period with vertical lines in the dates corresponding to the two splits considered. Figure 1 also plots the sample autocorrelations of returns and absolute returns and the cross-correlations between returns and absolute returns together with the corresponding 95% asymptotic confidence bounds for each of the subperiods considered. The bounds for the autocorrelations of returns are constructed using the heteroscedasticity correction proposed by Diebold (1988). Figure 1 shows that, as usual, returns oscillate around zero and show volatility clustering. The autocorrelations of absolute returns are positive and significant with the exception of those corresponding to the second out-of-sample period. Therefore, only in this out-of-sample period, corresponding to the last year of data, returns seem to be homoscedastic. Finally, the cross-correlations between returns and future absolute returns are always negative and significant suggesting asymmetric heteroscedasticity. However, the cross-correlations between returns and past absolute returns are not significant. Therefore, there is not evidence of volatility in the conditional mean of returns. Note that the evidence of conditional heteroscedasticity is stronger in both in-sample periods while the evidence of leverage is similar in the in-sample and out-of-sample periods.

## 3.2 One-step VaR forecasts

In this subsection, we describe the procedures based on directly forecasting the quantile of the distribution of returns, namely, Historical Simulation, CAViaR, nonparametric estimators of the return distribution and Extreme Value Theory.

### 3.2.1 Historical Simulation

Consider that  $R_t$ , the return at time  $t$ , has been observed for  $t = 1, \dots, T$ . The oldest procedure to forecast the VaR is based on Historical Simulation (HS) as follows

$$\widehat{VaR}_t^{HS} = -R_{(\omega)}, \quad (3)$$

where  $R_{(\omega)}$  is the  $\omega$ th-order statistic of returns with  $\omega = [T \times \alpha]$ . HS is by far the most popular procedure to forecast VaR used by commercial banks; see, for example, Pritsker (2006), Pérignon *et al.* (2008), Pérignon and Smith (2010b) and Berkowitz *et al.* (2011). For instance, Pérignon and Smith (2010b) document that almost three-quarters of banks that disclose their VaR method, report using HS. The popularity of HS is due to its simplicity and smoothness. However,  $\widehat{VaR}_t^{HS}$  can exhibit predictable jumps when large negative returns either enter into or drop out of the window used to obtain them. Furthermore, HS is based on assuming independent and identically distributed (iid) returns which is an inadequate assumption. Finally, note that empirical quantiles are not efficient estimators of extreme quantiles.

The HS estimator of VaR is illustrated implementing it to obtain one-step-ahead 1% VaR forecasts of the S&P500 returns over the two out-of-sample periods described in the previous subsection<sup>56</sup>. The top panel of Figure 2, that plots the corresponding forecasts, shows that  $\widehat{VaR}_t^{HS}$  is a stepwise function with a smooth evolution. Consider first the forecasts obtained using the first data split. Until August 2013 the samples used in the rolling window contain the observations corresponding to the years 2008 and 2009 and the VaR forecasts are highly influenced by these observations. Consequently, the  $\widehat{VaR}_t^{HS}$  has low variability and is very large in the first part of the out-of-sample period decreasing by the end, when the effect of the crisis of 2008 disappears. However, when the VaR is forecast using the second split of the data, the level estimated for the last year is nearly constant and very different from the VaR forecasts obtained before. Therefore, VaR forecasts obtained using HS are very sensitive to the sample used. The effect of high volatility periods may have long run effects. The top panel of Figure 2 also illustrates that too many observations in the moving window for HS results in sluggish adaptation to the dynamic changes in the true distribution of returns.

Recently, Zikovic and Aktan (2011) propose forecasting the VaR using Weighted Historical Simulation (WHS) by introducing Exponential Weighted Moving Average into the VaR forecast so that the most recent observations have larger weights with the smoothing parameter chosen by minimizing the Lopez (1999) size adjusted function, which is defined later in this survey. In an empirical application, they show that the optimal decay factor is rather constant over different time frames and always close to 0.99.

The corresponding one-step-ahead forecasts, denoted as  $\widehat{VaR}_t^{WHS}$ , are plotted in the bottom panel of Figure 2. First, observe that  $\widehat{VaR}_t^{WHS}$  does not depend on the particular split used in its estimation. The one-step-ahead forecasts obtained for the period May 22, 2013 to May 19, 2014 are the same regardless of whether they have been obtained using  $T = 1220$  observations, as in the first split, or  $T = 1970$  observations, as in the second split. Furthermore,  $\widehat{VaR}_t^{WHS}$  oscillate around a constant level which is clearly smaller than  $\widehat{VaR}_t^{HS}$ , implying lower risk. It is important to note that Zikovic and Aktan (2011) also find that  $\widehat{VaR}_t^{WHS}$  yields significantly lower forecasts than other alternative procedures. It could be worth to further investigate why this is the case. Finally,  $\widehat{VaR}_t^{WHS}$  have larger variability than  $\widehat{VaR}_t^{HS}$ .

### 3.2.2 CAViaR

An alternative VaR forecast procedure is based on representing directly the dynamic evolution of quantiles. The Conditional Autoregressive Value at Risk (CAViaR) was introduced by Engle and Manganelli (2004) as follows

$$\widehat{VaR}_t^{CAV} = \beta_0 + \beta_1 \widehat{VaR}_{t-1}^{CAV} + \beta_2 l(X_{t-1}), \quad (4)$$

where  $l(\cdot)$  is a function of a finite number of lagged observable variables,  $X_{t-1}$ , usually past returns. Three alternative specifications for  $l(\cdot)$  are proposed. First, they propose the adaptive

---

<sup>5</sup>Note that sample sizes of  $T = 1220$  or  $1970$  could seem to be too large to compute  $\widehat{VaR}_t^{HS}$ . Most authors use smaller sample sizes. However, some works use similar or even larger samples. For example, Angelidis *et al.* (2007) estimate  $\widehat{VaR}_t^{HS}$  with  $T = 1750$  observations while McNeil and Frey (2000) and Bao *et al.* (2006) consider sample sizes larger than 2000 observations.

<sup>6</sup>All estimates in this paper have been obtained using Matlab codes by the first author.

function, denoted by CAV(1), in which  $\beta_0 = 0$  and  $\beta_1 = 1$  and  $l(\cdot) = \left\{ \left[ 1 + \exp \left( G \left[ R_{t-1} + \widehat{VaR}_{t-1}^{CAV} \right] \right) \right] \right\}^{-1} - c$

where  $G$  is some positive finite number. The second specification proposed is the absolute value function, denoted by CAV(2), in which  $l(\cdot) = |R_{t-1}|$ . They also propose the asymmetric slope, denoted by CAV(3), in which  $l(\cdot) = \beta_3 (R_{t-1})^+ + \beta_4 (R_{t-1})^-$ , where  $(x)^+ = \max(x, 0)$  and  $(x)^- = -\min(x, 0)$ . Finally, they propose the Indirect GARCH specification which is given by

$$\widehat{VaR}_t^{IG} = \left( \beta_0 + \beta_1 \left( \widehat{VaR}_{t-1}^{IG} \right)^2 + \beta_2 R_{t-1}^2 \right)^{1/2}. \quad (5)$$

The Indirect GARCH in (5) is adequate when the volatility of returns is given by a GARCH(1,1) model and the standardized returns are iid.

The parameters of the CAViaR model can be estimated by the quantile regression approach developed by Koenker and Basset (1978); see also Komunjer (2005) who extends this approach to the tick-exponential Quasi Maximum Likelihood (QML) estimator.

Figure 3 plots the one-step-ahead 1% VaR forecasts for the two splits of the data considered, obtained using the adaptive (first row), absolute (second row), asymmetric (third row) and the Indirect GARCH (fourth row) functions. The adaptive and absolute forecasts are highly dependent on the estimation period used and rather different from the forecasts obtained when implementing the asymmetric slope function and the Indirect GARCH. Furthermore, these two latter forecasts are similar between them and do not depend on the estimation period in which they are based.

The CAViaR procedure has been extended in several directions. First, Gouriéroux and Jasiak (2008) propose a dynamic adaptive quantile procedure which ensures that the estimated quantiles do not cross at different VaR levels. They also introduce two alternative estimators of the parameters. First, an asymptotically efficient information-based estimation method which is obtained by maximizing the inverse KLIC measure and, second, an L-Moment estimator which is easier to compute but is not fully asymptotically efficient. A second extension of CAViaR has been proposed by De Rossi and Harvey (2009) who combine it with signal extraction, approximating some of the forms of the function  $l(\cdot)$  to the filtered estimators of time-varying quantiles. Third, several authors propose extending the Indirect GARCH in equation (5) by allowing for asymmetries. For example, Yu *et al.* (2010) and Gerlach *et al.* (2011) specify the equation of the VaR in (5) as a TGARCH model and propose EM-type and Monte Carlo Markov Chain (MCMC) estimators, respectively. The Monte Carlo results in Gerlach *et al.* (2011) suggest that there are not differences between forecasting the VaR using the MCMC procedure and the parameter estimator of Koenker and Bassett (1978). Furthermore, the asymmetric extension proposed does not seem to provide more accurate forecasts. More recently, Chen, Gerlach, Hwang and McAleer (2012) also propose a Threshold-CAViaR model for intra-day ranges estimated using a Bayesian approach via the link with the Skewed-Laplace distribution. Fourth, Huang *et al.* (2010) extend the CAViaR model by allowing the parameters in (4) to be a function of past returns. Finally, several authors propose introducing explanatory variables in the CAViaR equation. For example, Jeon and Taylor (2013) introduce implied volatility as a regressor while Rubia and Sanchis-Marco (2013) introduce liquidity and trading activity.

Taylor (2008a) propose the conditional autoregressive expectiles procedure based on using Asymmetric Least Squares, which is the least squares analogue of quantile regression. There is



a one-to-one mapping from expectiles to quantiles and, according to Taylor (2008a), estimating expectiles is more attractive from a computational point of view.

Analyzing whether all these extensions make a significant improvement in VaR forecasts should be considered in future research.

### 3.2.3 Nonparametric estimators of predictive distributions

The CAViaR procedure represents directly the evolution of quantiles, assuming a particular specification of this evolution. Alternatively, several authors propose nonparametric estimators of the returns distribution avoiding the effects of potential misspecification. These nonparametric methods are computationally more complicated but inferential gains can be obtained when the assumptions of the parametric models are wrong. Next, we describe some of the nonparametric procedures proposed to forecast VaR.

Chen and Tang (2005) propose forecasting the VaR by implementing kernel smoothing on the empirical distribution of returns in such a way that the estimator of the VaR is a weighted average of the order statistics around  $R_{(\omega)}$ . Taylor (2008b) proposes a related procedure based on forecasting quantiles using a double kernel smoothing estimator of the cumulative distribution function which provides greater accuracy for tail quantiles that are changing relatively quickly over time. This procedure is adapted from the double kernel estimator of Yu and Jones (1998) which is also considered by Cai and Wang (2008).

Alternatively, Geweke and Amisano (2011) propose a nonparametric model, which is a specific case of an artificial neural network model with two hidden layers, to obtain predictive distributions of daily returns over horizons of one to several trading days. Their model is estimated using MCMC methods. Also, Ferraty and Quintela-del-Río (in press) propose a nonparametric estimator when there are explanatory variables.

Finally, Xu (2013) proposes a fully nonparametric quantile regression model based on a double-smoothing local polynomial estimation of the conditional distribution function and the implementation of the empirical likelihood.

The nonparametric procedures are not illustrated in this survey due to their computational complexity.

### 3.2.4 Extreme Value Theory

The quantile of the distribution of returns can be estimated by implementing Extreme Value Theory (EVT) which models the tails of the distribution of returns without making any specific assumption concerning the center of the distribution; see Rocco (2014) for a detailed and very useful survey on EVT in finance. In applying EVT, one can adopt three different approaches, two of them parametric and the third non-parametric.

The first parametric approach is the *Block Maxima* (BM) that divides the sample into  $m$  subsamples of  $n$  observations each and picks the maximum of each subsample. When  $m$  and  $n$  tend to infinity, the limiting distribution of the adequately rescaled block maxima is one of three distributions which belong to the Generalized Extreme Value (GEV) distribution. The GEV distribution depends on a parameter,  $\xi$ , known as the shape parameter. The three distributions

just mentioned correspond to  $\xi > 0$  (Fréchet),  $\xi = 0$  (Gumbel) and  $\xi < 0$  (Weibull). In the context of fat tailed distributions, the quantity of interest is often the inverse of  $\xi$ , known as the tail index. The shape parameter can be estimated by Maximum Likelihood (ML) and the VaR is estimated based on the distribution of extremes; see, for example, Longin (2000). The BM approach has been implemented to forecast VaR by, for example, Diebold *et al.* (2000).

The second EVT parametric approach is the *Peak Over Threshold* (POT) method, according to which the observations exceeding a high threshold,  $u$ , are modelled separately from non-extreme observations. As  $u$  tends to infinity, the distribution of the exceedances, appropriately scaled, belongs to the Generalized Pareto Distribution (GPD) whose main parameter is  $\xi$ , the same shape parameter as that of the corresponding GEV distribution. In this case, the VaR forecast is given by

$$\widehat{VaR}_t^{GPD} = -R_{(k+1)} + \frac{\widehat{\sigma}}{\widehat{\xi}} \left( \left( \frac{\alpha}{k/T} \right)^{-\widehat{\xi}} - 1 \right), \quad (6)$$

where  $k$  is the number of observations over the threshold and  $\widehat{\xi}$  and  $\widehat{\sigma}$  are ML estimates of the shape and scale parameters of the GPD distribution, respectively.

Finally, it is also possible to nonparametrically estimate the shape parameter without assuming a particular model for the tail. Several estimators can accomplish this task being the most popular the Hill estimator which only works in the Fréchet case ( $\xi > 0$ ); see Hill (1975). The corresponding VaR forecast is given by

$$\widehat{VaR}_t^{HILL} = -R_{(k+1)} \left( \frac{\alpha}{k/T} \right)^{-\widehat{\xi}}, \quad (7)$$

where  $\widehat{\xi} = \frac{1}{k} \sum_{j=1}^k \left( \log(R_{(j)}) - \log(R_{(j+1)}) \right)$ ; see Peng and Qi (2006b) and Hill (2010) for the asymptotic properties of the tail index estimator for dependent, heterogeneous processes. In practice, the number of data points in the tails is limited leading to small sample biases. To address this problem, Huisman *et al.* (2001) propose a robust small sample bias-corrected estimator of  $\xi$  based on a linear regression of a set of Hill estimators obtained with different number of observations in the tail. Gomes and Pestana (2007), Gomes *et al.* (2007), Gomes *et al.* (2008) and Gomes *et al.* (2012) propose alternative estimators of  $\xi$  with minimum variance and reduced bias. Gouriéroux and Jasiak (2010a) point out that the accuracy of the Hill estimator and its extensions is rather poor due to the difficulty involved in estimating the probability of infrequent events. Another problem is that the Hill-type estimators depend on the number of observations in a very erratic way; see McNeil and Frey (2000) who show that the EVT method based on the GPD distribution gives more stable quantile estimates than the Hill estimator.

Any EVT approach entails choosing an adequate cut-off between the central part of the distribution and the tails. When working with threshold exceedances, the cut-off is induced by the number of observations in the tail,  $k$ , while in the BM procedure, it is implied by the choice of the number of blocks. The choice of the cut-off may have severe consequences on risk estimates. If it is too low, the VaR forecasts are biased and the asymptotic limit theorems do not apply. Conversely, if  $u$  is too large, the VaR forecasts have large standard deviations due to the limited number of observations over the threshold. Danielsson *et al.* (2001) and Ferreira *et al.* (2003) develop bootstrap methods for optimal threshold selection in the context of the Hill and GPD estimators, respectively. The former authors choose the threshold by minimizing the

asymptotic MSE of the Hill estimator. However, the selection of the threshold using bootstrap procedures is very time consuming. Alternatively, Gonzalo and Olmo (2004) propose a single step approach to threshold selection. Gencay and Selcuk (2004) determine  $u$  using a combination of the mean excess function and the Hill plots. Chavez-Demoulin *et al.* (2014) propose including a sensitivity analysis across several threshold values for a full POT application. Alternatively, Li *et al.* (2010) propose choosing  $u$  in such a way that the bias of the tail index estimator is reduced; see Scarrot and MacDonald (2012) for an excellent review of alternative procedures to choose or estimate the threshold.

Further to the problem of choosing an adequate threshold, Embrechts (2009) points out several caveats of EVT being specially important the fact that the rate of convergence in all EVT-based estimation procedures could be arbitrarily slow depending on the underlying distribution of returns. Furthermore, the asymptotic properties of EVT are based on the assumption of iid returns which is usually not satisfied in practice. Consequently, EVT needs to be modified complicating its implementation. In order to overcome this problem, several authors propose approaches based on self-exciting market point processes (SEMPP) which take into account the time between exceedances. There are two SEMPP models proposed in the literature: the Hawkes-POT model introduced by Chavez-Demoulin *et al.* (2005) and the ACD-POT model proposed by Herrera and Shipp (2013) who consider the time between exceedances as a stochastic process modeled as the ACD model of Engle and Russell (1998). Another related idea is to include the inter-exceedance times as covariates in the POT as proposed by Santos and Alves (2012a). Very recently, Chavez-Demoulin *et al.* (2014) propose a non-parametric extension to fit time-varying volatility in situations where the stationarity assumption can be violated by erratic changes in regime. They propose a Bayesian procedure allowing the intensity of the occurrence of exceedances to depend on past exceedances as well as on their size allowing the GPD distribution for the tails to vary over time.

### 3.3 Two-step VaR forecasts

The VaR forecast procedures described above model directly the quantile of the distribution of returns. Alternatively, the VaR can be forecast assuming particular specifications for the conditional distribution of returns. Consider the following model of returns

$$R_t = \mu_t + \varepsilon_t \sigma_t, \quad (8)$$

where  $\mu_t$  and  $\sigma_t$  are the conditional mean and standard deviation respectively, and  $\varepsilon_t$  is an iid sequence with variance 1. Thus, the one-step-ahead VaR conditional on information available at time  $t - 1$  is given by

$$VaR_t^\alpha = \mu_t + q_\varepsilon^\alpha \sigma_t, \quad (9)$$

where  $q_\varepsilon^\alpha$  is the  $\alpha$  quantile of the distribution of  $\varepsilon_t$ . Next, we describe alternative specifications proposed in the literature to represent  $\mu_t$ ,  $\sigma_t$  and  $q_\varepsilon^\alpha$ .

#### 3.3.1 Conditional mean

In order to forecast the VaR in equation (9) one needs to estimate the conditional mean,  $\mu_t$ . Given that the linear dependence of returns is usually very weak, most authors have represented

it by AR(1) or MA(1) models; see, for example, McNeil and Frey (2000), Kuester *et al.* (2006), Bali and Theodossiou (2007), Halbleib and Pohlmeier (2012) and Ardia and Hoogerheide (2014).

Consider again the series of S&P500 returns. The sample autocorrelations of returns plotted in Figure 1 are not significant in any subperiod. Given that the sample mean of returns is not significantly different from zero, the dependence in S&P500 returns seems to be adequately represented by a white noise in each of the two estimation subperiods and, consequently, we consider  $\mu_t = 0$ .

### 3.3.2 Conditional variance

When representing the conditional variance, it is very popular to choose models within the GARCH family and, in particular, the GARCH(1,1) model of Bollerslev (1986) given by

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \quad (10)$$

where  $\beta_0 > 0$ ,  $\beta_1 \geq 0$ ,  $\beta_2 \geq 0$ , and  $(\beta_1 + \beta_2) < 1$ ; see McNeil and Frey (2000), Engle (2003), Christoffersen and Gonçaves (2005), Giannopoulos and Tunaru (2005), Kuester *et al.* (2006) and Bali and Theodossiou (2007) among many others for VaR estimates obtained after fitting a GARCH(1,1) model to represent conditional variances. The popular RiskMetrics model is based on the Integrated GARCH (IGARCH) model which is given by equation (10) with  $\beta_0 = 0$  and  $\beta_1 + \beta_2 = 1$ . The Basel accords recommend to use the Gaussian IGARCH model with  $\beta_2 = 0.95$ ; see, for example, Gouriéroux and Jasiak (2010a).

The GARCH(1,1) model in (10) can be extended to cope with the asymmetric response of volatility to positive and negative returns. Among the many alternative asymmetric GARCH models, the Exponential GARCH (EGARCH) model of Nelson (1991) has often been shown as having an adequate performance when forecasting the VaR; see, for example, Bali and Theodossiou (2007). The volatility in an EGARCH (1,1) model is given by

$$\log(\sigma_t^2) = \beta_0 + \beta_1 [|\varepsilon_t| - \gamma \varepsilon_t] + \beta_2 \log(\sigma_{t-1}^2). \quad (11)$$

Some authors, as Giot and Laurent (2003) and Sajjad *et al.* (2008), consider modelling the volatility by fitting the APARCH model of Ding *et al.* (1993). However, the results obtained from the APARCH and EGARCH models are usually very similar; see Rodríguez and Ruiz (2012). Furthermore, Dark (2010) obtains estimates of the power parameter very close to one for a large range of stocks.

Consider again the S&P500 returns which, as mentioned above, are characterized by asymmetric conditional heteroscedasticity. Consequently, we fit the GARCH (1,1) and EGARCH(1,1) models to analyse the effects of misspecification of the conditional variance on VaR forecasts. The distributions of the errors considered when maximizing the quasi-log-likelihood are Gaussian, Student and Skewed-Student of Hansen (1994). Figure 4 plots rolling window estimates of the parameters of the EGARCH(1,1) model together with their 95% asymptotic confidence bounds for each of the two sample sizes considered when the Skewed-Student distribution is maximized<sup>7</sup>. Note that the estimates of  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  and the leverage parameter,  $\gamma$ , could be

---

<sup>7</sup>The results when the parameters are estimated by maximizing the Gaussian and Student log-likelihoods are similar and not reported to save space. Also, similar conclusions are obtained when estimating the GARCH parameters.

rather different when they are obtained using the first rolling window scheme, with  $T = 1220$ , and when using the second one, with  $T = 1970$ . However, the parameters of the Skewed-Student distribution are rather stable. These estimation results are in concordance with the comment by Gerlach *et al.* (2013): "The accuracy of parameter estimation seems particularly affected during the crisis period, where extreme outlying returns have big influence on parameter estimation and efficiency".

Chiu *et al.* (2005) compare the VaR forecasts obtained when an asymmetric GARCH model is fitted to represent the conditional variance with those obtained fitting the model proposed Maheu and McCurdy (2004) which includes jumps. In an empirical exercise, they conclude that omitting the presence of jumps leads to understate VaR forecasts and brings larger unexpected loss.

Some authors also suggest incorporating long-memory in the volatility when computing the VaR; see, for example, Degiannakis (2004), Caporin (2008) and Dark (2010). However, given that the VaR forecasts required by the Basel accords are short run, the inclusion of long-memory is expected not to make any fundamental difference; see, for example, So and Yu (2006) supporting this result. Finally, Sajjad *et al.* (2008) propose a Markov-switching APARCH model in which the volatility persistence can take different values depending on whether it is in a high or low volatility regime.

Although most of the literature has focused on GARCH-type models to represent the dynamic evolution of volatilities, several authors also consider stochastic volatility (SV) models. Some applications of SV models to the estimation of VaR can be found in Lehar *et al.* (2002), Fleming and Kirby (2003), González-Rivera *et al.* (2004) and Chen, Gerlach, Lin and Lee (2012). None of these papers find important differences on VaR forecasts by using GARCH or SV models. Furthermore, McAleer (2009) points out that the computational burden of SV can be quite severe while econometric software packages do not yet seem to have incorporated SV algorithms.

Bao and Ullah (2004) analyse the biases incurred when VaR forecasts obtained using parametric procedures with estimated parameters. Some authors propose corrections for these biases. When the GARCH model is fitted, Duan (2004) and Moraux (2011) propose a correction based on the Delta method assuming Gaussian returns. However, the Delta method is only appropriate when the parameter estimator is Normal; see Hall and Yao (2003). Lönnbark (2010) also proposes a bias correction that relies on the normality of the VaR and returns and, consequently, not very useful in empirical applications. Hartz *et al.* (2006) propose a bias-correction method for improving the VaR forecasting ability of the GARCH model with Normal errors. Recently, Gouriéroux and Zakoïan (2013) propose a correction of the  $1/T$  asymptotic bias in the coverage probabilities within the context of a parametric model with known error distribution that improves the accuracy of one-step-ahead VaR forecasts. Their procedure, which cannot be implemented when the conditional volatility follows a GARCH specification, also has an alternative bootstrap approach.

Regardless of the procedure implemented to forecast the VaR, Boucher *et al.* (2014) show that the specification bias can be very large, sometimes of the same order of magnitude of the VaR itself, and propose a general framework to take into account the specification and estimation uncertainty by empirically adjusting the VaR forecasts by outcomes of the backtesting procedures described in the next section. Using simulated data, they show that by dynamically adjusting

for estimation bias, the performance of every method is improved. The bias strongly depends upon the VaR level and, consequently, Boucher *et al.* (2014) do not recommend using procedures that use information on different quantiles to estimate the VaR as those used by, for example, Xiao and Koenker (2009), Wang *et al.* (2012) and Yi *et al.* (2014).

Finally, it is important to mention that, due to the availability of intradaily ultra-high frequency data (UHFD), there is a growing literature that proposes estimating the daily volatility using measures based on this type of observations; see Giot and Laurent (2004), Clements *et al.* (2008), Coroneo and Veredas (2012) and Fuertes and Olmo (2013). Brownlees and Gallo (2010) forecast daily VaR using realized volatility, bipower realized volatility, two-scales realized volatility, realized kernel volatility and the daily range. They conclude that VaR forecasts based on UHFD have improved properties with respect to those based on GARCH models with the simple daily range delivering very accurate forecasts. Furthermore, they show that the intra-daily sampling frequency plays a bigger role in fitting the tails of the distribution than the choice of the UHFD volatility measure with low frequencies performing better than high frequencies. Arroyo *et al.* (2011) base the VaR forecast on the histogram of intra-daily observations. More recently, Louzis *et al.* (2013) survey the literature on using UHFD for VaR forecasting and also compare several alternative estimators of the volatility; see also Huang and Lee (2013).

### 3.3.3 Conditional Quantile

The third component needed to compute the VaR in equation (9) is  $q_\varepsilon^\alpha$ , the  $\alpha\%$  quantile of the distribution of standardized returns. There are three main alternatives to obtain  $q_\varepsilon^\alpha$ . First,  $q_\varepsilon^\alpha$  is given when a particular distribution for  $\varepsilon_t$  is assumed. Second, one can implement the methods described in the previous subsection to compute the quantile of the returns standardized using the estimated conditional mean and variance. Third, the quantile of the distribution of  $\varepsilon_t$  can also be obtained by simulation using resampling methods.

**Assuming a given distribution for standardized returns** The most popular parametric distributions for standardized returns are the Gaussian, the Student and the Skewed-Student distribution of Hansen (1994); see, for example, Giot and Laurent (2003), Kuester *et al.* (2006), Bali and Theodosiou (2007), Sajjad *et al.* (2008), Pérignon and Smith (2010a), Halbleib and Pohlmeier (2012), Louzis *et al.* (2013) and Ardia and Hoogerheide (2014) for applications using these distributions.

Figures 5 and 6 plot the VaR forecasts obtained when the GARCH(1,1) and EGARCH(1,1) models, respectively, are fitted to the S&P500 returns and the errors are Normal (top panels), Student (middle panels) and Skewed-Student (bottom panels). First, note that neither the GARCH nor the EGARCH forecasts depend on the split used to estimate the parameters. Therefore, we can conclude that, regardless of the specific model fitted to the conditional variance and the error distribution, the differences in the estimated parameters, observed in Figure 4, do have remarkably small effects on VaR forecasts. This result is in concordance with González-Rivera *et al.* (2007) who analyse the implications of estimating the parameters of the RiskMetrics model under different loss functions and conclude that there are not important effects on the out-of-sample performance of VaR forecasts. Furthermore, Ardia and Hoogerheide (2014) recently show that the impact of the updating frequency (and, consequently, of the parameter estimates)

on the quality of VaR forecasts is remarkably small. It seems that the quality of the estimates of the parameters of the conditional variance does have very minor effects on VaR forecasts. Second, observe that the VaR forecasts generated when the GARCH model is fitted to estimate the conditional variance are smoother than those obtained when the EGARCH model is fitted. The GARCH forecasts are more similar to the CAViaR (absolute and asymmetric) and EVT (GPD and Hill) forecasts. Third, Figures 5 and 6 show that, regardless of the particular model for the conditional variance, the risk is larger when the Student distribution is assumed as the conditional distribution of returns. When this distribution is assumed to be the Skewed-Student distribution, the risk is forecast in between the risk under Normality and the risk of the Student errors. We observe larger differences between the VaR forecasts obtained under different assumptions of the error distribution during the August to December 2011 period. The differences between VaR forecasts are negligible during the last year of the out-of-sample period when the risk has been relatively small.

An alternative leptokurtic and asymmetric distribution considered in the context of VaR forecast is the Skewed-Generalized-t (SGT) distribution proposed by Theodosiou (1998); see Bali and Theodosiou (2007), Bali *et al.* (2008) and Chen and Hung (2011) for applications of the SGT distribution to VaR forecasting. The SGT distribution has the attractive of encompassing most of the distributions usually assumed for standardized returns as, for example, the Gaussian, Generalized Error Distribution (GED), Student and Skewed-Student distributions. However, in our experience, the maximization of the log-likelihood based on a SGT distribution is very complicated. Consequently, we will not consider further this distribution in this survey. Recently, Ergen (2015) also considers the Skewed-t distribution of Azzalini and Capitanio (2003).

The Laplace distribution, useful in defining quantiles and performing quantile regression, has also been considered to represent the distribution of the errors in the context of GARCH models. For example, Guermat and Harris (2001) propose it in the context of RiskMetrics while Chen and Gerlach (2013) propose a new distribution called asymmetric two-sided Weibull distribution that generalizes the asymmetric Laplace distribution previously proposed by Chen, Gerlach and Lu (2012) by allowing different Weibull distributions for positive and negative returns. Estimation and forecasting are carried out by a Bayesian MCMC procedure. However, in the empirical application, estimating the VaR of seven daily time series of financial returns, the results are comparable to those obtained when the asymmetric Student distribution is fitted. Taking into account the additional computational effort involved in using Bayesian procedures, it does not seem worth the additional effort.

Also, several authors propose modelling the distribution of  $\varepsilon_t$  by assuming mixtures of Normal or GED distributions; see, for example, Kuester *et al.* (2006) and Broda *et al.* (2013).

Finally, time-varying skewness and kurtosis have also been introduced in the distribution of  $\varepsilon_t$ . Bali *et al.* (2008) and Dark (2010) consider SGT and Skewed-Student distributions, respectively, with time-varying parameters which depend on past information. Wilhelmsson (2009) and Grigoletto and Lisi (2009) propose volatility models that allow for time-varying skewness and kurtosis. More recently, Grigoletto and Lisi (2011) model the distribution of  $\varepsilon_t$  by a Pearson Type IV which can be considered as a skewed version of the Student distribution in which the moments depend on parameters that evolve over time. Finally, in the context of RiskMetrics, Gerlach *et al.* (2013) propose an asymmetric Laplace distribution allowing for time-varying skewness and kurtosis and an asymmetric response of volatility to positive and

negative past returns. In the empirical application, they conclude that time-varying moments are not relevant to forecast VaR.

**Conditional Historical Simulation** Above, we describe HS as being one of the most popular VaR procedures among practitioners. However, HS has the important limitation of not taking into account the evolution of conditional variances that characterizes financial returns. To overcome this limitation, Hull and White (1998) propose implementing HS to returns standardized using the volatility estimated after fitting a GARCH-type model. Very recently, Dupuis *et al.* (in press) propose a robust estimator of the volatility based on weighted likelihood combined with HS choosing optimally the rolling window length and the smoothing parameter value.

Finally, Francq and Zakoïan (2015) propose a one-step QML estimator of the parameters of the conditional volatility and quantile, based on reparametrizing the model and derive its asymptotic distribution. When compared with the asymptotic distribution of the more popular two-step QML estimator, they show that none of the methods is superior in every situation. If the errors have a distribution admitting moments of any order, the two-step estimator may be superior. However, when the distribution of the errors does not have a finite fourth order moment, then the one-step estimator will be better.

The first panel of Figure 7 plots the VaR forecasts, denoted as Conditional Historical Simulation (CHS), obtained using HS after fitting the GARCH<sup>8</sup> model to estimate the conditional variance. First, we can observe that the VaR forecasts do not depend on the particular split used to obtain them. Second, it is possible to observe that the CHS one-step-ahead forecasts are very similar to those obtained when fitting the GARCH model assuming conditional Normality.

**Conditional EVT** Alternatively, instead of assuming a particular distribution of  $\varepsilon_t$ , Danielsson and de Vries (2000) and McNeil and Frey (2000) propose to estimate  $q_\varepsilon^\alpha$  in equation (9) implementing EVT to the standardized returns which are iid if the conditional mean and variance are correctly specified; see Chan and Gray (2006) for a nice description of conditional EVT (CEVT) and its application to forecast the VaR of daily electricity prices. In particular, McNeil and Frey (2000) propose filtering first the returns after estimating a GARCH model and then applying EVT to the tails of the innovations while bootstrapping the central part of the distribution. Jalal and Rockinger (2008) show that this procedure appears to perform a remarkable job in combination with a well chosen threshold estimation such as the one of Gonzalo and Olmo (2004). However, the results when implementing a GARCH model with conditionally Student errors are not so different.

Recently, Mancini and Trojani (2011) propose robustifying this procedure by estimating the GARCH parameters using the M-estimator of Mancini *et al.* (2005) and then fitting GPD to the tails of the residuals. The GPD is estimated using the robust estimator of Dupuis (1999) and Juarez and Schucani (2004) which, according to the authors, is less sensitive to the choice of the threshold than classical methods. Finally, resampling is carried out from this distribution.

---

<sup>8</sup>Note that, as in any of the conditional procedures described later, the same methodology could be applied to the returns standardized using any other of the alternative procedures available to estimate the conditional variance.



Finally, in the context of GARCH(1,1) and AR(1)-GARCH(1,1) models, Chan *et al.* (2012) and Chan *et al.* (2013) respectively, propose an estimator of the tail index based on the solution of a sample moment equation in which the unknown parameters are replaced by their QML estimates.

As an illustration, the first panel of Figure 8 plots the one-step-ahead 1% VaR forecasts obtained using CEVT with BM and blocks of  $n = 30$  observations. Note that the BM forecasts are larger than the GARCH forecasts plotted in Figure 5 and depend on the split of the sample used for their estimation. The middle and bottom panels of Figure 10 plot VaR forecasts obtained using CEVT with the GPD and Hill procedures, respectively, with  $k = 100$  as suggested by McNeil and Frey (2000). Note that the VaR forecasts obtained by both procedures are rather similar between them and similar to the forecasts obtained when fitting the GARCH model with Student errors. Furthermore, the GPD and Hill forecasts do not depend on the particular split used for their estimation. Finally, note that, once more, the main differences between the VaR forecasts obtained with the GPD and Hill procedures appear in the period between August and December 2011 when there is a clear increase in the risk of S&P500 returns.

To illustrate the role of the threshold on VaR forecasts, Figure 11 plots the one-step-ahead 1% VaR forecasts of the S&P500 returns obtained using EVT with the GPD (first column) and Hill (second column) estimators and the threshold estimated using the procedures proposed by Gonzalo and Olmo (2004) (first row) and Gencay and Selcuk (2004) (second row). We can observe differences between the VaR forecasts obtained in each case. First of all, note that, regardless of whether the VaR is computed using GPD or Hill, the risk is larger when the threshold is chosen implementing the procedure proposed by Gencay and Selcuk (2004). Furthermore, the forecasts of the VaR plotted in Figure 10 for  $k = 100$ , are somehow in between those obtained when implementing Gencay and Selcuk (2004) and Gonzalo and Olmo (2004). Also, it is important to point out that when implementing the former procedure, the VaR forecasts are the same regardless of the particular split chosen for the estimation period. However when implementing the procedure by Gonzalo and Olmo (2004), the VaR forecasts are different depending on the particular split used for the estimation.

**Filtered Historical Simulation** The quantile  $q_{\varepsilon}^{\alpha}$  can be estimated using bootstrap methods that do not assume any particular distribution of the errors; see Ruiz and Pascual (2002) for a review of the literature on using bootstrap procedures in financial time series and, in particular, on VaR forecasting. In particular, Barone-Adesi *et al.* (1999, 2002) propose a bootstrap method called Filtered Historical Simulation (FHS), based on using random draws with replacement from the standardized residuals which does not incorporate parameter uncertainty; see Engle (2003) and Pritsker (2006) for implementations. More recently, Pascual *et al.* (2006) propose a bootstrap procedure that allows incorporating the parameter uncertainty. As we will see in more detail below, bootstrap procedures have the advantage of allowing the construction of confidence intervals for the estimated VaR.

The second and third rows of Figure 7 plot the VaR forecasts obtained after implementing FHS without and with parameter uncertainty, respectively. First of all, observe that VaR forecasts do not depend on the particular split used. Second, the forecasts obtained with and without parameter uncertainty are different with the former being larger. Therefore, it seems that incorporating the parameter uncertainty may have an effect when using bootstrap to obtain the

conditional quantile of the distribution of conditional returns.

**Bayesian methods** Within the context of the VaR computed in a fully specified model, Hoogerheide and van Dijk (2010) propose estimating the model parameters and the density of returns using Bayesian methods based on importance sampling.

**Quantile regression** Xiao and Koenker (2009) propose a two-step approach of quantile regression estimation for linear GARCH models. In the first step, they use a quantile autoregression sieve approximation for the GARCH model by combining information over different quantiles. In the second stage, estimation of the GARCH model is carried out based on the first-step minimum distance estimation of the scale process.

### 3.3.4 Semiparametric methods

The two-step procedures considered in this subsection are based on estimating, first, the conditional mean and variance and, second, the quantile of the corresponding standardized returns. These procedures assume parametric specifications of the conditional mean and variance. However, semiparametric and nonparametric specifications of the conditional moments have also been considered in the literature. For example, Fan and Gu (2003) introduce a semiparametric model to estimate the volatility as a discretization of the IGARCH(1,1) model of RiskMetrics. In order to estimate the decay factor needed for the RiskMetrics methodology they propose two alternatives, one resulting in a data dependent decay factor which remains constant in the forecasting period, and the other which adapts automatically to changes in stock price dynamics, adding flexibility to the first decay factor. Additionally, Fan and Gu (2003) propose a symmetric nonparametric estimation approach to estimate the quantiles of the standardized residuals. Alternatively, Martins-Filho and Yao (2006), based on the two stage approach of McNeil and Frey (2000), propose a nonparametric estimation procedure for the conditional mean and variance using the local linear estimator of Fan (1992). Furthermore, they propose a method based on L-Moment theory instead of the GPD used by McNeil and Frey (2000).

Chen *et al.* (2008) propose estimating the local volatility adaptively and to model the standardized returns using the Generalized Hyperbolic distribution. According to these authors, this nonparametric adaptive methodology has the attractive of being able to estimate homogeneous volatility over a short time interval and reflects sudden changes in volatility.

Finally, Gouriéroux and Jasiak (2010b) also propose a semiparametric method based on a local approximation of the conditional density in a neighborhood of a predetermined extreme value. Given that the procedure requires a large number of observations around the quantile, they implement their procedure to compute the intradaily VaR.

## 3.4 Combining VaR forecasts

Recently, several authors propose forecast VaR combining several procedures. For example, McAleer *et al.* (2013a,b) consider mixing alternative risk models while McAleer *et al.* (2010) provide a procedure for choosing one risk model at the beginning of the forecast period that is modified depending on the past history of VaR violations. Alternatively, Jeon and Taylor

(2013) propose combining CAViaR forecasts with those obtained from a new estimator based on implied volatilities merging information from the historical time series of returns and the information supplied by the expectation of risk in the market. Finally, Halbleib and Pohlmeier (2012) consider several alternative VaR forecasts and point out that none of them is valid on its own. Consequently, also propose data-driven optimal combinations.

## 4 The uncertainty of VaR forecasts (VaR forecast intervals)

"Statistical point estimates should always be accompanied by confidence intervals as we are taught in Statistics 101", Embrechts (2009).

Boucher *et al.* (2014) argue that a key reason for the lack of accuracy of VaR forecasts in empirical applications is that they are subject to specification and estimation uncertainty; see also Jorion (2009) for the reasons of VaR forecasts failure during the global crisis. In spite of the importance of having measures of uncertainty associated with VaR forecasts, very few empirical papers report VaR forecast intervals. In this section, we review the literature on the uncertainty of VaR forecasts.

In the context of HS, Jorion (1996) and Ridder (1998) could be the first in pointing out the need to consider the estimation uncertainty associated with  $\widehat{VaR}_t^{HS}$  forecasts when the returns are assumed to be Gaussian; see also Dowd (2006) for the uncertainty of a related VaR measure based on order statistics and Gouriéroux and Jasiak (2010a) for the asymptotic distribution of  $\widehat{VaR}_t^{HS}$ . However, as far as we know, there are not analytical measures of the uncertainty associated with the WHS procedure. Furthermore, there are not available procedures to compute intervals for CAViaR forecasts.

There are also proposals to construct forecast intervals for VaR forecasts obtained using the Hill estimator. For example, Peng and Qi (2006a) carry out Monte Carlo experiments to compare the performance of three alternative intervals, namely, the Normal approximation, the likelihood ratio and the data tilting and conclude that the latter performs best. More recently, Chavez-Demoulin *et al.* (2014) provide forecast intervals for the Bayesian nonparametric POT procedure they propose. However, there are not proposals to construct forecast intervals for VaR's obtained using BM.

In the context of nonparametric procedures to compute directly the quantile of the distribution of returns, Chen and Tang (2005) emphasize the importance of the uncertainty of VaR forecasts and develop a procedure for its estimation based on a kernel estimation of the spectral density function of a series built using the smoother function. Xu (2013) also derives the asymptotic distribution of the nonparametric estimator. Forecast intervals for the VaR, automatically shaped by the data, can be constructed based on inversion of the empirical likelihood confidence intervals for the conditional cumulative distribution function.

In the context of two-step procedures, Taniai and Taniguchi (2008) take into account the parameter uncertainty using the asymptotic properties of the residual empirical process, when the VaR forecast are based on CHS after fitting the RiskMetrics model. More recently, Francq and Zakoïan (2015) obtain forecast intervals for VaR when the parameters are estimated either by the two-step or the one-step methods. If the errors belong to the class of regularly-varying

distributions with proportional tails, the asymptotic distribution of the extreme conditional quantiles is also available; see Chan *et al.* (2007). However, this distribution is unconditional. More recently, Gong *et al.* (2010) propose employing empirical likelihood to obtain conditional forecast intervals for the conditional quantile in the context of ARCH models. Also dealing with CHS, Spierdijk (in press) considers using bootstrap procedures based on extracting replicates with replacement from the assumed distribution of the standardized errors. She shows that, in spite of its inconsistency, accurate VaR forecast intervals are obtained for various symmetric distributions without fourth order moments. However, she shows that this bootstrap scheme fails when the error distribution is asymmetric and does not have finite fourth order moments. Alternatively, she proposes a residual subsample bootstrap as proposed by Sherman and Carlstein (2004) which seems to work in this latter case. The subsample bootstrap is based on extracting without replacement bootstrap samples of a size smaller than that of the original sample. The bootstrap samples are extracted from the empirical distribution of standardized residuals.

Chan *et al.* (2007) derive the asymptotic distribution of the CEVT quantile estimator of McNeil and Frey (2000) without assuming a specific parametric distribution of  $\varepsilon_t$ . They propose two alternative methods to construct VaR forecast intervals. The first method is based on the asymptotic Normality of the VaR estimator. Alternatively, they propose to construct confidence intervals by the tilting method of Hall and Yao (2003) and Peng and Qi (2006a). Note that the VaR forecast intervals constructed in this way do not incorporate the parameter uncertainty.

Several authors propose implementing bootstrap procedures to obtain forecast intervals for the VaR obtained using FHS; see Dowd (2007) and Christoffersen and Gonçalves (2005) who show that the forecast intervals for HS are too narrow and do not contain the true VaR with the desired frequency while the methods that properly account for the conditional variance dynamics imply forecast intervals with coverages close to the nominal. Bootstrap procedures are also implemented by Hartz *et al.* (2006). Also in the context of FHS and, based on the asymptotic behaviour of the GARCH residuals, Gao and Song (2008) derive the expression of the asymptotic variance of  $\widehat{VaR}_t^{FHS}$  and use it to construct forecast intervals based on the Normal approximation. Using simulated data, they show that the bootstrap method of Christoffersen and Gonçalves (2005) performs better than the asymptotic approximation they propose; see also Mancini and Trojani (2011) for an implementation of the procedure proposed by Christoffersen and Gonçalves (2005) to compute the uncertainty of VaR forecasts.

It is important to note that, when the VaR is obtained using the Bayesian procedure of Hoogerheide and van Dijk (2010), one can obtain their corresponding standard deviations. Finally, Li *et al.* (2011) also construct forecast intervals in the context of nonparametric estimation of the conditional mean and variance.

## 5 Backtesting point VaR forecasts

The validation of internal VaR forecasts is of paramount importance to guarantee that financial and insurance institutions have adequate capital to cope with large unexpected losses. In the light of the vast number of alternative VaR forecasting procedures available, model diagnosis is of essential importance for internal risk management. In order to assess the accuracy of VaR estimates, the Basel accords develop a statistical testing device known as backtesting.

According to their requirements, the backtesting should be based on at least 250 one-step-ahead VaR forecasts; see Campbell (2007) for a survey on backtesting procedures. In this section, we summarize the most popular backtesting procedures which are closely related with the literature on forecast interval evaluation. First, we briefly describe the procedures based on the binary hit variable. The lack of power of the most popular tests based on the hits sequence has also lead to propose testing for several VaR levels jointly. Finally, when several alternative estimators of the VaR are adequate according to the backtesting tests, one may additionally want to rank them. We also describe tests designed to compare and choose among alternative VaR forecast models.

Table 2 summarizes the tests described in this section.

## 5.1 VaR adequacy

### 5.1.1 Tests for a single VaR level

Many popular procedures for evaluating the performance of VaR forecasts are based on VaR failures. Consider the failure process given by  $I_t^\alpha = 1(R_t < -VaR_t^\alpha)$ ,  $t = T + 1, \dots, T + H$ , where  $1(\cdot)$  is the indicator function. A necessary condition for an optimal VaR forecast is that it is conditionally unbiased so that

$$E_{t-1}[I_t^\alpha] = \alpha. \quad (12)$$

Most traditional backtesting procedures are based on testing some of the implications of this condition. The most popular backtesting procedure, known as unconditional coverage (UC) test, was proposed by Kupiec (1995) and tests the null hypothesis  $H_0 : E[I_t^\alpha] = \alpha$ . However, this is not the hypothesis of interest in (12) and, consequently, the UC test ignores conditional coverage since violations can cluster over time. We should not be able to predict whether the VaR will be violated since if we could, then that information could be used to construct a better VaR forecast. Furthermore, Escanciano and Pei (2012) show that the unconditional coverage test is always inconsistent in detecting non-optimal VaR forecasts obtained using HS and FHS and propose an alternative data-driven weighted backtesting procedure with good power properties for these procedures. The lack of power of the UC test has also been documented by Pritsker (2006), de la Pena *et al.* (2007) and Pérignon and Smith (2008) among others. de la Pena *et al.* (2007) propose swapping the null and the alternative hypothesis so that the probability of choosing the wrong model is reduced.

Alternatively, Christoffersen (1998) proposes the very influential conditional coverage (CC) test, where the null hypothesis is  $H_0 : E[I_t^\alpha | I_{t-1}^\alpha] = \alpha$ . The likelihood ratio (LR) statistic for CC,  $LR_{cc}$ , is given by the addition of the unconditional statistic plus the independence statistic based on testing whether  $I_t^\alpha$  are iid  $Ber(\alpha)$  random variables against the alternative of first order Markov dependence. Note that the  $LR_{cc}$  test only takes into account the first order autocorrelation of the hit sequence. Furthermore, Escanciano and Olmo (2010) show that the use of standard unconditional and independence backtesting procedures can be misleading because they do not take into account the uncertainty associated with parameter estimation. They quantify this risk in a very general class of dynamic parametric VaR models and propose a correction of the standard backtests taking it into account. They show that one of the main determinants of the corrected asymptotic variance is the forecasting scheme used to generate the VaR forecasts, i.e. whether one uses recursive, rolling or fixed parameter estimates. Later,

Escanciano and Olmo (2011) analyze the effects of model misspecification on the UC and  $LR_{cc}$  tests and propose using a block-bootstrap procedure to implement robust backtests.

Table 3 reports the p-values of the  $LR_{cc}$  test implemented to the VaR forecasts obtained with each the procedures illustrated above and for the two different sample splits considered. Table 3 shows that, according to the  $LR_{cc}$  test, different models are rejected depending on the particular period considered. The procedures that are not rejected in any of the two periods are the CAViaR with the asymmetric function, the GARCH model with Skewed-Student errors, the EGARCH model with Student and Skewed-Student errors, the CHS, the CEVT with GPD and the bootstrap. On the other hand, HS, WHS, GARCH-t and FHS are rejected in both periods. For all other models, they are rejected or not depending on the particular period evaluated without any particular pattern.

It is important to note that, in spite of the many limitations described later in this survey, the Kupiec (1995) and Christoffersen (1998) tests are still the most popular ones both among practitioners and in the academic literature; see, for example, Chan and Gray (2006), Brownlees and Gallo (2010), Pérignon and Smith (2010b), Grigoletto and Lisi (2009, 2011), Zikovic and Aktan (2011), Chen, Gerlach, Hwang and McAleer (2012), Halbleib and Pohlmeier (2012), Rubia and Sanchis-Marco (2013), Dias (2013) and Liu and Tse (in press). Their popularity is due first to their simplicity and second, because they are implicitly incorporated in the Basel accords for determining market risk capital requirements.

Berkowitz *et al.* (2011) extend and unify the above tests showing that they can be interpreted as LM tests. They note that the centered hits sequence, defined as  $\delta_t^\alpha = I_t^\alpha - \alpha$ , form a martingale difference sequence and, consequently,

$$E[\delta_t^\alpha \otimes X_{t-1}] = 0, \quad (13)$$

where  $X_{t-1}$  includes any vector contained in the information set available at time  $t - 1$ . In particular, Berkowitz *et al.* (2011) propose the following Portmanteau test

$$LB(\tau) = (H)(H+2) \sum_{j=1}^{\tau} (H-j)^{-1} r_j^2, \quad (14)$$

where  $r_j$  is the order  $j$  sample autocorrelation of  $\delta_t^\alpha$ . Under the null,  $LB(\tau)$  is asymptotically  $\chi_{(\tau)}^2$ . For small sample sizes and for the 1% VaR, Berkowitz *et al.* (2011) show that the asymptotic critical values can be highly misleading. Alternatively, they rely on the Monte Carlo testing technique proposed by Dufour (2006).

Table 3 also reports the p-values of the  $LB(\tau)$  test in (14) for  $\tau = 5$  and 20 for the same procedures and periods considered above when using the asymptotic critical values. First of all, observe that the number of rejected procedures increases with respect to that obtained using the  $LR_{cc}$  test. This is clear in the first split considered. Therefore, it seems that considering the dependence between failures with more than one lag could be important to decide about the adequacy of VaR forecasts. In the very tranquil period (2), the results of the  $LR_{cc}$  and  $LB(\tau)$  tests are identical. Also note that results for the two values of  $\tau$  considered are very similar. Only the EGARCH models with Normal and Skewed errors and the bootstrap are not rejected in any of the out-of-sample periods considered.

The dynamic quantile (DQ) test, proposed by Engle and Manganelli (2004), can be obtained by testing the orthogonality restriction in (13) when  $X_{t-1}$  includes lags of  $\text{VaR}_t^\alpha$  and of  $\delta_t^\alpha$ .

Berkowitz *et al.* (2011) show that the DQ test appears to be the best backtest for the 1% VaR with all other tests having much lower powers. However, Herwartz (2009) analyzes the finite sample performance of the DQ test and shows that it is oversized in conventional sample sizes when using the asymptotic distribution.

The last column of Table 1 reports the p-values of the DQ test<sup>9</sup>. In period (1), the results are very similar to those of  $LR_{cc}$  while, in period (2), the number of models rejected is even smaller than when implementing the  $LR_{cc}$  test. Note that in the very tranquil period (2), only the WHS model is rejected. The models which are not rejected in any of the out-of-sample periods considered are the same as when implementing the  $LR_{cc}$  test plus CEVT with MB.

The fact that the DQ test is based on a linear regression with a binary dependent variable, has inspired some authors to propose VaR diagnostics based on logit regressions; see Clements and Taylor (2003) and Patton (2006) who propose an LR test. Herwartz and Waichman (2010) show how to use bootstrap and Monte Carlo procedures to approximate the finite sample distribution of the DQ and LR tests. Their Monte Carlo results show that when  $\alpha = 0.01$ , both tests have serious distortions if the asymptotic approximation is used to approximate the finite sample distribution even when  $H = 3000$ . However, using the DQ test and approximating its finite distribution by bootstrapping gets the best results. They also show that the power of both tests is very low regardless of how the finite sample distribution is approximated. Furthermore, in their Monte Carlo experiments, Gaglianone *et al.* (2011) show that the asymptotic critical values of the LM-type tests are only accurate if the sample size corresponds at least to four years of daily observations, i.e.,  $H \geq 1000$ .

The backtesting methodologies described above only focus on the number of VaR exceptions and totally disregard their magnitude. Wong (2010) criticizes the statistics proposed by Christoffersen (1998) because they are two-tailed and, consequently, they can reject a risk model for overconservative. However, note that, as mentioned above, a risk model can also be rejected for being overconservative as this is not desirable for financial institutions. Alternatively, Wong (2010) proposes the tail risk statistic defined as follows

$$TR = -\frac{1}{T} \sum_{i=1}^T (R_t - \alpha) I(R_t < \alpha). \quad (15)$$

The  $TR$  statistic tells a risk manager the size of aggregate tail losses a portfolio may incur over the period considered. The asymptotic distribution of the  $TR$  statistic is derived under the assumption of Normal returns. Consequently, Wong (2010) proposes using Berkowitz and O'Brien (2002) and Kerkhof and Melenberg (2004) and transforming the non-Normal density into a Normal one and applying the saddlepoint analysis to the transformed variable. However, this transformation complicates the empirical application of the  $TR$  statistic. Alternatively, Colletaz *et al.* (2013) propose the Risk Map as a new method for validating risk models which jointly accounts for the number and magnitude of extreme losses and graphically summarizes all information about the performance of a risk model. The Risk Map boils down to the intuition that a large loss not only exceeds the regular VaR with level  $\alpha$  but is also likely to exceed a VaR defined with a much lower probability. In order to test a risk model, Colletaz *et al.*

---

<sup>9</sup>The computation of the DQ test has been carried out implementing the Matlab codes of Simone Manganelli available at <http://www.simonemandanelli.org/Simone/Research.html>

(2013) propose testing whether the sequences of exceptions (with respect to the 1% VaR) and superexceptions (with respect to the lower quantile) satisfy the conditions of the unconditional coverage test of Kupiec (1995). The corresponding p-value is reported in a three-dimensional Risk Map.

Alternatively, instead of testing implications based on hits, several authors consider tests based on the duration sequence, i.e. the number of observations between two consecutive failures; see Christoffersen and Pelletier (2004) and Haas (2005). Under a correct VaR model, durations should have a geometric distribution with an average equal to the reciprocal of the coverage probability; see Candelon *et al.* (2011) for further developments of duration tests in the context of GMM. Santos and Alves (2012b) also propose an independence test based on durations. Very recently, Ziggel *et al.* (2014) propose a weighted test of the unconditional coverage and iid assumptions extending the duration test of Christoffersen and Pelletier (2004). The weighted test allows the user to choose the weight with which the test of unconditional coverage enters the joint test of conditional coverage. It is based on iid Bernoulli random variables and, according to the authors, it is very intuitive and easy to implement.

The tests described above are based on the LM principle and have low finite sample power against a variety of model misspecifications; see Christoffersen *et al.* (2001) and Gouriéroux and Jasiak (2010a) who point out that the count of exceedances may be misleading as an instrument of VaR control. The problem arises because the binary variables,  $\delta_t^\alpha$ , are constructed to represent rare events and, in finite samples, there could be few extreme observations leading to a lack of information needed to reject a misspecified model. LM tests sacrifice too much information implying a reduce power against misspecified VaR forecast models. Alternatively, tests based on the quantile regression methodology, which can be interpreted as Wald tests, have been proposed. For example, Gaglianone *et al.* (2011) propose a backtest that relies on the quantile regression of Koenker and Xiao (2002). They propose the following random coefficient model that can be used to construct a Wald test for the null hypothesis that a given VaR model is correctly specified

$$R_t = \beta_0(U_t) + \beta_1(U_t)VaR_t, \quad (16)$$

where  $U_t$  are iid  $U(0,1)$  variables and  $\beta_i(U_t)$ ,  $i = 0, 1$  are assumed to be comonotonic in  $U_t$ . The null hypothesis of adequate coverage can be written as a Mincer and Zarnowitz (1969) test as follows

$$H_0 : \begin{cases} \beta_0(\alpha) = 0 \\ \beta_1(\alpha) = 1. \end{cases} \quad (17)$$

The corresponding test, denoted as VaR Quantile Regression (VQR), uses more information to reject a misspecified model, which makes it to deliver more power in finite samples than the LM tests described above. The asymptotic distribution of the VQR test is based on the strict stationarity of  $R_t$  and  $VaR_t$ . The Monte Carlo experiments reported by Gaglianone *et al.* (2011) show that, when  $\alpha = 0.01$ , the empirical size of the VQR test is much larger than the nominal in small samples. However, when the sample size is as large as  $T = 2500$ , all tests have similar sizes close to the nominal. With respect to power, the DQ and VQR tests have similar powers in small samples with the second being preferred if  $T \geq 1000$ . Furthermore, it is important to note that extensions of the methodology proposed by Gaglianone *et al.* (2011) to multivariate VaR forecast models and/or to ten-step-ahead VaRs are not straightforward.

Finally, Hoogerheide *et al.* (2012) propose an extension of the forecast rationality and op-



timally tests proposed by Patton and Timmermann (2012) which involve no observations of the target variable. They show that the power of the new test is larger than that of the traditional unconditional and conditional coverage tests in the context of simple AR and ARCH models.

### 5.1.2 Tests for multiple VaR levels

Several authors show that the conclusions about which model is more adequate to forecast the VaR depend on the particular quantile of the distribution of returns being forecast; see, for example, Giacomini and Komunjer (2005), Kuester *et al.* (2006), Gerlach *et al.* (2011), Gaglianone *et al.* (2011) and Chen, Gerlach, Lin and Lee (2012). Therefore, when looking at the adequacy of VaR forecasts, it seems important to look for tests in which several quantiles are tested jointly.

Berkowitz (2001) proposes testing for the adequacy of the whole density. A weakness of this test is that models with a superior density forecast may not necessarily meet the requirements of risk managers that are mainly concerned about forecasting the tails. Consequently, Hurlin and Topkavi (2007) propose a new VaR validation procedure that is half way between focusing on one single VaR level and using the whole forecast density by looking at violations over multiple coverages. The proposed test extends the LM test of Berkowitz *et al.* (2011) in (14) to the multivariate case by considering the vector of centered hits which is defined as  $\delta_t = (\delta_t^{\alpha_1}, \dots, \delta_t^{\alpha_m})'$ . Joint conditional accuracy of VaR forecasts implies that  $Cov(\delta_t, \delta_{t-j}) = 0$  for  $j \neq 0$ . Consequently, Hurlin and Topkavi (2007) propose the following statistic

$$Q_m(J) = T \sum_{i=1}^J \left( \text{vec} \hat{R}_j \right)' \left( \hat{R}_0^{-1} \otimes \hat{R}_0^{-1} \right) \left( \text{vec} \hat{R}_j \right), \quad (18)$$

where  $\hat{R}_j = D^{1/2} \hat{C}_j D^{1/2}$  with  $D = I_m \odot \hat{C}_0$ , being  $\odot$  the element to element multiplication and  $\hat{C}_j = \frac{1}{H} \sum_{t=1}^H \delta_t \delta_{t-j}'$ ; see Sajjad *et al.* (2008) for an empirical application of this test. Herwartz (2009) shows that the test in (18) is oversized when using the asymptotic distribution. By using Monte Carlo critical values, the test in (18) turns out to be more powerful than the DQ test. However, a false specification of the tail features of the distribution of the errors is likely to come along with a failure of the VaR model which Herwartz (2009) shows that remains undetected when implementing the statistic in (18). Furthermore, Leccadito *et al.* (2014) point out that if the coverage rates considered in the test are very close to each other, then the matrix  $\hat{R}_0$  is singular and, consequently, the  $Q_m(J)$  test cannot be calculated. Gouriéroux and Jasiak (2008) also consider multilevel tests.

Alternatively, Perignon and Smith (2008) propose a multilevel test that extends the unconditional test of Kupiec (1995). However, the authors only present a graphical analysis rather than a formal testing procedure. Alternatively, Escanciano and Velasco (2010) propose omnibus specification tests extending the unconditional, conditional and independence tests of Kupiec (1995) and Christoffersen (1998) to test a possible continuum of quantiles. The asymptotic critical values are approximated by subsampling which accounts for parameter estimation uncertainty. More recently, Leccadito *et al.* (2014) propose two alternative multilevel tests. The first test is a multilevel generalization of the Christoffersen (1998) test while the second is a Pearson-type test based on the bivariate distribution of the total number of violations for all levels considered and its lag. They show that the proposed tests are superior with respect to

those of Hurlin and Topkavi (2007) and Pérignon and Smith (2008). However, there is not comparison with the tests proposed by Escanciano and Velasco (2010).

## 5.2 Comparing alternative VaR forecasts

The backtesting procedures described above are designed to test whether a particular procedure gives accurate VaR forecasts. However, when several accurate estimators are available, one often wants also to decide which estimator is best among them. With this goal, Lopez (1999) proposes to choose the procedure that minimizes  $C(m) = \sum_{t=T+1}^{T+H} C_t^{(m)}$  where

$$C_t^{(m)} = \begin{cases} f\left(R_t, VaR_t^{(m)}\right) & \text{if } R_t < VaR_t^{(m)}, \\ g\left(R_t, VaR_t^{(m)}\right) & \text{if } R_t \geq VaR_t^{(m)}, \end{cases}$$

where the index  $m$  stands for VaR procedure  $m$  and  $f(x, y)$  and  $g(x, y)$  are functions such that  $f(x, y) \geq g(x, y)$ ; see Lopez (1999) and Sener *et al.* (2012) for a description of alternative functions. Sarma *et al.* (2003) and Angelidis and Degiannakis (2007) use the Regulatory Loss function (RLF) in which  $f(R_t, VaR_t^{(m)}) = \left(R_t - VaR_t^{(m)}\right)^2$  and  $g\left(R_t, VaR_t^{(m)}\right) = 0$ . More recently, Sener *et al.* (2012) propose a loss function that penalizes for the magnitudes of the errors, the autocorrelations between the errors and excessive capital allocations.

It is important to note that, if quantile forecasts are to be assessed, Gneiting (2011) propose using a consistent scoring Generalized Piecewise Linear (GPL) family of functions. A function is consistent for the  $\alpha$  quantile if and only if it is of the following form

$$S(x, y) = \begin{cases} (I(x \geq y) - \alpha) \frac{1}{|b|} (x^b - y^b) & \text{if } b \in \mathbb{R} / \{0\} \\ (I(x \geq y) - \alpha) & \text{if } b = 0. \end{cases}$$

Gneiting (2011) points out that the loss functions proposed by Lopez (1999) and extended by Caporin (2008) are not of the GPL form and, consequently, this could explain why these functions are not able to properly distinguish between the true Data Generating Process (DGP) and alternative models to forecast the VaR; see Caporin (2008).

Table 4 reports the values of  $C(m)$  for RLF for those models that are not rejected by at least two backtests in each of the out-of-sample periods considered. In the first out-of-sample period, the minimum  $C(m)$  is obtained by the EGARCH model with Student errors. However, when looking at the values of  $C(m)$  in the second very tranquil out-of-sample period, all the models have similar results and therefore, it is not obvious whether one can identify a "best" model.

On the other hand, Giacomini and Komunjer (2005) and Bao *et al.* (2006) compare competing VaR forecasts using  $C(m)$  with  $C_t^{(m)}$  defined with the following predictive quantile loss function (PQLF)

$$C_t^{(m)} = [\alpha - 1 (R_t < VaR_t^\alpha)] [R_t - VaR_t^\alpha]. \quad (19)$$

Table 4 reports the corresponding values of  $C(m)$ . It is remarkable that, regardless of the out-of-sample period, the values of  $C(m)$  are very similar among all models and periods considered. For example, only the CAViaR with the Absolute Value function and the EVT with BM seems to be slightly higher in periods (1) and (2).

Giacomini and Komunjer (2005) also propose to compare two VaR forecasts versus their combination using a conditional quantile forecast encompassing test of superior predictive ability.

A rejection of the test provides statistical evidence that the combination outperforms the two individual forecasts.

Alternatively, when trying to establish the superiority between two models, Sarma *et al.* (2003) propose testing  $H_0 : \{\theta = 0\}$  against  $H_1 : \{\theta < 0\}$ , where  $\theta$  is the median of the distribution of the loss differential between procedure  $i$  and procedure  $j$ ,  $z_t = C_t^{(i)} - C_t^{(j)}$ . Under the null hypothesis, the exact distribution of  $S_{ij} = \sum_{t=T+1}^{T+H} 1(z_t \geq 0)$  is binomial with parameters  $(H, 0.5)$  while the asymptotic distribution is given by

$$\frac{S_{ij} - 0.5H}{\sqrt{0.25H}} \stackrel{a}{\sim} N(0, 1); \quad (20)$$

see Diebold and Mariano (1995). If  $H_0$  is rejected, model  $i$  is significantly better than model  $j$  for the chosen loss function. Note that the statistic in (20) can be obtained as the t-statistic of the regression of  $z_t$  on a constant using the Newey and West (1987) heteroscedasticity autocorrelation consistent standard errors.

An alternative to the test in (20) is the test of conditional predictive ability, proposed by Giacomini and White (2006), which takes into account the estimation uncertainty due to model selection. The one-step-ahead conditional predictive ability (CPA) statistic is given by

$$CPA = H \left( H^{-1} \sum_{t=T}^{T+H-1} h_t \Delta C_{t+1}^{(m)} \right)' \widehat{\Omega}_H^{-1} \left( H^{-1} \sum_{t=T}^{T+H-1} h_t \Delta C_{t+1}^{(m)} \right), \quad (21)$$

where  $\widehat{\Omega}_H$  is a consistent estimate of the variance of  $h_t \Delta C_{t+1}^{(m)}$  and  $h_t$  is a  $q \times 1$  vector given by  $h_t = (1, \Delta C_t^{(m)})'$ . Under the null hypothesis of equal conditional predictive ability, CPA is asymptotically distributed as a  $\chi_q^2$ .

On the other hand, Bao *et al.* (2006) and Angelidis and Degiannakis (2007) propose to compare alternative VaR forecasts using the test of superior predictive ability (SPA) of Hansen (2005). The null hypothesis, that the benchmark model ( $m = 0$ ) is not inferior than the alternatives, is tested with the following statistic

$$SPA = \max \left[ \max_{m=1, \dots, M} \frac{H^{1/2} \bar{z}_m}{\widehat{\omega}_m}, 0 \right], \quad (22)$$

where  $\widehat{\omega}_m^2$  is a consistent estimator of  $\omega_m^2 = \text{var}(H^{1/2} \bar{z}_m)$  and  $\bar{z}_m = H^{-1} \sum_{t=1}^H (C_t^{(0)} - C_t^{(m)})$ . The estimation of  $\omega_m^2$  and the p-values of SPA can be obtained using the stationary bootstrap of Politis and Romano (1994) with the optimal block-size chosen by the block selection algorithm proposed by Politis and White (2004, 2009). Recently, Sener *et al.* (2012) propose a predictive ability test that does not require a benchmark model allowing for the simultaneous comparison of several procedures.

Table 4 reports the p-values of the SPA test when, for each out-of-sample period, each model is selected as benchmark against all the others. We can observe that, in period (1), when the RLF function is used, none of the models is rejected as inferior against the others if the nominal size of the test is 1%. On the other hand, the results are completely different when the PQLF is implemented. In this case, several models are rejected depending on the particular period considered. Consequently, we can conclude that the ranking of the models is highly dependent on the period considered and the function used for it.

### 5.3 A final remark on backtesting

The literature on forecasting and backtesting VaR largely assumes that the appropriate data are used. However, Frésard *et al.* (2011), using information from annual reports of the largest 200 US and international commercial banks, document that a large fraction of them artificially boost the performance of their models by polluting their returns with extraneous profits such as intraday revenues, fees, commissions, net interest income and revenues from market making or underwriting activities. They find that over the period 2005-2008, less than 6% of the largest commercial banks in the world evaluate their VaR models using the appropriate uncontaminated data. They also show that all available backtesting procedures are highly sensitive to data contamination. For example, using the "traffic light" approach developed by the Basel Committee, 23.5% of the VaR models are rejected when tested with uncontaminated data whereas 10.8% are rejected when tested with returns that include both fees and intraday trading revenues. Therefore, data contamination has dramatic implications for model validation and can lead to the acceptance of misspecified VaR models and significantly reduced regulatory capital.

## 6 VaR forecasting in practice

"Overall, this crisis has reinforced the importance of risk management", Jorion (2009).

In spite of its limitations, the main advantages of VaR are simplicity, wide applicability and universality. Although a myriad of procedures are currently available for forecasting and testing VaR, no consensus has been reached on which procedures are best. As the illustration above shows, for a given returns series, the ranking of VaR forecast procedures depend on the period of time considered, the procedures compared and the measures used to compare them. Dias (2013) also points out the importance of market capitalization and the particular period when ranking the performance of different procedures. She concludes that VaR is always wrongly estimated if we do not take crises and non-crises periods into account and that market capitalization has a positive effect on the estimation of VaR. There are many papers comparing the performance of alternative VaR forecasts with very mixed results; see, for example, Angelidis *et al.* (2007). In this section, we focus on those that consider the performance of VaR forecasts during the recent financial crisis.

In practice, banks appear to be wary of being overly optimistic about their level of market risk during tranquil periods. In fact, empirical evidence presented by Berkowitz and O'Brien (2002), Pérignon *et al.* (2008) and Pérignon and Smith (2010b) suggests that they systematically overestimate their VaR; see also Pérignon and Smith (2010a) for an analysis of the potential causes of this overestimation. The opposite situation seems to occur in times of stressed market conditions. During the 2007/2008 financial crisis, banks systematically underestimated their VaR; see, for example, Chen, Gerlach, Lin and Lee (2012). This alternation of over- and underestimation of market risk levels may, at least to some extent, be due to the fact that VaR measures typically are calibrated using historical data. Following a period of calm financial markets, the VaR estimates and the accompanying CR can decline to low levels, but then they might underestimate risk during a period of stress that lies ahead.

With respect to the comparison of VaR forecasts obtained using different procedures, Chen, Gerlach, Lin and Lee (2012) compare empirically a range of parametric models combined with four error distributions implemented to four Asia-Pacific stock markets. They conclude that GARCH models always outperform stochastic volatility models when estimated using MCMC methods. Furthermore, asymmetric models were favoured pre-crisis while models with Skewed-Student errors are ranked best during and post crisis. In this latter period, all models forecast VaR less accurately and anti-conservatively. Also, Gerlach *et al.* (2013) conclude that allowing for skewness in the error distribution may be important in the post-crisis period. However, in the pre-crisis period models with Student errors or symmetric Laplace distributions can do a good job. Similar conclusions are obtained by Halbleib and Pohlmeier (2012) who conclude that models with Skewed-Student distribution can do a good job during the crisis period. They also show that the CEVT procedure can have a reasonable performance.

Recently, Boucher *et al.* (2014) conclude that the RiskMetrics and GARCH-based models are among the preferred ones. However, many authors conclude that methods assuming specific densities may yield wrong forecasts when the model for the conditional moments is misspecified and conclude that EVT procedures produce the most accurate forecasts of extreme losses. For example, Zikovic and Aktan (2011) carry out an extensive empirical comparison of procedures implemented to daily returns of seven stock indexes and two commodities observed from 04/01/2000 to 02/01/2009. They conclude that the CEVT is satisfactory with the WHS being a viable alternative. In the context of exchange rates, Wang *et al.* (2010) conclude that, when compared with HS, EVT is more appropriate to forecast the VaR of the Yuan. On the contrary, other authors as, for example, Sener *et al.* (2012), rank EVT as having the worst performance in a more extensive analysis of series and procedures. Sener *et al.* (2012) conclude that VaR forecasts based on asymmetric CAViaR and EGARCH models have the best performance. The preference of CAViaR is also established by Chen, Gerlach, Hwang and McAleer (2012) and Chen and Lu (2012). Also, Tolikas (2014) have shown that when EVT is based on the GEV distribution, the corresponding VaR forecasts underestimate risk. Alternatively, he proposes using the Generalized Logistic distribution for the extremes which performs better during the crisis period. Chen and Gerlach (2013) also conclude that differences between models are dominated by the error distribution.

## 7 Conclusions

In this paper, we review recent contributions on forecasting and backtesting the VaR measure of risk. The obvious benefit of VaR is that it is easily and intuitively understood by non-specialists. The different procedures and tests have been illustrated by estimating the VaR of a series of daily S&P500 returns observed over a period that covers the recent global financial crisis.

In the empirical illustration considered in this survey, when looking at the adequacy of alternative procedures to obtain VaR forecasts, the results of a particular test could be different depending on the number of out-of-sample observations and the particular period being analyzed. There is not a procedure that clearly outperforms others and, except the EGARCH model with Skewed-Student errors, all of them are rejected by one test and/or out-of-sample period. It seems that rather simple forecasts based on modelling the evolution of conditional variance with asymmetric GARCH-type models and asymmetric leptokurtic errors are among the most

competitive ones. Time varying skewness and kurtosis of the distribution of standardized returns does not seem to improve VaR forecasts. Introducing bias correction, the uncertainty of VaR forecasts and their validation are important topics that still deserve more research to reach more conclusive results on the performance of alternative procedures.

It is important to note that although this paper is devoted to surveying VaR forecasting procedures, the recent Basel accords suggest using Expected Shortfall (ES) in place of VaR; see Acerbi and Tasche (2002) for a definition of ES. Therefore, ES is likely to gain prominence in the future; see, for example, Wu and Xiao (2002), Yamai and Yoshida (2005), Komunjer (2007), Chen (2008), Chen *et al.* (2012), Lönnbark (2013), Stoyanov *et al.* (2013), Brandtner (in press) and Xu (in press) for some recent references on forecasting ES. However, Gneiting (2011) shows that it is not possible to measure the adequacy of ES using a consistent scoring function.

## References

- [1] Acerbi, C. and D. Tasche (2002), Expected shortfall: a natural coherent alternative to value at risk, *Economic Notes*, 31, 379-388.
- [2] Angelidis, T. and S. Degiannakis (2007), Backtesting VaR models: a two-stage procedure, *Journal of Risk Model Validation*, 1(2), 1-22.
- [3] Angelidis, T., A. Benos and S. Degiannakis (2007), A robust VaR model under different time periods and weighting schemes, *Review of Quantitative Finance and Accounting*, 28, 187-201.
- [4] Ardia, D. and L.F. Hoogerheide (2014), GARCH models for daily stock returns: Impact of estimation frequency on value-at-risk and expected shortfall forecasts, *Economics Letters*, 123(2), 187-190.
- [5] Arroyo, J., G. González-Rivera, C. Maté and A. Muñoz San Roque (2011), Smoothing methods for histogram-valued time series. An application to value-at-risk, *Statistical Analysis and Data Mining*, 4(2), 216-228.
- [6] Asai, M. and M. McAleer (2008), A portfolio index GARCH model, *International Journal of Forecasting*, 24(3), 449-461.
- [7] Azzalini, A. and A. Capitanio (2003), Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution, *Journal of the Royal Statistical Society*, B65, 367-389.
- [8] Bali, T.G. and P. Theodossiou (2007), A conditional-sgt-VaR approach with alternative GARCH models, *Annals of Operations Research*, 151, 241-267.
- [9] Bali, T.G., H. Mo and Y. Tang (2008), The role of autoregressive conditional skewness and kurtosis in the estimation of VaR, *Journal of Banking & Finance*, 32(2), 269-282.
- [10] Bao, Y. and A. Ullah (2004), Bias of value-at-risk estimator, *Finance Research Letters*, 1, 241-249.

- 
- [11] Bao, Y., T.-H. Lee and B. Saltoglu (2006), Evaluating predictive performance of value-at-risk models in emerging markets: a reality check, *Journal of Forecasting*, 25, 101-128.
- [12] Barone-Adesi, G., K. Giannopoulos and L. Vosper (1999), VaR without correlations for portfolios of derivative securities, *Journal of Future Markets*, 19(5), 583-602.
- [13] Barone-Adesi, G., K. Giannopoulos and L. Vosper (2002), Backtesting derivative portfolios with filtered historical simulation (FHS), *European Financial Management*, 8, 31-58.
- [14] Berkowitz, J. (2001), Testing density forecast, with applications to risk management, *Journal of Business & Economic Statistics*, 19, 465-474.
- [15] Berkowitz, J. and J. O'Brien (2002), How accurate are value-at-risk models at commercial banks?, *Journal of Finance*, 57, 1093-1111.
- [16] Berkowitz, J., P. Christoffersen and D. Pelletier (2011), Evaluating Value-at-Risk models with desk-level data, *Management Science*, 57(2), 2213-2227.
- [17] Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, 31, 307-327.
- [18] Boucher, C.M., J. Daniélsson, P.S. Kouonchou and B.B. Maillet (2014), Risk models-at-risk, *Journal of Banking & Finance*, 44, 72-92.
- [19] Brandtner, M. (in press), Conditional value-at-risk, spectral measures and (non)diversification in portfolio selection problems - a comparison with mean-variance analysis, *Journal of Banking & Finance*.
- [20] Broda, S.A., M. Haas, J. Krause, M.S. Paoletta and S.C. Steude (2013), Stable mixture GARCH models, *Journal of Econometrics*, 172, 292-306.
- [21] Brownlees, C.T. and G.M. Gallo (2010), Comparison of volatility measures: A risk management perspective, *Journal of Financial Econometrics*, 8, 29-56.
- [22] Cai, Z. and X. Wang (2008), Nonparametric estimation of conditional VaR and expected shortfall, *Journal of Econometrics*, 147, 120-130.
- [23] Campbell, S.D. (2007), A review of backtesting and backtesting procedures, *Journal of Risk*, 9(2).
- [24] Candelon, B., G. Colletaz, C. Hurlin and S. Tokpavi (2011), Backtesting value-at-risk: A GMM duration-based approach, *Journal of Financial Econometrics*, 9(2), 314-343.
- [25] Caporin, M. (2008), Evaluating value-at-risk measures in the presence of long-memory conditional volatility, *Journal of Risk*, 10, 79-110.
- [26] Chan, K.F. and P. Gray (2006), Using extreme value theory to measure value-at-risk for daily electricity spot prices, *International Journal of Forecasting*, 22(2), 283-300.
- [27] Chan, N.H., L. Peng and R. Zhang (2012), Interval estimation of the tail index of a GARCH model, *TEST*, 21, 546-565.

- [28] Chan, N.H., S.-J. Deng, L. Peng and Z. Xia (2007), Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations, *Journal of Econometrics*, 137(2), 556-576.
- [29] Chan, N.H., D. Li, L. Peng and R. Zhang (2013), Tail index of an AR(1) model with ARCH(1) errors, *Econometric Theory*, 29(5), 920-940.
- [30] Chavez-Demoulin, V., A. Davidson and A. McNeil (2005), A point process approach to value-at-risk estimation, *Quantitative Finance*, 5(2), 227-234.
- [31] Chavez-Demoulin, V., P. Embrechts and S. Sardy (2014), Extreme-quantile tracking for financial time series, *Journal of Econometrics*, 181(1), 44-52.
- [32] Chen, C.W.S., R. Gerlach, E.M.H. Lin and W.C.W. Lee (2012), Bayesian forecasting for financial risk management, pre and post the global financial crisis, *Journal of Forecasting*, 31, 661-687.
- [33] Chen, C.W.S., R. Gerlach, B.B.K. Hwang and M. McAleer (2012), Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range, *International Journal of Forecasting*, 28(3), 557-574.
- [34] Chen, Q. and R.H. Gerlach (2013), The two-sided Weibull distribution and forecasting financial tails, *International Journal of Forecasting*, 29, 527-540.
- [35] Chen, Q., R. Gerlach and Z. Lu (2012), Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution, *Computational Statistics & Data Analysis*, 56, 3498-3516.
- [36] Chen, S.X. (2008), Nonparametric estimation of expected shortfall, *Journal of Financial Econometrics*, 87-1007.
- [37] Chen, S.X. and C.Y. Tang (2005), Nonparametric inference of value-at-risk for dependent financial returns, *Journal of Financial Econometrics*, 3(2), 227-255.
- [38] Chen, W.-H. and J.-C. Hung (2011), Skewness and leptokurtosis in GARCH-type VaR estimation of petroleum and metal asset returns, *Journal of Empirical Finance*, 18, 160-173.
- [39] Chen, Y. and J. Lu (2012), Value at risk estimation, in Duan, J.-C., W.K. Härdle and J.E. Gentle (eds.), *Handbook of Computational Finance*, Springer-Verlag, Berlin.
- [40] Chen, Y., W. Härdle and S.-O. Jeong (2008), Nonparametric risk management with Generalized Hyperbolic distributions, *Journal of the American Statistical Association*, 103(483), 910-923.
- [41] Chiu, C.L., M.C. Lee, J.C. Hung (2005), Estimation of value-at-risk under jump dynamics and asymmetric information, *Applied Financial Economics*, 15, 1095-1106.
- [42] Christoffersen, P.F. (1998), Evaluating interval forecasts, *International Economic Review*, 39, 841-862.
- [43] Christoffersen, P.F. (2009), Value-at-risk models, in T. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch (eds.), *Handbook of Financial Time Series*, Springer Verlag.



- [44] Christoffersen, P.F. (2012), *Elements of Financial Risk Management*, 2nd. ed., Academic Press, Amsterdam.
- [45] Christoffersen, P.F. and S. Goncalves (2005), Estimation risk in financial risk management, *Journal of Risk*, 7, 1-28.
- [46] Christoffersen, P.F. and D. Pelletier (2004), Backtesting value-at-risk: a duration-based approach, *Journal of Financial Econometrics*, 2(1), 84-108.
- [47] Christoffersen, P.F., J. Hahn and A. Inoue (2001), Testing and comparing value-at-risk measures, *Journal of Empirical Finance*, 8, 325-342.
- [48] Clements, M.P. and N. Taylor (2003), Evaluating interval forecasts of high-frequency financial data, *Journal of Applied Econometrics*, 18, 445-456.
- [49] Clements, M.P., A.B. Galvao and J.H. Kim (2008), Quantile forecast of daily exchange returns from forecasts of realized volatility, *Journal of Empirical Finance*, 15, 729-750.
- [50] Colletaz, G., C. Hurlin and C. Pérignon (2013), The risk map: a new tool for validating risk models, *Journal of Banking & Finance*, 37(10), 3843-3856.
- [51] Coroneo, L. and D. Veredas (2012), A simple two-component model for the distribution of intraday returns, *European Journal of Finance*, 18(9), 775-797.
- [52] Danielsson, J. (2011), *Financial Risk Forecasting*, Wiley.
- [53] Danielsson, J. and C.G. de Vries (2000), Value at risk and extreme returns, *Annales d'Economie et Statistique*, 60, 239-270.
- [54] Danielsson, J., L. de Haan, L. Peng and C.G. de Vries (2001), Using a bootstrap method to choose the sample fraction in tail index estimation, *Journal of Multivariate Analysis*, 76, 226-248.
- [55] Danielsson, J., B.N. Jorgensen, M. Sarma and C.G. de Vries (2006), Comparing downside risk measures for heavy tailed distributions, *Economics Letters*, 92(2), 202-208.
- [56] Danielsson, J., B.N. Jorgensen, G. Samorodnitsky, M. Sarma and C.G. de Vries (2013), Fat tails, VaR and subadditivity, *Journal of Econometrics*, 172, 283-291.
- [57] Dark, J.G. (2010), Estimation of time varying skewness and kurtosis with an application to Value at Risk, *Studies in Nonlinear Dynamics and Econometrics*, 14(2).
- [58] de la Pena, V.H., R. Rivera and J. Ruiz-Mata (2007), Quality control of risk measures: backtesting VaR models, *Journal of Risk*, 9(2), 39-54.
- [59] De Rossi, G. and A.C. Harvey (2009), Quantiles, expectiles and splines, *Journal of Econometrics*, 152, 179-85.
- [60] Degiannakis, S. (2004), Volatility forecasting: evidence from a fractionally integrated asymmetric power ARCH skewed-t model, *Applied Financial Economics*, 14, 1333-1342.
- [61] Dias, A. (2013), Market capitalization and value-at-risk, *Journal of Banking & Finance*, 37(12), 5248-5260.

- [62] Diebold, F.X. (1988), *Empirical Modeling of Exchange Rate Dynamics*, Springer-Verlag, New York.
- [63] Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business & Economic Statistics*, 13(3), 253-263.
- [64] Diebold, F.X., T. Schuermann and J.D. Stroughair (2000), Pitfalls and opportunities in the use of extreme value theory in risk management, *Journal of Risk Finance*, 1(2), 30-35.
- [65] Ding, Z., C. Granger and R. Engle (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 83-106.
- [66] Dowd, K. (2006), Using order statistics to estimate confidence intervals for probabilistic risk measures, *Journal of Derivatives*, 14(2), 77-81.
- [67] Dowd, K. (2007), *Measuring Market Risk*, 2nd ed., Wiley, Chichester.
- [68] Dowd, K. and D. Blake (2006), After VaR: the theory, estimation, and insurance applications of quantile-based risk measures, *The Journal of Risk and Insurance*, 73(2), 193-229.
- [69] Duan, J. (2004), Maximum likelihood estimation using price data of the derivative contract, *Mathematical Finance*, 4, 155-167.
- [70] Dufour, J.-M. (2006), Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics, *Journal of Econometrics*, 133(2), 443-477.
- [71] Dupuis, D.J. (1999), Exceedances over high thresholds: A guide to threshold selection, *Extremes*, 1, 251-261.
- [72] Dupuis, D.J., N. Papageorgiou and B. Rémillard (in press), Robust conditional variance and value-at-risk estimation, *Journal of Financial Econometrics*.
- [73] Embrechts, P. (2009), Linear correlation and EVT: properties and caveats, *Journal of Financial Econometrics*, 7(1), 30-39.
- [74] Embrechts, P., C. Klüppelberg and T. Mikosch (2000), *Modelling Extremal Events for Insurance and Finance*, Springer, London.
- [75] Engle, R.F. (2003), Risk and volatility: econometric models and financial practice, *Nobel Lecture*.
- [76] Engle, R.F. and S. Manganelli (2004), CAViaR: Conditional autoregressive value at risk by regression quantiles, *Journal of Business & Economic Statistics*, 22(4), 367-381.
- [77] Engle, R.F. and J. Russell (1998), Autoregressive conditional duration: a new model for irregularly spaced transaction data, *Econometrica*, 66, 1127-1162.
- [78] Ergen, I. (2015), Two-step methods in VaR prediction and the importance of fat tails, *Quantitative Finance*, 15(6), 1013-1030.
- [79] Escanciano, J.C. and J. Olmo (2010), Backtesting parametric value-at-risk with estimation risk, *Journal of Business & Economic Statistics*, 28(1), 36-51,

- [80] Escanciano, J.C. and J. Olmo (2011), Robust backtesting tests for value-at-risk models, *Journal of Financial Econometrics*, 9, 132-161.
- [81] Escanciano, J.C. and P. Pei (2012), Pitfalls in backtesting Historical Simulation VaR models, *Journal of Banking & Finance*, 36, 2233-2244.
- [82] Escanciano, J.C. and C. Velasco (2010), Specification tests of parametric dynamic conditional quantiles, *Journal of Econometrics*, 159, 209-221.
- [83] Fan, J. (1992), Desing-adaptive nonparametric regression, *Journal of the American Statistical Association*, 87, 99-1004.
- [84] Fan, J. and J. Gu (2003), Semiparametric estimation of value at risk, *Econometrics Journal*, 6, 260-289.
- [85] Ferraty, F. and A. Quintela-del-Río (in press), Conditional VaR and expected shortfall: A new functional approach, *Econometric Reviews*.
- [86] Ferreira, A., L. de Haan and L. Peng (2003), On optimizing the estimation of high quantiles of a probability distribution, *Statistics*, 37(5), 401-434.
- [87] Fleming, J. and C. Kirby (2003), A closer look at the relation between GARCH and stochastic autoregressive volatility, *Journal of Financial Econometrics*, 1, 365-419.
- [88] Francq, C. and J.-M. Zakoïan (2015), Risk-parameters estimation in volatility models, *Journal of Econometrics*, 184(1), 158-173.
- [89] Frésard, L., C. Pérignon and A. Wilhelmsson (2011), The pernicious effects of contaminated data in risk management, *Journal of Banking & Finance*, 35(10), 2569-2583.
- [90] Fuertes, A.-M. and J. Olmo (2013), Optimally harnessing inter-day and intra-day information for Value-at-Risk prediction, *International Journal of Forecasting*, 29 (1), 28-42.
- [91] Gaglianone, W.P., L.R. Lima, O. Linton and D.R. Smith (2011), Evaluating Value-at-Risk models via quantile regression, *Journal of Business & Economic Statistics*, 29(1), 150-160.
- [92] Gao, F. and F. Song (2008), Estimation risk in GARCH VaR and ES estimates, *Econometric Theory*, 24 (5), 1404-1424.
- [93] Gençay, R. and F. Selçuk (2004), Extreme value theory and value-at-risk: Relative performance in emerging markets, *International Journal of Forecasting*, 20(2), 287-303.
- [94] Gerlach, R.H., C.W.S. Chen and N.Y.C. Chan (2011), Bayesian time-varying quantile forecasting for value-at-risk in financial markets, *Journal of Business & Economic Statistics*, 29(4), 481-492.
- [95] Gerlach, R., Z. Lu and H. Huang (2013), Exponentially smoothing the skewed Laplace distribution for value-at-risk forecasting, *Journal of Forecasting*, 32, 534-550.
- [96] Geweke, J. and G. Amisano (2011), Hierarchical Markov normal mixture with applications to financial asset returns, *Journal of Applied Econometrics*, 26(1), 1-29.
- [97] Giacomini, R. and I. Komunjer (2005), Evaluation and combination of conditional quantile forecasts, *Journal of Business & Economic Statistics*, 23, 416-431.

- [98] Giacomini, R. and H. White (2006), Tests of conditional predictive ability, *Econometrica*, 74(6), 1545-1578.
- [99] Giannopoulos, K. and R. Tunaru (2005), Coherent risk measures under filtered historical simulation, *Journal of Banking & Finance*, 29(4), 979-996.
- [100] Giot, L. and S. Laurent (2003), Value-at-risk for long and short trading positions, *Journal of Applied Econometrics*, 18(6), 641-663.
- [101] Giot, L. and S. Laurent (2004), Modeling daily Value-at-Risk using realized volatility and ARCH-type models, *Journal of Empirical Finance*, 11, 379-398.
- [102] Gneiting, T. (2011), Making and evaluating point forecasts, *Journal of the American Statistical Association*, 106 (494), 746-762.
- [103] Gomes, M.I. and D. Pestana (2007), A sturdy reduced bias extreme quantile (VaR) estimator, *Journal of the American Statistical Association*, 102(477), 280-292.
- [104] Gomes, M.I., L. de Haan and L. Henriques Rodrigues (2008), Tail index estimation for heavy-tailed models: Accommodation of bias in weighted log-excesses, *Journal of the Royal Statistical Society, B*, 70(1), 31-52.
- [105] Gomes, M.I., M.J. Matins and M.M. Neves (2007), Improving second order reduced-bias tail index estimator, *Review of Statistics*, 5(2), 177-207.
- [106] Gomes, M.I., F. Figueiredo, L. Henriques Rodrigues and M.C. Miranda (2012), A computational sturdy of a quasi-PORT methodology for VaR based on second-order reduced-bias estimation, *Journal of Statistical Computation and Simulation*, 82(4), 587-602.
- [107] Gong, Y., L. Zhouping and L. Peng (2010), Empirical likelihood intervals for conditional value-at-risk in ARCH-GARCH models, *Journal of Time Series Analysis*, 31(2), 65-75.
- [108] González-Rivera, G., T.-H. Lee and S. Mishra (2004), Forecasting volatility: a reality check based on option pricing, utility function, value-at-risk, and predictive likelihood, *International Journal of Forecasting*, 20(4), 629-645.
- [109] González-Rivera, G., T.-H. Lee and E. Yoldas (2007), Optimality of the RiskMetrics VaR model, *Finance Research Letters*, 4,137-145.
- [110] Gonzalo, J. and J. Olmo (2004), Which extreme values are really extreme?, *Journal of Financial Econometrics*, 2, 349-369.
- [111] Gouriéroux, C. and J. Jasiak (2008), Dynamic quantile models, *Journal of Econometrics*, 147, 198-205.
- [112] Gouriéroux, C. and J. Jasiak (2010a), Value at risk, in Ait-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, North-Holland, Amsterdam.
- [113] Gouriéroux, C. and J. Jasiak (2010b), Local likelihood density estimation and value-at-risk, *Journal of Probability and Statistics*.
- [114] Gouriéroux, C. and J.-M. Zakoïan (2013), Estimation-adjusted VaR, *Econometric Theory*, 29(4), 735-770.

- 
- [115] Granger, C.W.J. (2002), Some comments on risk, *Journal of Applied Econometrics*, 17, 447-456.
- [116] Grigoletto, M. and F. Lisi (2009), Looking for skewness in financial time series, *Econometrics Journal*, 12(2), 310-323.
- [117] Grigoletto, M. and F. Lisi (2011), Practical implications of higher moments in risk management, *Statistical Methods and Applications*, 20, 487-506.
- [118] Guermat, C. and R. Harris (2001), Robust conditional variance estimation and value-at-risk, *Journal of Risk*, 4, 25-41.
- [119] Haas, M. (2005), Improved duration-based backtesting of value-at-risk, *Journal of Risk*, 8(2), 17-38.
- [120] Halbleib, R. and W. Pohlmeier (2012), Improving the value at risk forecasts: Theory and evidence from financial crisis, *Journal of Economic Dynamics & Control*, 36(8), 1212-1228.
- [121] Hall, P. and Q. Yao (2003), Inference in ARCH and GARCH models with heavy-tailed errors, *Econometrica*, 71, 285-317.
- [122] Hansen, B.E. (1994), Autoregressive conditional density estimation, *International Economic Review*, 35, 705-730.
- [123] Hansen, P.R. (2005), A test for superior predictive ability, *Journal of Business & Economic Statistics*, 23(4), 365-380.
- [124] Hartz, C., S. Mittnik and M. Paoletta (2006), Accurate value-at-risk forecasting based on the Normal-GARCH model, *Computational Statistics & Data Analysis*, 51(4), 2295-2312.
- [125] Herrera, R. and B. Schipp (2013), Value at risk forecasts by extreme value models in a conditional duration framework, *Journal of Empirical Finance*, 23, 33-47.
- [126] Herwartz, H. (2009), Exact inference in diagnosing Value-at-Risk estimates - A Monte Carlo evidence, *Economics Letters*, 103, 160-162.
- [127] Herwartz, H. and I. Waichman (2010), A comparison of bootstrap and Monte-Carlo testing approaches to value-at-risk diagnosis, *Computational Statistics*, 25, 725-732.
- [128] Hill, B.M. (1975), A simple general approach to inference about the tail of a distribution, *The Annals of Statistics*, 3(5), 1163-1174.
- [129] Hill, J.B. (2010), On tail index estimation for dependent heterogeneous data, *Econometric Theory*, 26(5), 1398-1436.
- [130] Hoogerheide, L. and H.K. van Dijk (2010), Bayesian forecasting of Value at Risk and Expected Shortfall using adaptive importance sampling, *International Journal of Forecasting*, 26, 231-247.
- [131] Hoogerheide, L., F. Ravazzolo and H.K. van Dijk (2012), Comment, *Journal of Business & Economic Statistics*, 30(1), 30-33.

- [132] Huang, D., B. Yu, Z. Luz, F. Fabozzi, S. Focardi and M. Fukushima (2010), Index-exciting Caviar: a new empirical time-varying risk model, *Studies in Nonlinear Dynamics and Econometrics*, 14(1).
- [133] Huang, H. and T.-H. Lee (2013), Forecasting value-at-risk using high frequency information, *Econometrics*, 1, 127-140.
- [134] Huisman, R. K.G. Koedijk, C.J.M. Kool and F. Palm (2001), Tail index estimates in small samples, *Journal of Business & Economic Statistics*, 19, 208-216.
- [135] Hull, J. and A. White (1998), Incorporating volatility updating into the historical simulation method for value at risk, *Journal of Risk*, 1, 5-19.
- [136] Hurlin, C. and S. Tokpavi (2007), Backtesting value-at-risk accuracy: a simple new test, *Journal of Risk*, 9(2), 19-37.
- [137] Jalal, A. and M. Rockinger (2008), Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data, *Journal of Empirical Finance*, 15, 868-877.
- [138] Jeon, J. and J.W. Taylor (2013), Using CAViaR models with implied volatility for value-at-risk estimation, *Journal of Forecasting*, 32, 62-74.
- [139] Jorion, P. (1996), Risk<sup>2</sup>: Measuring the risk in value at risk, *Financial Analysts Journal*, 52, 47-56.
- [140] Jorion, P. (2002), How informative are value-at-risk disclosures?, *Accounting Review*, 77(4), 911-931.
- [141] Jorion, P. (2006), *Value-at-Risk: The New Benchmark for Managing Financial Risk*, 3rd ed., McGraw-Hill, New York.
- [142] Jorion, P. (2009), Risk management lessons from the credit crisis, *European Financial Management*, 15(5), 923-933.
- [143] Juarez, S.F. and W.R. Schucany (2004), Robust and efficient estimation for Generalized Pareto estimation, *Extremes*, 7, 237-251.
- [144] Kerkhof, J. and B. Melenberg (2004), Backtesting for risk-based regulatory capital, *Journal of Banking & Finance*, 28, 1845-1865.
- [145] Koenker, R. and G. Basset (1978), Regression quantiles, *Econometrica*, 46, 33-50.
- [146] Koenker, R. and Z. Xiao (2002), Inference on the quantile regression process, *Econometrica*, 70(4), 1583-1612.
- [147] Komunjer, I. (2005), Quasi-maximum likelihood estimation for conditional quantiles, *Journal of Econometrics*, 128, 137-164.
- [148] Komunjer, I. (2007), Asymmetric power distribution: theory and applications to risk measurement, *Journal of Applied Econometrics*, 22, 891-921.
- [149] Krause, A. (2003), Exploring the limitations of value at risk: How good is it in practice?, *The Journal of Risk and Finance*, Winter, 19-28.

- [150] Kuester, K., S. Mittik and M.S. Paoletta (2006), Value-at-risk prediction: A comparison of alternative strategies, *Journal of Financial Econometrics*, 4(1), 53-89.
- [151] Kupiec, P. (1995), Techniques for verifying the accuracy of risk measurement models, *Journal of Derivatives*, 2, 173-184.
- [152] Leccadito, A., S. Boffelli and G. Urga (2014), Evaluating the accuracy of value-at-risk forecasts: New multilevel tests, *International Journal of Forecasting*, 30, 206-216.
- [153] Lehar, A., M. Scheicher and C. Schittenkopf (2002), GARCH vs. stochastic volatility: option pricing and risk management, *Journal of Banking & Finance*, 26, 323-345.
- [154] Li, Z., Y. Gong and L. Peng (2011), Empirical likelihood intervals for conditional value-at-risk in heteroscedastic regression models, *Scandinavian Journal of Statistics*, 38(4), 781-787.
- [155] Lien, D., X. Yan and K. Ye (2014), Alternative approximations to value-at-risk, *Communications in Statistics-Simulation and Computation*, 43(10), 2225-2240.
- [156] Li, D., L. Peng and J. Yan (2010), Bias reduction for high quantiles, *Journal of Statistical Planning and Inference*, 140(6), 2433-2441.
- [157] Liu, S. and Y.-K. Tse (in press), Intraday value-at-risk: An asymmetric autoregressive conditional duration approach, *Journal of Econometrics*.
- [158] Longin, F. (2000), From value at risk to stress testing: the extreme value approach, *Journal of Banking & Finance*, 24, 1097-1130.
- [159] Lönnbark, C. (2010), A corrected value-at-risk predictor, *Applied Economic Letters*, 17(12), 1193-1196.
- [160] Lönnbark, C. (2013), On the role of the estimation error in prediction expected shortfall, *Journal of Banking & Finance*, 37(3), 847-853.
- [161] Lopez, A.J. (1999), Regulatory evaluation of value-at-risk models, *Journal of Risk*, 1, 37-64.
- [162] Louzis, D.P., S. Xanthopoulos-Sisinis and A.P. Refenes (2013), The role of high-frequency intra-daily data, daily range and implied volatility in multiperiod value-at-risk forecasting, *Journal of Forecasting*, 32(6), 561-576.
- [163] Maheu, J.M. and T.H. McCurdy (2004), News arrival, jump dynamics and volatility components for individual stock returns, *Journal of Finance*, 59, 755-793.
- [164] Mancini, L. and F. Trojani (2011), Robust value-at-risk prediction, *Journal of Financial Econometrics*, 9(2), 281-313.
- [165] Mancini, L., E. Ronchetti and F. Trojani (2005), Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models, *Journal of the American Statistical Association*, 100, 628-641.
- [166] Martens, M., D. van Dijk and M. Pooter (2009), Modeling and forecasting S&P500 volatility: long memory, level shifts, leverage effects, day-of-the-week seasonality and macroeconomic announcements, *International Journal of Forecasting*, 25, 282-303.

- [167] Martins-Filho, C. and F. Yao (2006), Estimation of value-at-risk and expected shortfall based on nonlinear models of returns dynamics and extreme value theory, *Studies in Non-linear Dynamics & Econometrics*, 10(2), 1-43.
- [168] McAleer, M. (2009), The ten commandments for optimizing value-at-risk and daily capital charges, *Journal of Economic Surveys*, 23(5), 831-849.
- [169] McAleer, M. and B. da Veiga (2008), Single index and portfolio models for forecasting value-at-risk thresholds, *Journal of Forecasting*, 27, 217-235.
- [170] McAleer, M., J.-A. Jiménez-Martín and T. Pérez-Amaral (2010), What happened to risk management during the 2008-09 financial crisis?, in Kolb, R.W. (ed.), *Lessons from the Financial Crisis: Causes, Consequences, and our Economic Future*, Wiley: New York.
- [171] McAleer, M., J.-A. Jiménez-Martín and T. Pérez-Amaral (2013a), Has the Basel Accord improved risk management during the global financial crisis?, *The North American Journal of Economics and Finance*, 26(C), 250-265.
- [172] McAleer, M., J.-A. Jiménez-Martín and T. Pérez-Amaral (2013b), International evidence on GFC-robust forecasts for risk management under the Basel Accords, *Journal of Forecasting*, 32, 267-288.
- [173] McNeil, A.J. and R. Frey (2000), Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, *Journal of Empirical Finance*, 7, 271-300.
- [174] McNeil, A.J., R. Frey and P. Embrechts (2005), *Quantitative Risk Management: Concepts, Techniques, Tools*, Princeton University Press.
- [175] Mincer, J. and V. Zarnowitz (1969), The evaluation of economic forecasts and expectations, in Mincer, J. (ed.), *Economic Forecasts and Expectations*, National Bureau of Economic Research, New York.
- [176] Moraux, F. (2011), How valuable is your VaR? Large sample confidence intervals for normal VaR, *Journal of Risk Management in Financial Institutions*, 4(2), 189-200.
- [177] Nelson, D. (1991), Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, 59, 347-370.
- [178] Newey, W.K. and K.D. West (1987), Hypothesis testing with efficient method of moments estimation, *International Economic Review*, 28(3), 777-787.
- [179] Pascual, L., E. Ruiz and J. Romo (2006), Bootstrap prediction for returns and volatilities in GARCH models, *Computational Statistics & Data Analysis*, 50(9), 2293-2312.
- [180] Patton, A. (2006), Modelling asymmetric exchange rate dependence, *International Economic Review*, 47, 527-556.
- [181] Patton, A. and A. Timmermann (2012), Forecast rationality tests based on multi-horizon bounds, *Journal of Business & Economics Statistics*, 30(1), 1-17.
- [182] Peng, L. and Y. Qi (2006a), Confidence regions for high quantiles of a heavy distribution, *Annals of Statistics*, 34(4), 1964-1986.



- [183] Peng, L. and Y. Qi (2006b), A new calibration method of constructing empirical likelihood-based confidence intervals for the tail index, *Australian and New Zealand Journal of Statistics*, 48(1), 59-66.
- [184] Pérignon, C. and D.R. Smith (2008), A new approach to comparing VaR estimation method, *The Journal of Derivatives*, 16(2), 54-66.
- [185] Pérignon, C. and D.R. Smith (2010a), Diversification and Value-at-Risk, *Journal of Banking & Finance*, 34, 55-66.
- [186] Pérignon, C. and D.R. Smith (2010b), The level and quality of Value-at-risk disclosure by commercial banks, *Journal of Banking & Finance*, 34, 362-377.
- [187] Pérignon, C., Z.Y. Deng and Z.Y. Wang (2008), Do banks overstate their value-at-risk?, *Journal of Banking & Finance*, 32, 783-794.
- [188] Politis, D.N. and J.P. Romano (1994), The stationary bootstrap, *Journal of the American Statistical Association*, 89(428), 1303-1313.
- [189] Politis, D.N. and H. White (2004), Automatic block-length selection for the dependent bootstrap, *Econometric Reviews*, 23(1), 53-70.
- [190] Politis, D.N. and H. White (2009), Correction to "Automatic block-length selection for the dependent bootstrap", *Econometric Reviews*, 28(4), 372-375.
- [191] Pritsker, M. (2006), The hidden dangers of historical simulation, *Journal of Banking & Finance*, 30(2), 561-582.
- [192] Ridder, T. (1998), Basics of statistical VaR-estimation, in Bol, G., G. Nakhaeizadeh and K.-H. Vollmer (eds.), *Risk Measurement, Econometrics and Neural Networks*, Physica-Verlag, New York.
- [193] Rocco, M. (2014), Extreme value theory in finance: A survey, *Journal of Economic Surveys*, 28(1), 82-108.
- [194] Rodríguez, M.J. and E. Ruiz (2012), Revisiting several popular GARCH models with leverage effect: differences and similarities, *Journal of Financial Econometrics*, 10(4), 637-668.
- [195] Rubia, A. and L. Sanchis-Marco (2013), On downside risk predictability through liquidity and trading activity: A dynamic quantile approach, *International Journal of Forecasting*, 29(1), 202-219.
- [196] Ruiz, E. and L. Pascual (2002), Bootstrapping financial time series, *Journal of Economic Surveys*, 16, 271-300.
- [197] Sajjad, R., J. Coakley and T. Nankervis (2008), Markov Switching GARCH modelling of Value at Risk, *Studies in Nonlinear Dynamics and Econometrics*, 12(3).
- [198] Sandström, A. (2011), *Handbook of Solvency for Actuaries and Risk Managers Theory and Practice*, Chapman & Hall, Boca Raton.
- [199] Santos A.P.P. and M.F. Alves (2012a), Forecasting value-at-risk with a duration-based POT method, *Mathematics and Computers in Simulation*, 94, 295-309.

- [200] Santos, A.P.P. and M.F. Alves (2012b), A new class of independence tests for interval forecast evaluation, *Computational Statistics & Data Analysis*, 56(11), 3366-3380.
- [201] Sarma, M., S. Thomas and A. Shah (2003), Selection of value at risk models, *Journal of Forecasting*, 22, 337-358.
- [202] Scarrott, C. and A. MacDonald (2012), A review of extreme value threshold estimation with uncertainty quantification, *REVSTAT, Statistical Journal*, 10(1), 33-60.
- [203] Sener, E., S. Baronyan and L.A. Mengütürk (2012), Ranking the predictive performances of value-at-risk estimation methods, *International Journal of Forecasting*, 28(4), 849-873.
- [204] Sherman, M. and E. Carlstein (2004), Confidence intervals based on estimators with unknown rates of convergence, *Computational Statistics & Data Analysis*, 46, 123-139.
- [205] So, M. and P. Yu (2006), Empirical analysis of GARCH models in value at risk estimation, *Journal of International Markets, Institutions & Money*, 18, 180-197.
- [206] Spierdijk, L. (in press), Confidence intervals for ARMA-GARCH value-at-risk: The case of heavy tails and skewness, *Computational Statistics & Data Analysis*.
- [207] Stoyanov, S.V., S.T. Rachev and F.J. Fabozzi (2013), CVaR sensitivity with respect to tail thickness, *Journal of Banking & Finance*, 37, 977-988.
- [208] Taniai, H. and M. Taniguchi (2008), Statistical estimation errors of VaR under ARCH returns, *Journal of Statistical Planning and Inference*, 139, 3568-3577.
- [209] Taylor, J.W. (2008a), Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics*, 6(2), 231-252.
- [210] Taylor, J.W. (2008b), Using exponentially weighted quantile regression to estimate value at risk and expected shortfall, *Journal of Financial Econometrics*, 6(3), 382-406.
- [211] Theodossiou, P. (1998), Financial data and the skewed generalized t distribution, *Management Science*, 44, 1650-1661.
- [212] Tolikas, K. (2014), Unexpected tail in risk measurement: Some international evidence, *Journal of Banking & Finance*, 40, 476-493.
- [213] Wang, H.J., D. Li and X. He (2012), Estimation of high conditional quantiles for heavy-tailed distributions, *Journal of the American Statistical Association*, 107(500), 1453-1464.
- [214] Wang, Z., W. Weitao, C. Chen and Y. Zhou (2010), The exchange rate risk of Chinese Yuan: Using VaR and ES based on extreme value theory, *Journal of Applied Statistics*, 37(2), 265-282.
- [215] Wilhelmsson, A. (2009), Value at risk with time varying variance, skewness and kurtosis—the NIG-ACD model, *The Econometrics Journal*, 12, 82-104.
- [216] Wong, W.K. (2010), Backtesting value-at-risk based on tail losses, *Journal of Empirical Finance*, 17(3), 526-538.
- [217] Wu, G. and Z. Xiao (2002), An analysis of risk measures, *Journal of Risk*, 4, 53-75.

- 
- [218] Xiao, Z. and R. Koenker (2009), Conditional quantile estimation and inference for GARCH models, *Journal of the American Statistical Association*, 104(488), 1696-1712.
- [219] Xu, K.-L. (2013), Nonparametric inference for conditional quantiles of time series, *Econometric Theory*, 29(4), 673-698.
- [220] Xu, K.-L. (in press), Model-free inference for tail risk measures, *Econometric Theory*.
- [221] Yamai, Y. and T. Yoshihara (2005), Value-at-risk versus expected shortfall: a practical perspective, *Journal of Banking & Finance*, 29, 997-1015.
- [222] Yi, Y., X. Feng and Z. Huang (2014), Estimation of extreme value-at-risk: An EVT approach for quantile GARCH model, *Economics Letters*, 124(3), 378-381.
- [223] Yu, K. and M.C. Jones (1998), Local linear quantile regression, *Journal of the American Statistical Association*, 93(441), 228-237.
- [224] Yu, P.L.H., W.K. Li and S. Jin (2010), On some models for value-at-risk, *Econometric Reviews*, 29(5-6), 622-641.
- [225] Ziggel, D., T. Berens, G.N.F. WeiB and D. Wied (2014), A new set of improved value-at-risk backtests, *Journal of Banking & Finance*, 14, 29-41.
- [226] Zikovic, S. and B. Aktan (2011), Decay factor optimization in time weighted simulation-Evaluating VaR performance, *International Journal of Forecasting*, 27(4), 1147-1159.

Table 1. Summary of VaR forecast procedures

Procedures	Mnemonics	Selected recent papers	Extensions	
<b>Direct (one-step) forecasts of the quantile of returns</b>				
Historical Simulation	HS	Pritsker (2006)		
Weighted Hist.Simul.	WHS	Zikovic and Aktan (2011)		
			Dy. Adaptive Quantile	Gouri. and Jasiak (2008)
Adaptive	CAV(1)		Signal extraction	DRos. and Harvey (2009)
Absol. Value	CAV(2)	Engle and Manganelli (2004)	Asymmetry	Gerlach et al. (2011)
CAViaR			Time-varying param.	Huang et al. (2010)
Asym. slope	CAV(3)		Explanatory variables	Rub. and Sanc-Mar (2013)
Indirect GARCH	IG		Cond. Autoreg. Expect.	Taylor (2008a)
Kernel smoothing		Cai and Wang (2008)		
Nonparametric	Art.neural network	Ferra. and Quin.-R. (in press)		
	Quantile regression	Xu (2013), Yan et al. (2013)		
Block Maxima	EVT-BM	Diebold et al. (2000)		
Ext. Value Th.	Peak over threshold	EVT-POT	Chav.-Dem. et al. (2014)	
	Hill	EVT-H	Hill (1975)	
<b>(Two-step) forecasts: conditional variance and quantile of standardized returns</b>				
<b>Conditional variance</b>				
GARCH-type		Bali and Theodosiou (2007)		
Stochastic Volatility	SV	Chen, Ger., Lin and Lee (2012)		
Nonparametric				
Realized volatility		Louzis et al. (2013)		
<b>Quantile of standardized returns</b>				
Given distribution				
Conditional Historical Simulation	CHS	Hull and White (1998)	Robust	Dupuis et al. (in press)
Conditional Extreme Value Theory	CEVT	McNeil and Frey (2000)	Robust	Mancini and Trojani (2011)
Filtered Historical Simulation	FHS	Bar.-Ad. et al. (1999, 2002)		
Bayesian		Hoog. and van Dijk (2010)		
Quantile regression				

Table 2 Summary of VaR backtesting procedures

Procedures		Mnemonics	Selected recent papers	Extensions	
<b>Adequacy</b>					
Single level	Number of Violations	Unconditional coverage	UC	Kupiec (1995)	Parameter uncertainty
		Conditional coverage	CC	Christoffersen (1998)	Bootstrap
	Magnitude of violations	Likelihood ratio	LR	Christoffersen (1998)	Logit regressions
		Portmanteau	LB	Berkowitz et al. (2011)	
		Dynamic Quantile	DQ	Engle and Manganelli (2004)	
		Tail risk		Wong (2010)	
	Durations	Risk map	TR	Colletaz et al. (2013)	
				Christoffersen and Pelletier (2004)	
	Quantile regression		VQR	Gaglionone et al. (2011)	
	Forecast rationality			Hoogerheide et al. (2012)	
Multiple levels	Whole density		Berkowitz (2001)		
	Violations		Hurlin and Topkavi (2007)		
	Omnibus		Escanciano and Velasco (2010)		
	Perason-type		Leccacito et al. (2014)		
<b>Ranking</b>					
Regulatory Loss Function		RLF	Lopez (1999)		
Generalized Piecewise Linear		GPL	Gneiting (2011)		
Predictive Quantile Loss Function		PQLF	Giacomini and Komunjer (2005)		
Median of loss differential			Sarma et al. (2003)		
Conditional predictive ability		CPA	Giacomini and White (2006)		
Superior predictive ability		SPA	Bao et al. (2007)	No benchmark Sener	

Table 3. p-values of backtesting tests of S&P500 one-step-ahead 1% VaR forecasts obtained using alternative procedures during different periods: (1) 28th May 2010 - 19th May 2014 ( $H = 1000$ ); (2) 22nd May 2013 - 19th May 2014 ( $H = 250$ ). CAViaR computed using (1) Adaptive; (2) Absolute value; (3) Asymmetric. The models rejected when the significance level is 10% appear in bold.

	<i>LRCC</i>		<i>LB(5)</i>		<i>LB(20)</i>		<i>DQ</i>	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
<i>HS</i>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.02</b>	0.33
<i>WHS</i>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.02</b>	<b>0.00</b>	<b>0.08</b>	<b>0.00</b>	<b>0.00</b>
<i>CAViaR(1)</i>	0.16	<b>0.08</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	0.28
<i>CAViaR(2)</i>	0.38	<b>0.00</b>	1.00	<b>0.00</b>	0.13	<b>0.00</b>	<b>0.00</b>	0.28
<i>CAViaR(3)</i>	0.57	0.56	<b>0.02</b>	1.00	<b>0.01</b>	1.00	0.62	0.60
<i>IG</i>	<b>0.03</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.08</b>	0.28
<i>GARCH - N</i>	<b>0.05</b>	0.34	<b>0.04</b>	0.99	0.52	0.94	<b>0.02</b>	0.25
<i>GARCH - t</i>	<b>0.03</b>	<b>0.08</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.07</b>	0.28
<i>GARCH - S</i>	0.75	0.56	<b>0.01</b>	1.00	<b>0.08</b>	1.00	0.71	0.58
<i>EGARCH - N</i>	<b>0.01</b>	0.63	0.76	1.00	0.62	1.00	<b>0.01</b>	0.31
<i>EGARCH - t</i>	0.21	0.93	1.00	1.00	<b>0.01</b>	1.00	0.12	0.70
<i>EGARCH - S</i>	0.90	0.93	0.12	1.00	0.55	1.00	0.99	0.68
<i>CHS</i>	0.90	0.56	<b>0.00</b>	1.00	<b>0.00</b>	1.00	0.12	0.53
<i>CEVT - MB</i>	0.43	<b>0.08</b>	1.00	<b>0.00</b>	1.00	<b>0.00</b>	1.00	0.28
<i>CEVT - GPD</i>	0.38	0.56	<b>0.00</b>	1.00	0.14	1.00	0.45	0.58
<i>CEVT - Hill</i>	<b>0.03</b>	0.56	<b>0.00</b>	1.00	<b>0.00</b>	1.00	<b>0.07</b>	0.58
<i>FHS</i>	<b>0.08</b>	<b>0.08</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.00</b>	<b>0.04</b>	0.28
<i>Bootstrap</i>	0.76	0.56	<b>0.01</b>	1.00	0.76	1.00	0.82	0.59

Table 4. Comparison of VaR forecasts obtained by alternative procedures. The values reported for  $C(m)$  are the statistic obtained for each model. The minimum  $C(m)$  appears in bold. The values reported for the SPA test are p-values of the null that the model is inferior to all other models considered. In bold the models not rejected as inferior when the significance level is 10%.

	28th May 2010-19th May 2014				22nd May 2013-19th May 2014			
	$C(m)$		SPA		$C(m)$		SPA	
	RLF	PQLF	RLF	PQLF	RLF	PQLF	RLF	PQLF
<i>CAViaR(1)</i>	-	-	-	-	<b>0.02</b>	0.17	1.00	<b>0.00</b>
<i>CAViaR(2)</i>	2.22	0.54	0.47	<b>0.00</b>	-	-	-	-
<i>CAViaR(3)</i>	9.24	0.32	0.61	0.85	<b>0.00</b>	0.06	0.93	0.99
<i>GARCH - N</i>	-	-	-	-	0.35	0.06	0.36	1.00
<i>GARCH - t</i>	-	-	-	-	0.35	0.06	1.00	<b>0.00</b>
<i>GARCH - S</i>	4.42	0.33	0.43	0.65	0.03	0.06	0.89	0.95
<i>EGARCH - N</i>	8.45	0.33	<b>0.06</b>	0.70	1.29	0.06	0.65	0.89
<i>EGARCH - t</i>	<b>0.66</b>	0.33	<b>0.04</b>	<b>0.00</b>	0.23	0.07	0.72	<b>0.01</b>
<i>EGARCH - S</i>	2.25	<b>0.31</b>	<b>0.02</b>	0.99	0.76	0.07	0.63	0.41
<i>CHS</i>	5.56	0.33	0.38	0.75	0.07	0.06	0.78	0.95
<i>CEVT - BM</i>	2.77	0.45	0.98	<b>0.00</b>	0.08	0.10	1.00	<b>0.00</b>
<i>CEVT - GPD</i>	4.70	0.33	0.48	0.74	0.01	0.06	0.91	0.71
<i>CEVT - Hill</i>	-	-	-	-	0.02	0.06	0.79	0.72
<i>FHS</i>	-	-	-	-	0.37	0.37	1.00	<b>0.00</b>
<i>Bootstrap</i>	5.67	0.33	0.65	0.89	0.04	0.06	0.83	0.98

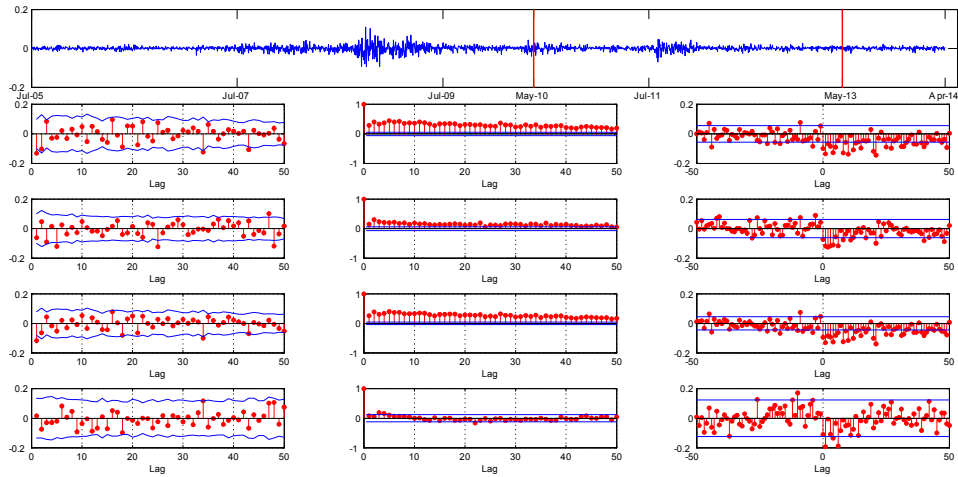


Figure 1: S&P500 returns (first row) observed from 25th July 2005 up to 19th May 2014 with vertical lines separating the in-sample and out-of-sample periods considered. Sample autocorrelations of returns (first column), sample autocorrelations of absolute returns (second column) and cross-correlations between returns and absolute returns (third column) for returns observed during: i) 25th July 2005 up to 27th May 2010 (second row); ii) 28th May 2010 up to 19th May 2014 (third row); iii) 25th July 2005 up to 21th May 2013 (fourth row); iv) 22th May 2013 up to 19th May 2014 (fifth row).



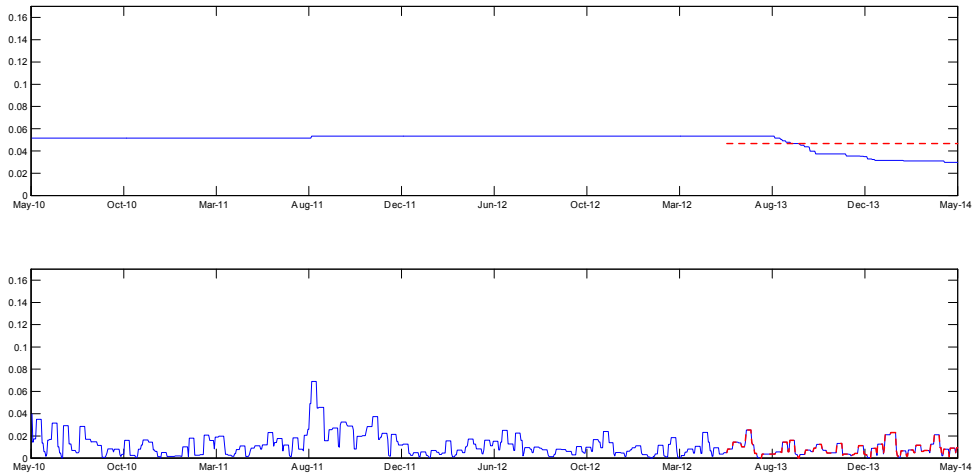


Figure 2: Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line), obtained using HS (top panel) and WHS (bottom panel) procedures.

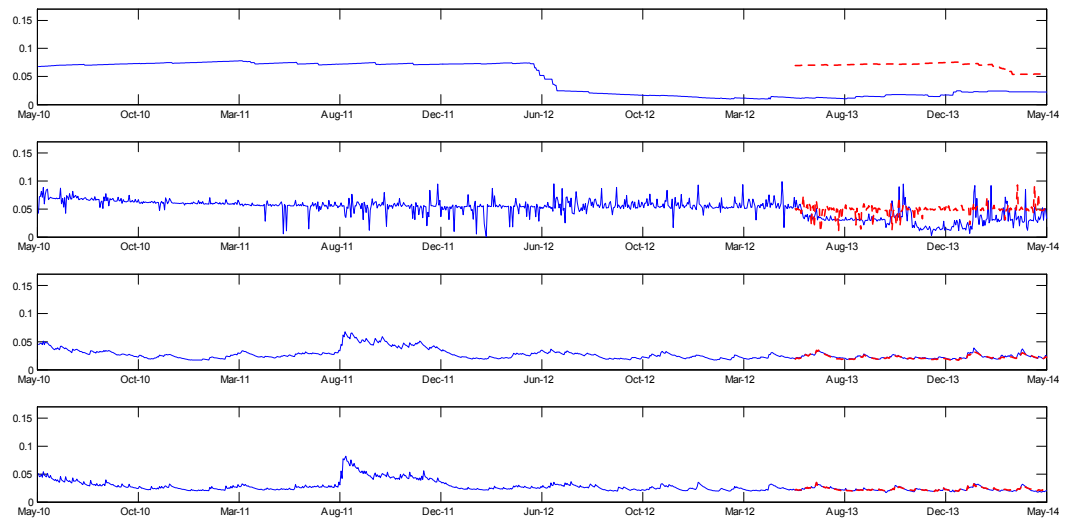


Figure 3: Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line), obtained using CAViaR with the adaptive function (first panel), the absolute value function (second panel) asymmetric function (third panel) and Indirect GARCH (fourth panel).

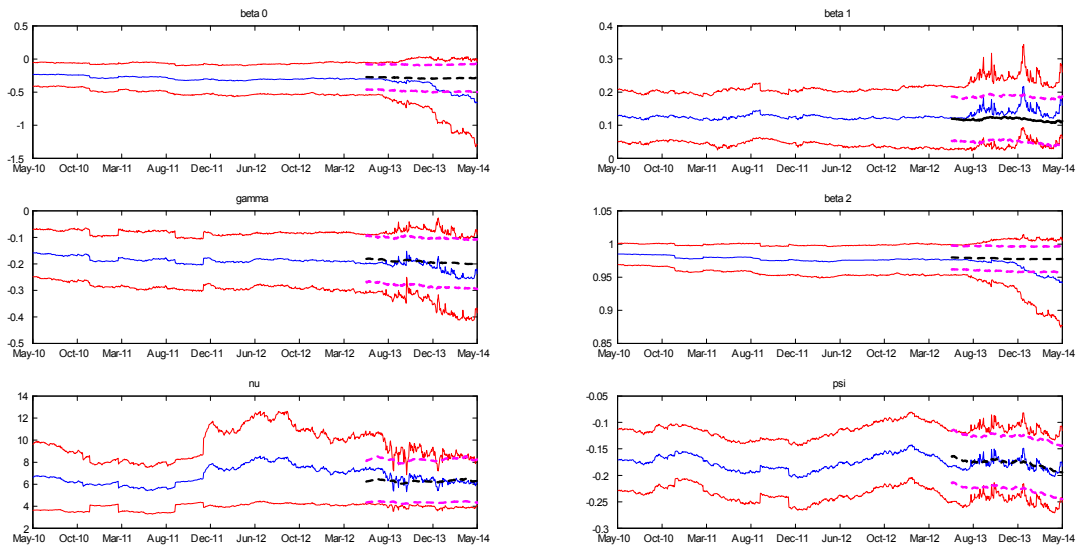


Figure 4: Rolling window estimates and 95% confidence intervals of the EGARCH(1,1) parameters obtained maximizing the Skewed-Student likelihood fitted to S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line)

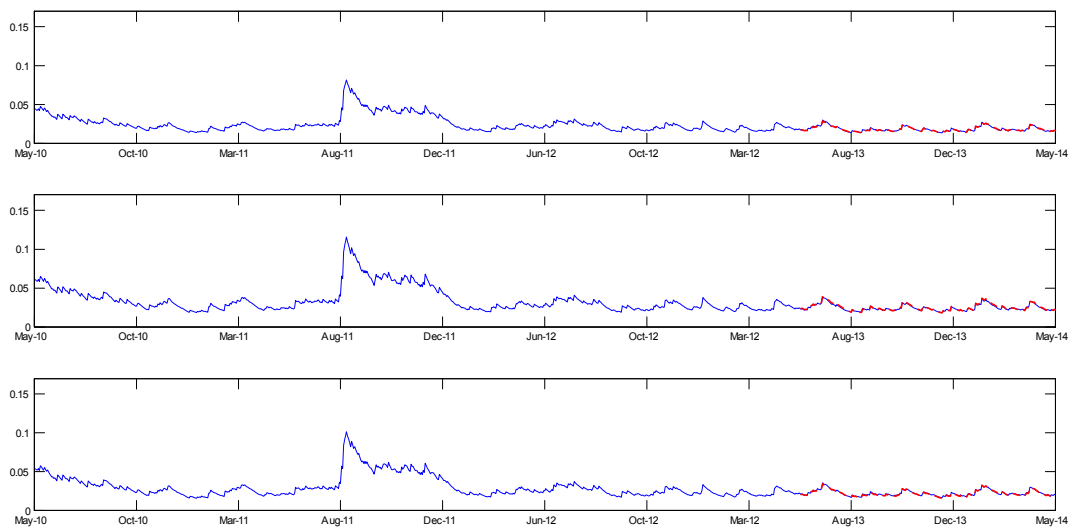


Figure 5: Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line), obtained after fitting a GARCH model assuming: i) Normal conditional returns (top panel); ii) Student conditional returns (middle panel); and iii) Skewed-t conditional returns (bottom panel).

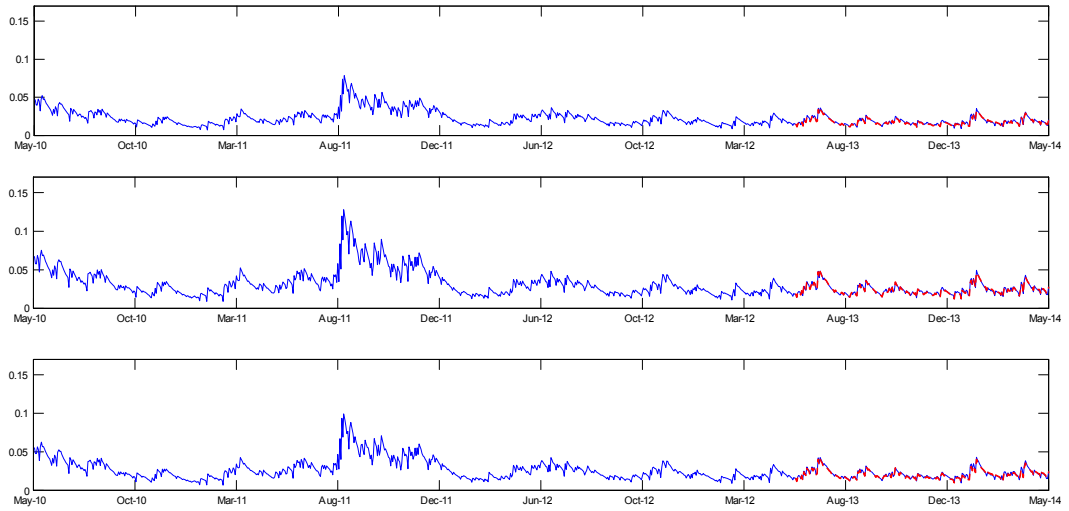


Figure 6: Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line) obtained after fitting the EGARCH model assuming: i) Normal conditional returns (top panel); ii) Student- $\nu$  conditional returns (middle panel); and iii) Skewed-t conditional returns (bottom panel).

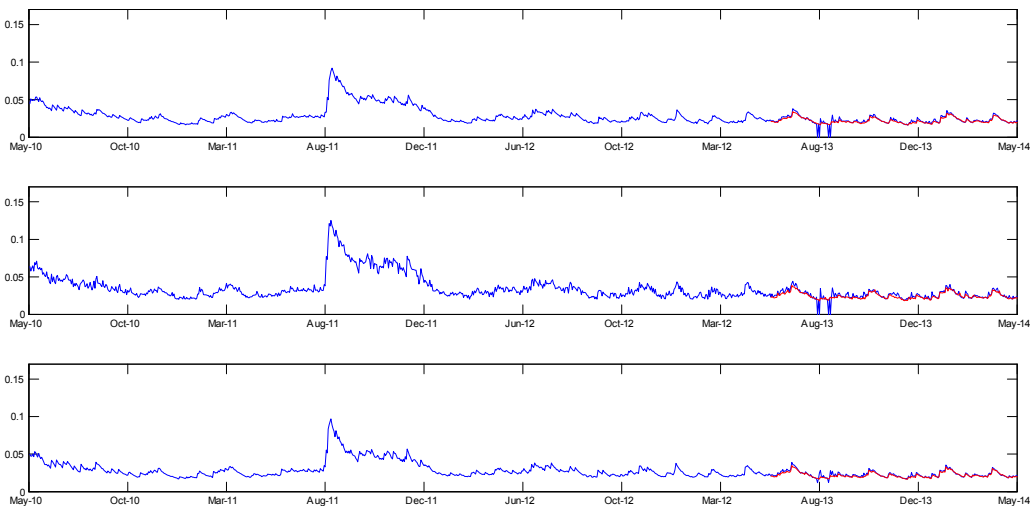


Figure 7: Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line), obtained using: a) CHS (top panel); b) FHS (middle panel); and c) Bootstrap (bottom panel).

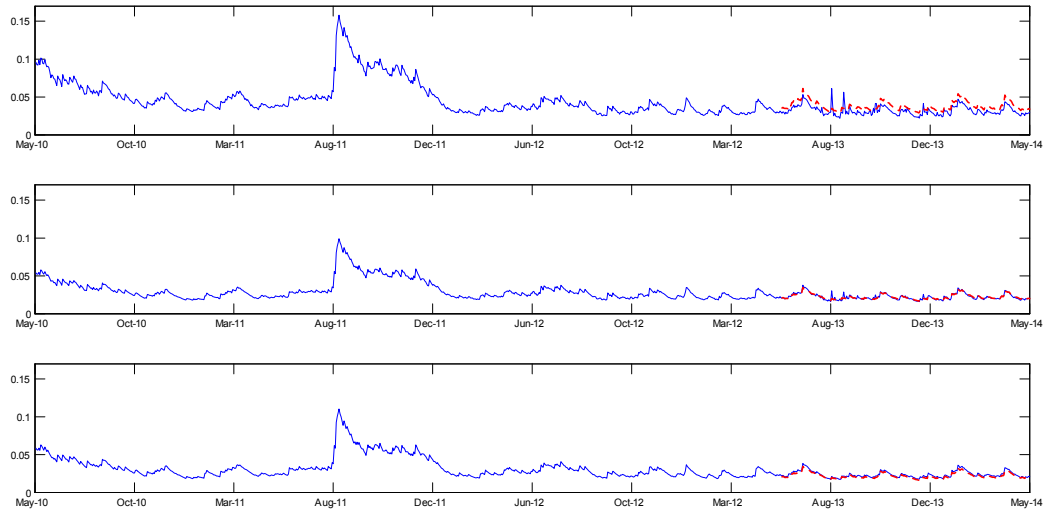


Figure 8: Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on i)  $T=1220$  and  $H=1000$  (continuous line) and ii)  $T=1970$  and  $H=250$  (discontinuous line), obtained using CEVT with i) BM (first row), ii) GDP (second row), and iii) Hill estimator.

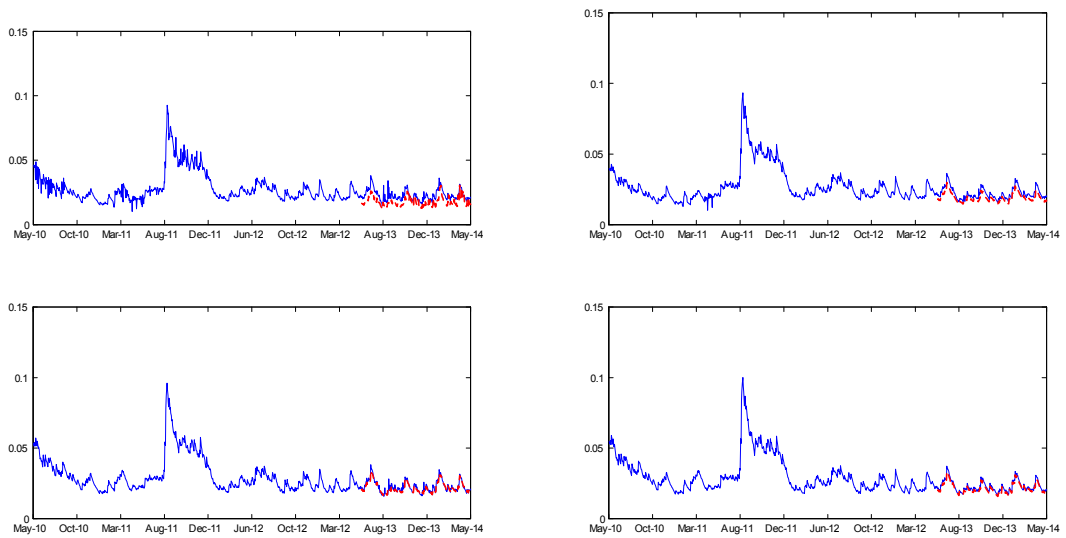


Figure 9: One-step-ahead 1% VaR forecasts of S&P500 returns obtained using CEVT with GPD (first column) and Hill (second) column procedures with the threshold chosen as proposed by Gonzalo and Olmo (2004) (first row) and Gencay and Selcuk (2004) (second row). Continuous and discontinuous lines represent the forecasts obtained with the first and second splits of the sample, respectively.