International Journal of Forecasting 32 (2016) 475–501

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Review

Frontiers in VaR forecasting and backtesting

Maria Rosa Nieto [a], Esther Ruiz [b,*]

[a] School of Business and Economics, Universidad de Anahuac México Norte, Mexico
[b] Department of Statistics, Universidad Carlos III de Madrid, Spain

ARTICLE INFO

Keywords:
Backtesting
Extreme value theory
GARCH
Quantile
Risk

ABSTRACT

The interest in forecasting the 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 testing ground when fitting alternative models for representing the dynamic evolution of time series of financial returns. There are vast numbers of alternative methods for constructing and evaluating VaR forecasts. In this paper, we survey the new benchmarks proposed in the recent literature.

Contents

* Corresponding author.
E-mail address: ortega@est-econ.uc3m.es (E. Ruiz).

http://dx.doi.org/10.1016/j.ijforecast.2015.08.003

1. Introduction

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

The 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 has attracted a great deal of attention in the financial econometrics literature, due to its relevance for financial and insurance institutions. Some adverse results in the past have forced the agencies that regulate financial activity to look for a quantitative way of defining the risk associated with a given position in the market; see Granger (2002) for alternative definitions and measures of risk. The Basel accords explicitly recognize the role of VaR as a measure that financial institutions must implement and report in order to monitor their financial risk and determine the amount of capital that is subject to regulatory control. Consequently, VaR is now established as the most popular risk measure for controlling and managing market risk. Although VaR has been analyzed mainly 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 that are operating in the European Union; see Dowd and Blake (2006) and Sandström (2011) for descriptions of applications of VaR in the insurance sector. 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, the forecasting of VaR is also important because it is implemented routinely as an empirical check for alternative models for forecasting conditional means and variances; see for example Asai and McAleer (2008), Brownlees and Gallo (2010), Grigoletto and Lisi (2009), Martens, van Dijk, and Pooter (2009), and Wilhelmsson (2009), 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 Chen and Lu (2012), Christoffersen (2009), Gourieroux and Jasiak (2010a), and Kuester, Mittik, and Paolella (2006) for previous surveys, and Christoffersen (2012), Dowd (2007), Danielsson (2011), Embrechts, Klüppelberg, and Mikosch (2000), Jorion (2006), and McNeil, Frey, and Embrechts (2005) for comprehensive textbooks. Given that the number of recent contributions related to VaR forecasting and testing is extremely large, we attempt to focus this survey by describing only univariate models, putting aside the interesting discussion on multivariate VaR forecasts. Furthermore, we consider only the VaR on the left tail of the distribution of returns, as it has attracted 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 do 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 that correspond to returns over a holding period of one day.[1] 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 the 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 bellow it. By regulatory convention, the VaR is positive; consequently, it is given by

$$\mathrm{VaR}_t^{\alpha} = -\sup\left[r \mid P\left[R_t \le r\right] \le \alpha\right]. \quad (1)$$

The probability in Eq. (1) is usually defined with respect to the distribution of returns, conditional on the information available at time $t-1$.[2]

The Basel accords describe a standard approach to obtaining VaR forecasts which is known to produce VaR estimates that are larger than necessary, leading to excessively high capital requirements (CR); see for example McAleer (2009), Pérignon, Deng, and Wang (2008) 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, it is undesirable to use the standard approach, given that regulatory capital involves an opportunity cost. Hence, they have an incentive to use their own VaR forecasts; see Pérignon and Smith (2008), who present empirical evidence that the use of internal models is widespread among large financial institutions. Although internal VaR models are subject to supervisory approval based on qualitative

[1] We focus on the horizon allowed by the Basel accords. However, the VaR horizon required in other sectors could be different.

[2] Some authors instead define the VaR with respect to the marginal distribution of returns; see for example Lien, Yan, and Ye (2014).

and quantitative standards, financial institutions enjoy a considerable degree of freedom as to the precise nature of their models. However, this flexibility does not imply that they are tempted to pursue the lowest possible VaR forecasts. The relationship between VaR and 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 that exceed the VaR). Specifically, the regulatory capital required to be held on day $t+1$ is determined as follows:

$$CR_{t+1} = \max\left\{ \mathrm{VaR}_t^\alpha, (3+k)\overline{\mathrm{VaR}}_t^\alpha \right\}, \tag{2}$$

where $\overline{\mathrm{VaR}}_t^\alpha = \frac{1}{60}\sum_{j=0}^{59} \mathrm{VaR}_{t-j}^\alpha$. The penalty k ranges between zero and one, and its value is determined by the number of VaR violations over the last 250 business days. The expression for CR in Eq. (2) suggests that lower capital charges could be achieved by lower VaR forecasts. However, this need not be the case, as lower VaR forecasts are likely to be 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, large numbers of violations may also damage the institution's reputation; see McAleer and da Veiga (2008) for a discussion. Furthermore, an underestimation of the own level of risk may lead to an amount of CR that is insufficient to cover potential losses, thus increasing the risk of bankruptcy. Finally, note that the Basel accords establish that financial institutions may be forced to adopt the standard approach if ten or more VaR violations occur over a period of 250 business days. On the other hand, an exaggeration of the own level of risk implies an excessive amount of CR, which affects the profitability of the bank directly. Another consequence that is at least as undesirable is the fact 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 both 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 their investment decisions. As a consequence, financial institutions have incentives to pursue VaR forecasts that are as accurate as possible.

The obvious benefit of VaR is that it is easy and intuitive for non-specialists to understand; see Embrechts et al. (2000), Jorion (2006) and Krause (2003). However, as a risk measure, the VaR has been criticized 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. However, Danielsson, Jorgensen, Samorodnitsky, Sarma, and de Vries (2013) show that, for 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 VaR subadditivity, and conclude that the VaR is subadditive for most practical applications. Consequently, there is a very large class of distributions and practical situations for which one does not have to worry about subadditivity violations of the VaR, meaning that 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 that exceed itself. It is possible to have two distributions with the same VaR, but with totally different losses that exceed VaR. However, Danielsson, Jorgensen, Sarma, and de Vries (2006) show that, for heavy-tailed distributions, such as those observed in financial returns, the choice of the downside risk measure does not seem to matter much, as they all 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 begin by surveying procedures that are based on forecasting the α quantile of the distribution of returns directly (one-step procedures). Next, we describe two-step procedures based on estimating first the conditional mean and variance, then the conditional quantile.[3] The various procedures are summarized in Table 1.

3.1. Data and design of the empirical study

The procedures described in this paper are illustrated by using them to forecast and backtest the 1% VaR of a series of S&P500 returns corresponding to a holding period of one day, observed daily from July 25, 2005, to May 19, 2014.[4] 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. When choosing T, Kuester et al. (2006) point out that the estimation of the models that are usually used for forecasting VaR should be based on long 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 that adapt to the market conditions easily. 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 work

3 Note that this classification is not the usual one in the literature, according to which VaR forecast procedures are classified as parametric, semiparametric or non-parametric. However, we prefer to describe, first, those procedures that estimate the quantile of the distribution of returns directly 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.

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

Table 1
Summary of VaR forecast procedures.

Procedures		Mnemonics	Selected papers	Extensions	
Direct (one-step) forecasts of the quantile of returns					
Historical Simulation		HS	Pritsker (2006)		
Weighted Hist.Simul.		WHS	Zikovic and Aktan (2011)		
	Adaptive	CAV(1)		Dy. Adaptive Quantile	Gourieroux and Jasiak (2008)
	Absol. Value	CAV(2)		Signal extraction	De Rossi and Harvey (2009)
CAViaR	Asym. slope	CAV(3)	Engle and Manganelli (2004)	Asymmetry	Gerlach et al. (2011)
	Indirect GARCH	IG		Time-varying param.	Huang et al. (2010)
				Explanatory variables	Rubia and Sanchis-Marco (2013)
				Cond. Autoreg. Expect.	Taylor (2008a)
	Kernel smoothing		Cai and Wang (2008)		
Nonparametric	Art.neural network		Ferraty and Quintela-del-Río (in press)		
	Quantile regression		Xu (2013)		
	Block Maxima	EVT-BM	Diebold et al. (2000)		
Ext. Value Th.	Peak over threshold	EVT-POT	Chavez-Demoulin et al. (2014)		
	Hill	EVT-H	Hill (1975)		
(Two-step) forecasts: conditional variance and quantile of standardized returns					
Conditional variance					
GARCH-type			Bali and Theodossiou (2007)		
Stochastic Volatility		SV	Chen, Gerlach, Lin et al. (2012)		
Nonparametric			Chen et al. (2008)		
Realized volatility			Louzis et al. (2013)		
Quantile of standardized returns					
Given distribution			Many authors		
Conditional historical simulation		CHS	Hull and White (1998)	Robust	Dupuis et al. (2015)
Conditional extreme value theory		CEVT	McNeil and Frey (2000)	Robust	Mancini and Trojani (2011)
Filtered historical simulation		FHS	Barone-Adesi et al. (1999, 2002)		
Bayesian			Hoogerheide and van Dijk (2010)		
Quantile regression			Xiao and Koenker (2009)		

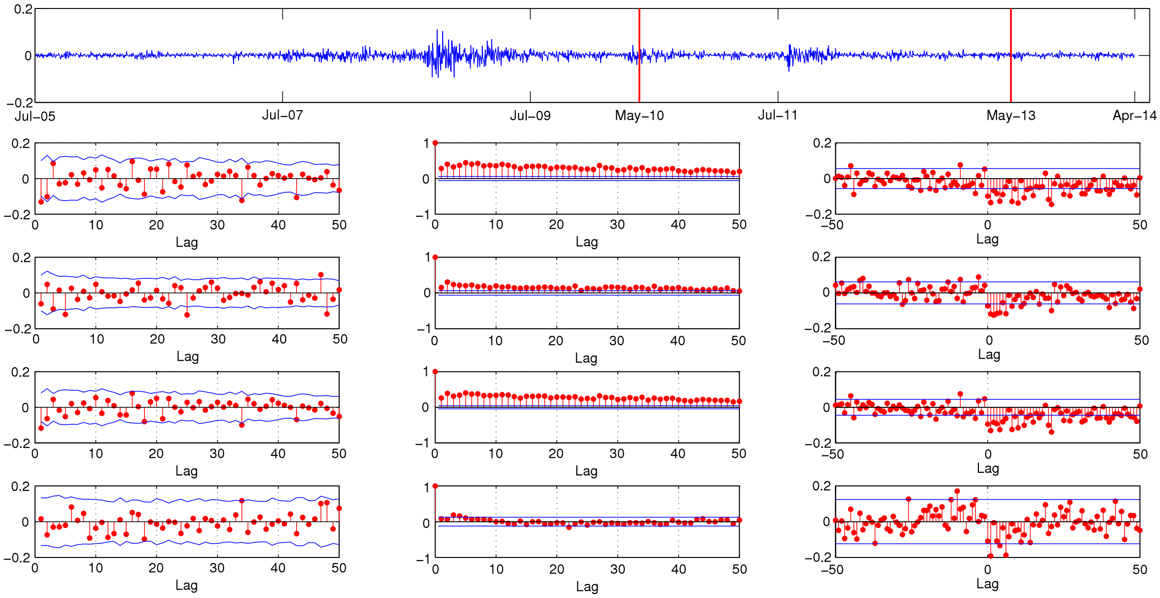

Fig. 1. S&P500 returns (first row) observed from 25th July 2005 to 19th May 2014, with vertical lines separating the in-sample and out-of-sample periods. Sample autocorrelations of returns (first column) and absolute returns (second column), and cross-correlations between returns and absolute returns (third column) computed using data from: (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).

adequately when the ratio between H and T is small; Engle and Manganelli (2004), Escanciano and Olmo (2010) and Escanciano and Pei (2012) consider $H/T = 0.1$, while Boucher, Daníelsson, Kouonchou, and Maillet (2014) choose $H/T = 250/1040$. Consequently, to illustrate the effects of differences in 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 of the VaR forecast procedures considered are implemented using a rolling window scheme, so that the size of the sample used for estimation remains fixed. Then, the VaR is forecast one-step-ahead through the out-of-sample period.

Fig. 1 plots the S&P500 returns for the full observation period, with vertical lines on the dates corresponding to the two splits considered. Fig. 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). Fig. 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, it is only in this out-of-sample period, corresponding to the last year of data, that 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 no evidence of volatility in the conditional mean of returns. Note that the evidence of conditional heteroscedasticity is stronger in the two 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 forecasting the quantile of the distribution of returns directly, 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 $t = 1, \ldots, T$. The oldest procedure for forecasting the VaR is based on historical simulation (HS) as follows

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

where $R_{(\omega)}$ is the ωth-order statistic of returns, with $\omega = [T \times \alpha]$. HS is by far the most popular procedure for forecasting VaR among commercial banks; see for example Berkowitz, Christoffersen, and Pelletier (2011), Pérignon and Smith (2008, 2010b) and Pritsker (2006). For instance, Pérignon and Smith (2010b) document that almost three-quarters of the banks that disclose their VaR method report using HS. The popularity of HS is due to its simplicity and smoothness. However, $\widehat{\mathrm{VaR}}_t^{\mathrm{HS}}$ 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 the assumption of independent and identically distributed (iid) returns, which is not an adequate

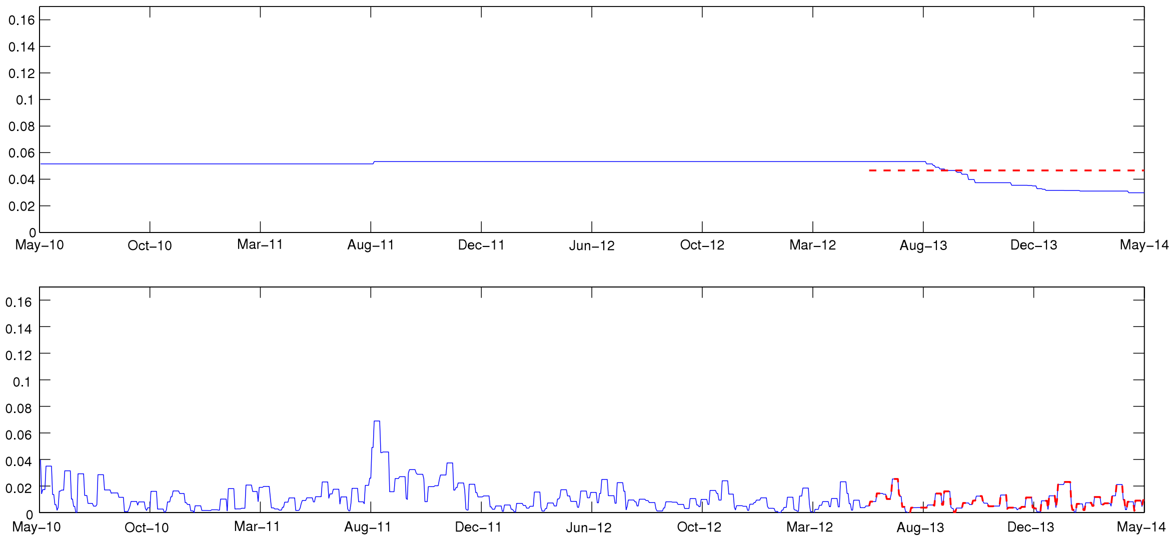

Fig. 2. Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line), and (ii) $T = 1970$ and $H = 250$ (discontinuous line), obtained using HS (top panel) and WHS (bottom panel).

assumption. Finally, note that empirical quantiles are not efficient estimators of extreme quantiles.

The HS estimator of VaR is illustrated by 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.[5,6] The top panel of Fig. 2, which plots the corresponding forecasts, shows that $\widehat{\mathrm{VaR}}_t^{HS}$ 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 influenced strongly by these observations. As a consequence, $\widehat{\mathrm{VaR}}_t^{HS}$ has a 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, as high volatility periods may have long lasting effects. The top panel of Fig. 2 also illustrates that having too many observations in the moving window for HS results in a sluggish adaptation to the dynamic changes in the true distribution of returns.

Recently, Zikovic and Aktan (2011) proposed forecasting the VaR using weighted historical simulation (WHS), by introducing an 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 quite constant over different time frames, and always close to 0.99.

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

3.2.2. CAViaR

An alternative VaR forecast procedure is based on representing the dynamic evolution of quantiles directly. The conditional autoregressive value at risk (CAViaR) was introduced by Engle and Manganelli (2004) as follows:

$$\widehat{\mathrm{VaR}}_t^{CAV} = \beta_0 + \beta_1 \widehat{\mathrm{VaR}}_{t-1}^{CAV} + \beta_2 \ell\left(X_{t-1}\right), \tag{4}$$

where $\ell(\cdot)$ is a function of a finite number of lagged observable variables, X_{t-1}, usually past returns. Three alternative specifications for $\ell(\cdot)$ are proposed: (i) the adaptive function, denoted CAV(1), in which $\beta_0 = 0$, $\beta_1 = 1$ and $\ell(\cdot) = \left\{\left[1 + \exp\left(c\left[R_{t-1} + \widehat{\mathrm{VaR}}_{t-1}^{CAV}\right]\right)\right]^{-1} - \alpha\right\}$,

[5] Note that sample sizes of $T = 1220$ or 1970 could seem to be too large for computing $\widehat{\mathrm{VaR}}_t^{HS}$, and most authors use smaller sample sizes. However, some works have used similar sized or even larger samples. For example, Angelidis, Benos, and Degiannakis (2007) estimate $\widehat{\mathrm{VaR}}_t^{HS}$ with $T = 1750$ observations, while Bao, Lee, and Saltoglu (2006) and McNeil and Frey (2000) consider sample sizes of more than 2000 observations.

[6] All of the estimates in this paper have been obtained using Matlab codes written by the first author.

where c is some positive finite number; (ii) the absolute value function, denoted CAV(2), in which $\ell(\cdot) = |R_{t-1}|$; (iii) the asymmetric slope, denoted CAV(3), in which $\ell(\cdot) = \beta_3 (R_{t-1})^+ + \beta_4 (R_{t-1})^-$, where $(x)^+ = \max(x, 0)$ and $(x)^- = -\min(x, 0)$. Finally, they propose the indirect GARCH specification, which is given by

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

The indirect GARCH in Eq. (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 using 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.

Fig. 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 indirect GARCH (fourth row) functions. The adaptive and absolute forecasts depend strongly on the estimation period used and differ considerably from the forecasts obtained when implementing the asymmetric slope function and the indirect GARCH. On the other hand, the latter two forecasts are similar to each other and do not depend on the estimation period in which they are based.

The CAViaR procedure has been extended in several directions. First, Gourieroux 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 that 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 $\ell(\cdot)$ to the filtered estimators of time-varying quantiles. Third, several authors propose extending the Indirect GARCH in Eq. (5) by allowing for asymmetries. For example, Gerlach, Chen, and Chan (2011) and Yu, Li, and Jin (2010) specify the equation of the VaR in Eq. (5) as a TGARCH model and propose EM-type and Monte Carlo Markov Chain (MCMC) estimators, respectively. The Monte Carlo results of Gerlach et al. (2011) suggest that there are no differences between using the MCMC procedure or the parameter estimator of Koenker and Basset (1978) to forecast the VaR. Furthermore, the asymmetric extension proposed does not seem to provide more accurate forecasts. More recently, Chen, Gerlach, Hwang, and McAleer (2012) also proposed a Threshold-CAViaR model for intraday 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 Eq. (4) to be a function of past returns. Finally, several authors propose the introduction of explanatory variables into the CAViaR equation. For example, Jeon and Taylor (2013) introduce the implied volatility as a regressor, while Rubia and Sanchis-Marco (2013) introduce liquidity and trading activity.

Taylor (2008a) proposes 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.

The question of whether these extensions significantly improve VaR forecasts should be analyzed in future research.

3.2.3. Nonparametric estimators of predictive distributions

The CAViaR procedure represents the evolution of quantiles directly, assuming a particular specification of this evolution. Alternatively, several authors have proposed nonparametric estimators of the returns distribution, to avoid the effects of potential misspecification. These nonparametric methods are more complicated computationally, but can result in inferential gains when the assumptions of the parametric models are wrong. Next, we describe some of the nonparametric procedures that have been proposed for forecasting the 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 a 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, for obtaining predictive distributions of daily returns over horizons of one to several trading days. Their model is estimated using MCMC methods. Ferraty and Quintela-del-Río (in press) also 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.

These 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 of the use of EVT in finance. When applying EVT, one of three different approaches can be adopted, of which two are parametric and the third is nonparametric.

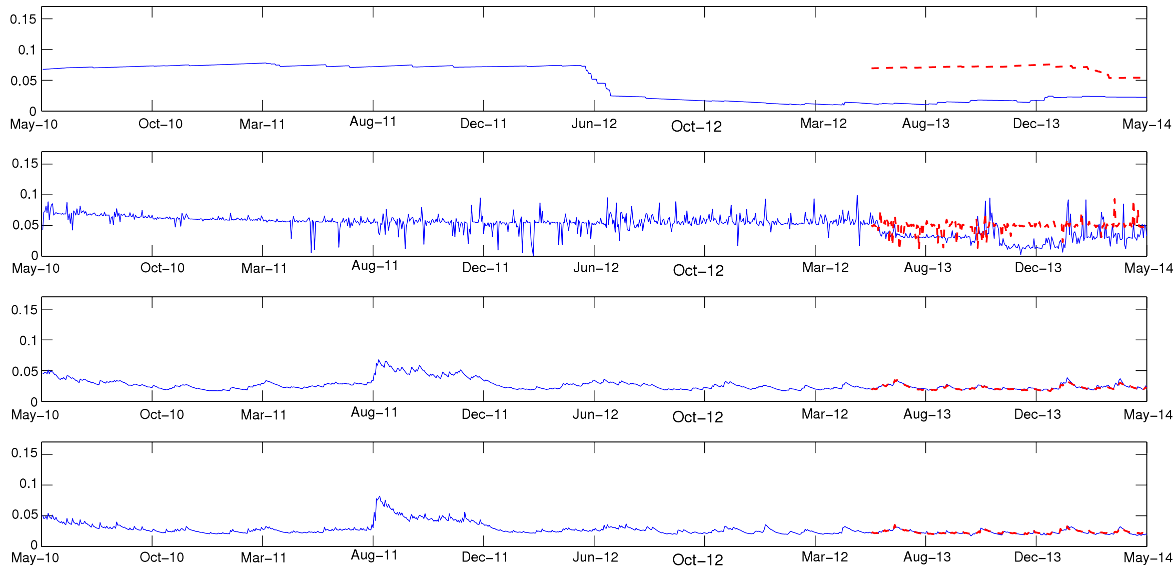

Fig. 3. Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line), and (ii) $T = 1970$ and $H = 250$ (discontinuous line), obtained using CAViaR with the adaptive function (first panel), the absolute value function (second panel), the asymmetric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

The first parametric approach is the *Block Maxima* (BM), which 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 that belong to the Generalized Extreme Value (GEV) distribution. The GEV distribution depends on a parameter ξ, 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 ξ, known as the tail index. The shape parameter can be estimated by maximum likelihood (ML), and the VaR is estimated based on the distributions of extremes; see for example Longin (2000). The BM approach has been implemented for forecasting VaR by Diebold, Schuermann, and Stroughair (2000), for example.

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

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

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 estimate the shape parameter nonparametrically without assuming any particular model for the tail. There are several estimators that can be used to accomplish this task, with the most popular being the Hill estimator, which only works in the Fréchet case ($\xi > 0$); see Hill (1975). The corresponding VaR forecast is given by

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

where $\widehat{\xi} = \frac{1}{k}\sum_{j=1}^{k}\left(\log(R_{(j)}) - \log\left(R_{(j+1)}\right)\right)$; see Hill (2010) and Peng and Qi (2006b) 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, Koedijk, Kool, and Palm (2001) propose a robust small sample bias-corrected estimator of ξ, based on a linear regression of a set of Hill estimators obtained using different numbers of observations in the tail. Gomes, de Haan, and Rodrigues (2008), Gomes, Figueiredo, Rodrigues, and Miranda (2012), Gomes, Matins, and Neves (2007), and Gomes and Pestana (2007) propose alternative estimators of ξ that minimize the variance and reduce the bias. Gourieroux and Jasiak (2010a) point out that the accuracy of the Hill estimator and its extensions is rather poor, due to the difficulty of 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 quantile estimates that are more stable than those from 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 for the risk estimates. If it is too low, the VaR forecasts will be biased and the asymptotic limit theorems do not apply. Conversely, if u is too large, the VaR forecasts will have large standard deviations due to the limited numbers of observations over the threshold. Danielsson, de Haan, Peng, and de Vries (2001) and Ferreira, de Haan, and Peng (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. Gençay and Selçuk (2004) determine u using a combination of the mean excess function and the Hill plots. Chavez-Demoulin, Embrechts, and Sardy (2014) propose the inclusion of a sensitivity analysis across several threshold values for a full POT application. As another alternative, Li, Peng, and Yan (2010) propose choosing u in such a way as to reduce the bias of the tail index estimator; see Scarrott and MacDonald (2012) for an excellent review of alternative procedures for choosing or estimating the threshold.

In addition to the problem of choosing an adequate threshold, Embrechts (2009) points out several caveats of EVT, and in particular 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. As a consequence, EVT needs to be modified, which complicates 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, Davidson, and McNeil (2005) and the ACD-POT model proposed by Herrera and Schipp (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 was proposed by Santos and Alves (2012a). Very recently, Chavez-Demoulin et al. (2014) propose a non-parametric extension that can fit the 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 the quantile of the distribution of returns directly. Alternatively, the VaR can be forecast assuming particular specifications for the conditional distribution of returns. Consider the following model of returns

$$R_t = \mu_t + \varepsilon_t \sigma_t, \tag{8}$$

where μ_t and σ_t are the conditional mean and standard deviation respectively, and ε_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

$$\mathrm{VaR}_t^\alpha = \mu_t + q_\varepsilon^\alpha \sigma_t, \tag{9}$$

where q_ε^α is the α quantile of the distribution of ε_t. Next, we describe alternative specifications proposed in the literature for representing μ_t, σ_t and q_ε^α.

3.3.1. Conditional mean

In order to forecast the VaR in Eq. (9), one needs to estimate the conditional mean, μ_t. Given that the linear dependence of returns is usually very weak, most authors have represented it using AR(1) or MA(1) models; see for example Ardia and Hoogerheide (2014), Bali and Theodossiou (2007), Halbleib and Pohlmeier (2012), Kuester et al. (2006) and McNeil and Frey (2000).

Consider again the series of S&P500 returns. The sample autocorrelations of returns plotted in Fig. 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 represented adequately by a white noise in each of the two estimation subperiods; 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

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

where $\beta_0 > 0$, $\beta_1 \geq 0$, $\beta_2 \geq 0$, and $(\beta_1 + \beta_2) < 1$; see Bali and Theodossiou (2007), Christoffersen and Goncalves (2005), Engle (2003), Giannopoulos and Tunaru (2005), Kuester et al. (2006) and McNeil and Frey (2000), 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 Eq. (10) with $\beta_0 = 0$ and $\beta_1 + \beta_2 = 1$. The Basel accords recommend the use of the Gaussian IGARCH model with $\beta_2 = 0.95$; see for example Gourieroux and Jasiak (2010a).

The GARCH(1,1) model in Eq. (10) can be extended to cope with the asymmetric response of the 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 to have an adequate performance when forecasting the VaR; see for example Bali and Theodosiou (2007). The volatility in an EGARCH (1,1) model is given by

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

Some authors, such as Giot and Laurent (2003) and Sajjad, Coakley, and Nankervis (2008), consider modelling the volatility by fitting the APARCH model of Ding, Granger, and Engle (1993). However, the results obtained from the APARCH and EGARCH models are usually very similar; see Rodríguez and Ruiz (2012). Furthermore, Dark (2010)

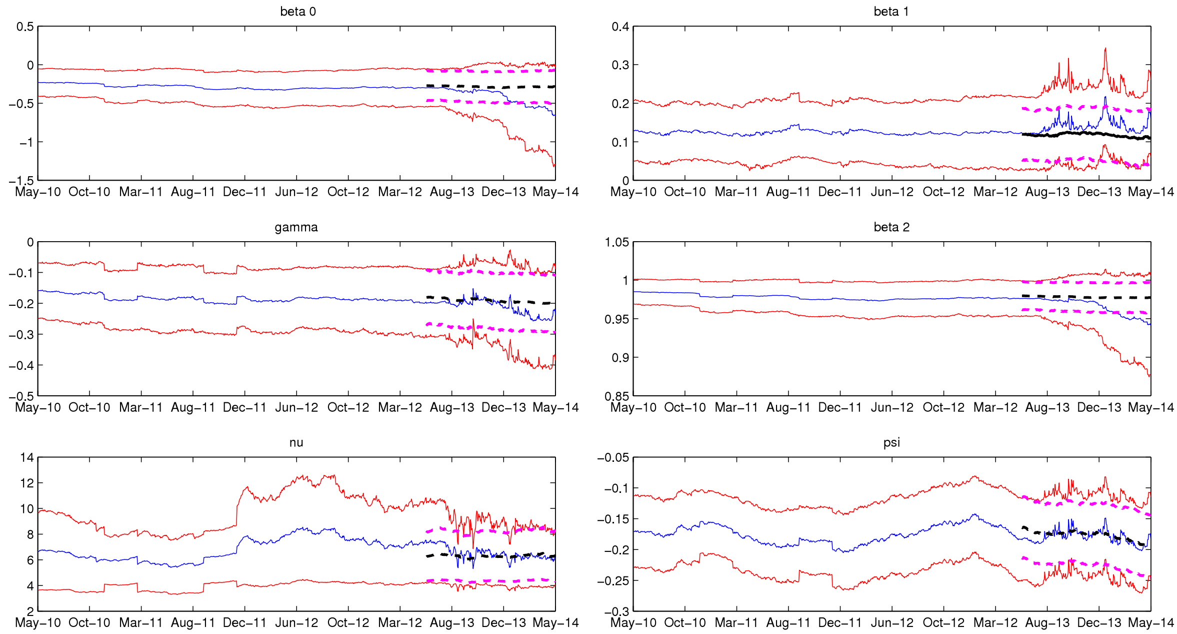

Fig. 4. Rolling window estimates and 95% confidence intervals of the EGARCH(1,1) parameters obtained by maximizing the Skewed-Student likelihood fitted to S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line), and (ii) $T = 1970$ and $H = 250$ (discontinuous line).

obtains estimates of the power parameter that are very close to one for a large range of stocks.

Consider again the S&P500 returns, which, as was 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 the Gaussian and Student distributions, as well as the Skewed-Student distribution of Hansen (1994). Fig. 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.[7] Note that the estimates of β_0, β_1, β_2 and the leverage parameter, γ, could be rather different depending on whether they are obtained using the first rolling window scheme, with $T = 1220$, or 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, Lu, and Huang (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, Lee, and Hung (2005) compare the VaR forecasts obtained when an asymmetric GARCH model is fitted to represent the conditional variance with those obtained when fitting the model proposed by Maheu and McCurdy (2004), which includes jumps. In an empirical exercise, they conclude that omitting the presence of jumps leads to VaR forecasts that are understated, and brings larger unexpected losses.

Some authors also suggest incorporating long-memory in the volatility when computing the VaR; see for example Caporin (2008), Dark (2010) and Degiannakis (2004). 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; for support of this result, see for example So and Yu (2006). 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 the use of GARCH-type models for representing the dynamic evolution of volatilities, several authors have also considered stochastic volatility (SV) models. Some applications of SV models to the estimation of VaR are provided by Chen, Gerlach, Lin, and Lee (2012), Fleming and Kirby (2003), González-Rivera, Lee, and Mishra (2004) and Lehar, Scheicher, and Schittenkopf (2002). None of these papers find important differences between the VaR forecasts obtained using GARCH and SV models. Furthermore, McAleer (2009) points out that the computational burden of SV can be quite severe, while econometric software packages do not seem to have incorporated SV algorithms yet.

Bao and Ullah (2004) analyse the biases incurred when VaR forecasts are 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 appropriate only when the

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

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 therefore is not very useful in empirical applications. Hartz, Mittnik, and Paolella (2006) propose a bias-correction method for improving the VaR forecasting ability of the GARCH model with normal errors. Recently, Gourieroux and Zakoïan (2013) proposed a correction of the $1/T$ asymptotic bias in the coverage probabilities within the context of a parametric model with a 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 for forecasting the VaR, Boucher et al. (2014) show that the specification bias can be very large, sometimes of the same order of magnitude as the VaR itself, and propose a general framework that can take into account the specification and estimation uncertainty by adjusting the VaR forecasts empirically based on the outcomes of the backtesting procedures described in the next section. Using simulated data, they show that dynamically adjusting for estimation bias improves the performance of every method. The bias depends strongly upon the VaR level; consequently, Boucher et al. (2014) recommend avoiding procedures that estimate the VaR using information on different quantiles, such as those used by Wang, Li, and He (2012), Xiao and Koenker (2009) and Yi, Feng, and Huang (2014), for example.

Finally, it is important to mention that, due to the availability of intradaily ultra-high frequency data (UHFD), there is a growing body of literature that proposes estimating the daily volatility using measures based on this type of observation; see Clements, Galvao, and Kim (2008), Coroneo and Veredas (2012), Fuertes and Olmo (2013) and Giot and Laurent (2004). Brownlees and Gallo (2010) forecast the daily VaR using the 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 relative to those based on GARCH models, with the simple daily range delivering very accurate forecasts. Furthermore, they show that the intradaily 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, González-Rivera, Maté, and Muñoz San Roque (2011) base their VaR forecast on the histogram of intra-daily observations. More recently, Louzis, Xanthopoulos-Sisinis, and Refenes (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 Eq. (9) is q_ε^α, the α% quantile of the distribution of standardized returns. There are three main alternatives for obtaining q_ε^α. First, q_ε^α is given when a particular distribution for ε_t is assumed. Second, one can implement the methods described in the previous subsection to compute the quantile of the returns standardized by the estimated conditional mean and variance. Third, the quantile of the distribution of ε_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 and Student distributions, and the Skewed-Student distribution of Hansen (1994); see for example Ardia and Hoogerheide (2014), Bali and Theodossiou (2007), Giot and Laurent (2003), Halbleib and Pohlmeier (2012), Kuester et al. (2006), Louzis et al. (2013), Pérignon and Smith (2010a) and Sajjad et al. (2008) for applications using these distributions.

Figs. 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 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 Fig. 4, have remarkably small effects on the VaR forecasts. This result is in accordance with the findings of González-Rivera, Lee, and Yoldas (2007), who analyse the implications of estimating the parameters of the RiskMetrics model under different loss functions and conclude that there are no important effects on the out-of-sample performances 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 has only a very minor effect on VaR forecasts. Second, observe that the VaR forecasts generated when the GARCH model is used to estimate the conditional variance are smoother than those obtained when the EGARCH model is used. The GARCH forecasts are more similar to the CAViaR (absolute and asymmetric) and EVT (GPD and Hill) forecasts. Third, Figs. 5 and 6 show that, regardless of the particular model used for the conditional variance, the risk is larger when the Student distribution is assumed for the conditional distribution of returns. When it is assumed to be the Skewed-Student distribution, the risk is forecast to be in between the risks under Normal and Student errors. We observe larger differences between the VaR forecasts obtained under different assumptions of the error distribution during the period from August to December 2011. The differences between the VaR forecasts during the last year of the out-of-sample period, when the risk was relatively small, are negligible.

An alternative leptokurtic and asymmetric distribution that has been considered in the context of VaR forecasting is the Skewed-Generalized-t (SGT) distribution proposed by Theodossiou (1998); see Bali, Mo, and Tang (2008), Bali and Theodossiou (2007) and Chen and Hung (2011) for applications of the SGT distribution to VaR forecasting. The SGT distribution has the attractive feature of encompassing most of the distributions that are usually assumed for

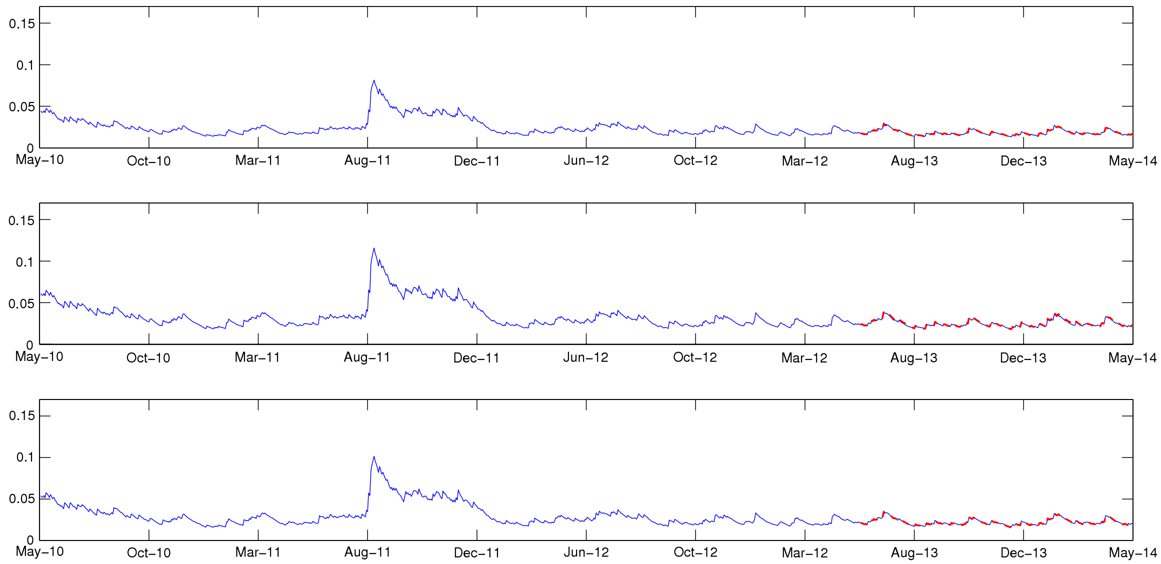

Fig. 5. Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line), and (ii) $T = 1970$ and $H = 250$ (discontinuous line), obtained after fitting a GARCH model assuming: (i) normal conditional returns (top panel); (ii) Student conditional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

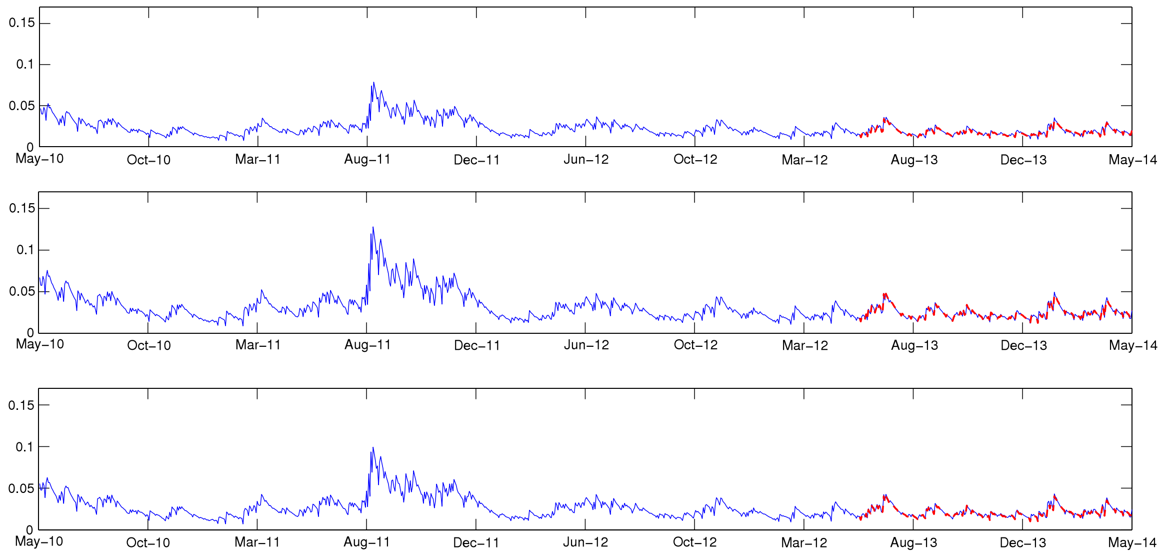

Fig. 6. Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line) and (ii) $T = 1970$ and $H = 250$ (discontinuous line) obtained after fitting the EGARCH model assuming: (i) normal conditional returns (top panel); (ii) Student-ν conditional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

standardized returns, such as the Gaussian, Generalized Error Distribution (GED), Student and Skewed-Student distributions, for example. However, our experience suggests that the maximization of the log-likelihood based on a SGT distribution is very complicated. Consequently, we will not consider this distribution further in this survey. Recently, Ergen (2015) also considered the Skewed-t distribution of Azzalini and Capitanio (2003).

The Laplace distribution, which is useful for performing quantile regression, has also been considered for representing the distribution of 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 the asymmetric two-sided Weibull distribution, which generalizes the asymmetric Laplace distribution proposed by Chen, Gerlach, and Lu (2012) by allowing different Weibull distributions for positive and negative returns. The estimation and forecasting are carried out using a Bayesian MCMC procedure. However, in their empirical application of estimating the VaRs of seven daily time series of financial returns, the results are comparable to those obtained when the asymmetric Student distribution is fitted. Given the additional computational effort that is involved when using Bayesian procedures, it does not seem worth the additional effort.

Several authors also propose modelling the distribution of ε_t by assuming mixtures of Normal or GED distributions; see for example Broda, Haas, Krause, Paolella, and Steude (2013) and Kuester et al. (2006).

Finally, time-varying skewness and kurtosis have also been introduced in the distribution of ε_t. Bali et al. (2008) and Dark (2010) consider SGT and Skewed-Student distributions, respectively, with time-varying parameters that depend on past information. Grigoletto and Lisi (2009) and Wilhelmsson (2009) propose volatility models that allow for time-varying skewness and kurtosis. More recently, Grigoletto and Lisi (2011) model the distribution of ε_t using 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 that allows for time-varying skewness and kurtosis and an asymmetric response of the volatility to positive and negative past returns. In the empirical application, they conclude that time-varying moments are not relevant for forecasting 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 that it does not take into account the evolution of the conditional variances that characterizes financial returns. To overcome this limitation, Hull and White (1998) propose applying HS to returns standardized using the volatility estimated after fitting a GARCH-type model. Very recently, Dupuis, Papageorgiou, and Rémillard (2015) propose a robust estimator of the volatility based on the weighted likelihood combined with HS for choosing the rolling window length and the smoothing parameter value optimally.

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 neither method is superior in every situation. If the errors have distributions that admit 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, the one-step estimator will be better.

The first panel of Fig. 7 plots the VaR forecasts, denoted as Conditional Historical Simulation (CHS), obtained using HS after applying the GARCH[8] model to the estimation of the conditional variance. First, we can see that the VaR forecasts do not depend on the particular split used to obtain them. Second, it is possible to see 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 ε_t, Danielsson and de Vries (2000) and McNeil and Frey (2000) propose to estimate q_ε^α in Eq. (9) by applying EVT to the standardized returns, which are iid if the conditional mean and variance are specified correctly; see Chan and Gray (2006) for a nice description of the conditional EVT (CEVT) and its application to the forecasting of the VaR of daily electricity prices. In particular, McNeil and Frey (2000) propose first filtering the returns from estimating a GARCH model, 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 when combined with a well-chosen threshold estimation, such as that of Gonzalo and Olmo (2004). However, the results obtained when implementing a GARCH model with conditionally Student t errors are not very different.

Recently, Mancini and Trojani (2011) proposed increasing the robustness of this procedure by estimating the GARCH parameters using the M-estimator of Mancini, Ronchetti, and Trojani (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 Schucany (2004), which the authors claim is less sensitive to the choice of the threshold than classical methods. Finally, resampling is carried out based on this distribution.

Finally, in the context of GARCH(1,1) and AR(1)-GARCH(1,1) models, Chan, Li, Peng, and Zhang (2013) and Chan, Peng, and Zhang (2012), 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 Fig. 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 Fig. 5. and depend on the split of the sample used for their estimation. The middle and bottom panels of Fig. 8 plot the VaR forecasts obtained using CEVT with the GPD and Hill procedures, respectively, for $k = 100$, as suggested by McNeil and Frey (2000). Note that the VaR forecasts obtained using the two procedures are rather similar, and are also 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 influence of the threshold on VaR forecasts, Fig. 9 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 with the threshold estimated using the procedures proposed by Gonzalo and Olmo (2004) (first row) and Gençay and Selçuk (2004) (second row). We can see 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

[8] Note that, as in any of the conditional procedures described later, the same methodology could also be applied to the returns standardized using any of the other procedures that are available for estimating the conditional variance.

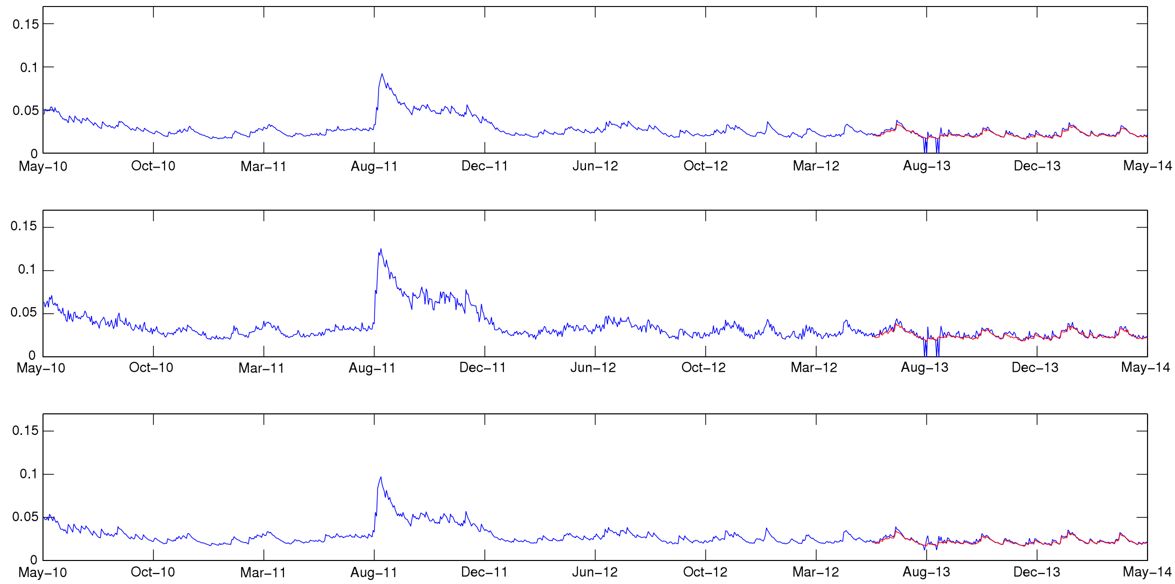

Fig. 7. Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line), and (ii) $T = 1970$ and $H = 250$ (discontinuous line), obtained using: (a) CHS (top panel); (b) FHS (middle panel); and (c) bootstrap (bottom panel).

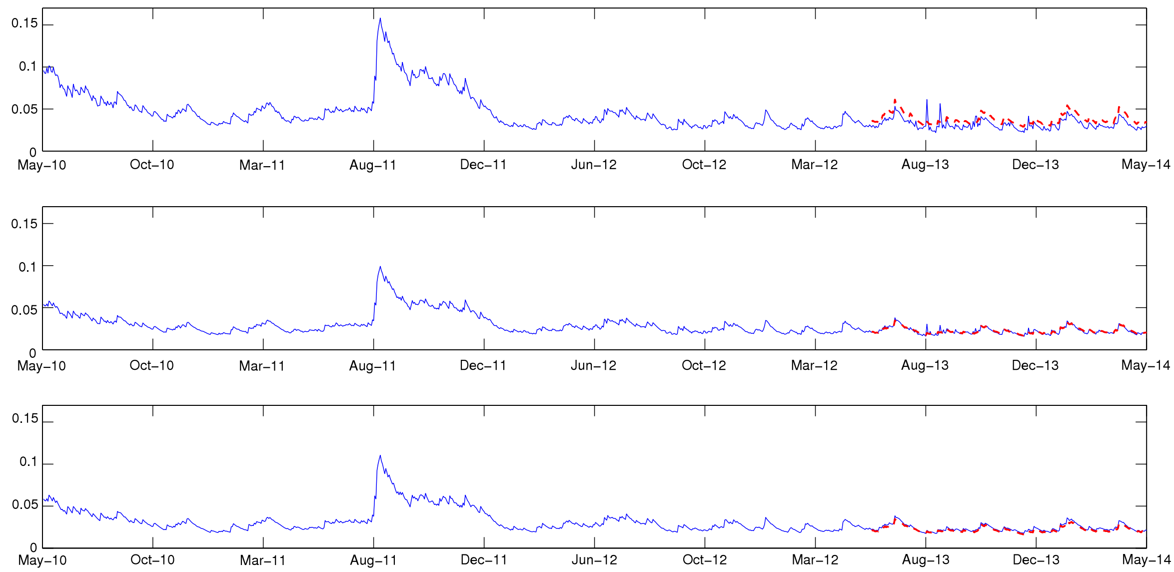

Fig. 8. Rolling window one-step-ahead 1% VaR forecasts of S&P500 returns based on (i) $T = 1220$ and $H = 1000$ (continuous line), and (ii) $T = 1970$ and $H = 250$ (discontinuous line), obtained using CEVT with (i) BM (first row), (ii) GDP (second row), and (iii) Hill estimator.

the threshold is chosen using the procedure proposed by Gençay and Selçuk (2004). Furthermore, the forecasts of the VaR plotted in Fig. 9 for $k = 100$ are in between those obtained when implementing the procedures of Gençay and Selçuk (2004) and Gonzalo and Olmo (2004). It is also 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 of Gonzalo and Olmo (2004), the VaR forecasts vary depending on the particular split used for the estimation.

Filtered historical simulation. The quantile q_ε^α 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 for VaR forecasting specifically. In particular, Barone-Adesi, Giannopoulos, and Vosper (1999, 2002) propose a bootstrap method known as filtered historical simulation (FHS), which is based on the idea of using random draws with replacement from the standardized residuals and does not incorporate parameter uncertainty; see Engle (2003) and Pritsker (2006) for implementations. More recently, Pascual, Ruiz, and Romo

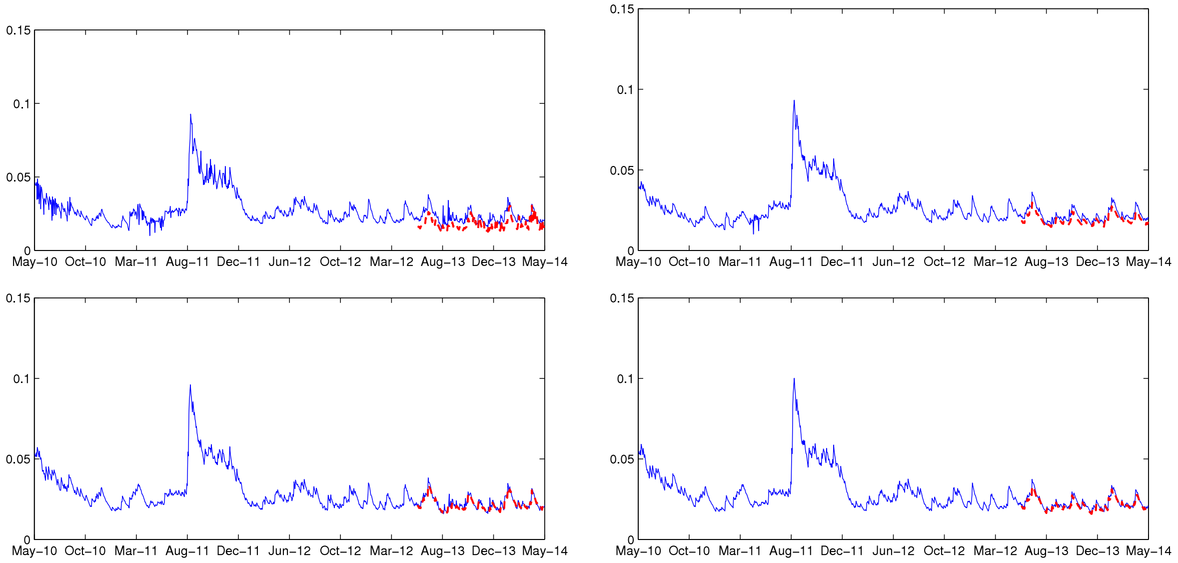

Fig. 9. One-step-ahead 1% VaR forecasts of S&P500 returns obtained using CEVT with the GPD (first column) and Hill (second column) procedures, with the threshold chosen as proposed by Gonzalo and Olmo (2004) (first row), or Gençay and Selçuk (2004) (second row). Continuous and discontinuous lines represent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

(2006) propose a bootstrap procedure that allows the parameter uncertainty to be incorporated. As we will see in more detail below, bootstrap procedures have the advantage that they allow the construction of confidence intervals for the estimated VaR.

The second and third rows of Fig. 7 plot the VaR forecasts obtained after implementing FHS without and with parameter uncertainty, respectively. First of all, observe that the 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 the 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 to 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 for estimating 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 that results in a data-dependent decay factor which remains constant in the forecasting period, and one that adapts automatically to changes in the stock price dynamics, adding flexibility to the first decay factor. In addition, Fan and Gu (2003) also propose a symmetric nonparametric estimation approach for estimating 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 modelling the standardized returns using the Generalized Hyperbolic distribution. According to these authors, this nonparametric adaptive methodology has the attractive feature of being able to estimate the homogeneous volatility over short time periods, thus reflecting sudden changes in the volatility.

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

3.4. Combining VaR forecasts

Recently, several authors have proposed the forecasting of VaR by combining several procedures. For example, McAleer, Jiménez-Martín, and Pérez-Amaral (2013a,b) consider mixing alternative risk models, while McAleer, Jiménez-Martín, and Pérez-Amaral (2010) provide a procedure for choosing one risk model at the beginning of the forecast period that is then 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 that is based on implied volatilities which merge information from the historical time series of returns with 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, they 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 for failure of VaR forecasts 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) were the first to point out the need to consider the estimation uncertainty associated with $\widehat{\mathrm{VaR}}_t^{HS}$ forecasts when the returns are assumed to be Gaussian; see also Dowd (2006) for a discussion of the uncertainty of a related VaR measure based on order statistics, and Gourieroux and Jasiak (2010a) for the asymptotic distribution of $\widehat{\mathrm{VaR}}_t^{HS}$. However, as far as we know, there are no analytical measures of the uncertainty associated with the WHS procedure. Furthermore, there are no procedures available for computing intervals for CAViaR forecasts.

There are also proposals for the construction of forecast intervals for VaR forecasts obtained using the Hill estimator. For example, Peng and Qi (2006a) carry out Monte Carlo experiments to compare the performances of three alternative intervals, namely the normal approximation, the likelihood ratio and data tilting, and conclude that data tilting performs best. More recently, Chavez-Demoulin et al. (2014) provide forecast intervals for the Bayesian nonparametric POT procedure that they propose. However, there are no proposals for the construction of forecast intervals for VaR values obtained using BM.

In the context of nonparametric procedures for computing the quantile of the distribution of returns directly, 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, shaped by the data automatically, can be constructed based on the 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 the parameter uncertainty into account using the asymptotic properties of the residual empirical process, when the VaR forecasts 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 using either the two-step or one-step method. 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, Deng, Peng, and Xia (2007). However, this distribution is unconditional. Gong, Zhouping, and Peng (2010) subsequently propose the employment of the 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 can be 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 suggested 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 for ε_t. They propose two alternative methods for constructing VaR forecast intervals. The first method is based on the asymptotic normality of the VaR estimator. Alternatively, they propose that confidence intervals be constructed using 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 parameter uncertainty.

Several authors propose the implementation of bootstrap procedures in order to produce forecast intervals for the VaR obtained using FHS; see Christoffersen and Goncalves (2005) and Dowd (2007), 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 account for the conditional variance dynamics properly imply forecast intervals with coverages that are close to the nominal. Bootstrap procedures are also implemented by Hartz et al. (2006). In the context of FHS, and based on the asymptotic behaviour of the GARCH residuals, Gao and Song (2008) derive an expression for the asymptotic variance of $\widehat{\text{VaR}}_t^{\text{FHS}}$ and use it to construct forecast intervals based on the Normal approximation. Using simulated data, they show that the bootstrap method of Christoffersen and Goncalves (2005) performs better than the asymptotic approximation that they propose; see also Mancini and Trojani (2011) for an application of the procedure proposed by Christoffersen and Goncalves (2005) to the computation of 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 the corresponding standard deviations. Finally, Li, Gong, and Peng (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, in order to ensure that financial and insurance institutions have adequate capital to cope with large unexpected losses. In light of the vast number of alternative VaR forecasting procedures available, model diagnosis is of critical importance for internal risk management. The Basel accords asses the accuracy of VaR estimates by developing 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 of backtesting procedures. In this section, we summarize the most popular backtesting procedures, which are closely related to 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 led to the proposal of testing for several VaR levels jointly. Finally, when several alternative estimators of the VaR are adequate according to the backtesting tests, one may want to rank them. We also describe tests that are designed for comparing and choosing 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 $I_t^\alpha = 1\left(R_t < -\text{VaR}_t^\alpha\right), t = T+1, \ldots, T+H$, where $1(\cdot)$ is the indicator function. A necessary condition for an optimal VaR forecast is that it is conditionally unbiased, so that

$$\underset{t-1}{E}\left[I_t^\alpha\right] = \alpha. \tag{12}$$

Most traditional backtesting procedures are based on testing some of the implications of this condition. The most popular backtesting procedure, known as the unconditional coverage (UC) test, was proposed by Kupiec (1995) and tests the null hypothesis $H_0 : E\left[I_t^\alpha\right] = \alpha$. However, this is not the hypothesis of interest in Eq. (12), and, as a consequence, the UC test ignores the 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, that information could be used to construct better VaR forecasts. 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 de la Pena, Rivera, and Ruiz-Mata (2007), Pérignon and Smith (2008) and Pritsker (2006), among others. de la Pena et al. (2007) suggest swapping the null and alternative hypotheses in order to reduce the probability of choosing the wrong model.

Alternatively, Christoffersen (1998) proposes the very influential conditional coverage (CC) test, where the null hypothesis is $H_0 : E\left[I_t^\alpha | I_{t-1}^\alpha\right] = \alpha$. The likelihood ratio (LR) statistic for CC, LR_{cc}, is given by the addition of the unconditional statistic and the independence statistic based on testing whether I_t^α 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 that takes 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 applied to the VaR forecasts obtained from 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 which particular period is considered. The procedures that are not rejected in either period 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. All other models 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 tests of Christoffersen (1998) and Kupiec (1995) are still the most popular ones both among practitioners and in the academic literature; see, for example, Brownlees and Gallo (2010), Chan and Gray (2006), Chen, Gerlach, Hwang et al. (2012), Dias (2013), Grigoletto and Lisi (2009,

Table 2
Summary of VaR backtesting procedures.

Procedures			Mnemonics	Selected papers	Extensions
Adequacy					
Single level	No. of Violations	Unconditional coverage	UC	Kupiec (1995)	Param. uncert., Escanciano and Olmo (2010)
		Conditional coverage	CC	Christoffersen (1998)	Bootstrap, Escanciano and Olmo (2011)
		Likelihood ratio	LR_{CC}	Christoffersen (1998)	Logit regressions, Patton (2006)
		Portmanteau	LB	Berkowitz et al. (2011)	
		Dynamic Quantile	DQ	Engle and Manganelli (2004)	
	Magnitude violations	Tail risk	TR	Wong (2010)	
		Risk map		Colletaz et al. (2013)	
	Durations			Christoffersen and Pelletier (2004)	
	Quantile regression		VQR	Gaglianone et al. (2011)	
	Forecast rationality			Hoogerheide et al. (2012)	
Multiple levels	Whole density			Berkowitz (2001)	
	Violations			Hurlin and Tokpavi (2007)	
	Omnibus			Escanciano and Velasco (2010)	
	Pearson-type			Leccadito et al. (2014)	
Ranking					
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. (2006)	Without benchmark, Sener et al. (2012)

Table 3
Comparison of backtesting results obtained using alternative procedures in two different periods.

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

Notes: The table shows p-values from backtesting tests of S&P500 one-step-ahead 1% VaR forecasts obtained using alternative procedures during two different periods: (1) 28th May 2010–19th May 2014 ($H = 1000$); and (2) 22nd May 2013–19th May 2014 ($H = 250$). Models rejected at the 10% significance level are shown in bold.

2011), Halbleib and Pohlmeier (2012), Liu and Tse (2015), Pérignon and Smith (2010b), Rubia and Sanchis-Marco (2013), and Zikovic and Aktan (2011). Their popularity is due, first, to their simplicity, and second, because they are incorporated implicitly in the Basel accords for determining market risk capital requirements.

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

$$E\left[\delta_t^{\alpha} \otimes X_{t-1}\right] = 0, \tag{13}$$

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

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

where r_j is the order j sample autocorrelation of δ_t^α. Under the null, $\text{LB}(\tau)$ is asymptotically $\chi^2_{(\tau)}$. 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 $\text{LB}(\tau)$ test in Eq. (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 procedures rejected increases relative 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 for decisions about the adequacy of VaR forecasts. In the very tranquil period (period 2), the results of the LR_{cc} and $\text{LB}(\tau)$ tests are identical. Also note that the results for the two values of τ 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 Eq. (13) when X_{t-1} includes lags of VaR_t^α and of δ_t^α. Berkowitz et al. (2011) show that the DQ test appears to be the best backtest for the 1% VaR, with all of the 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 3 reports the p-values of the DQ test.[9] In period 1, the results are very similar to those of LR_{cc}, while the number of models rejected in period 2 is even smaller than when implementing the LR_{cc} test. Note that in the very tranquil period (period 2), only the WHS model is rejected. The models that 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 both tests have very low powers, regardless of how the finite sample distribution is approximated. Furthermore, in their Monte Carlo experiments, Gaglianone, Lima, Linton, and Smith (2011) show that the asymptotic critical values of the LM-type tests are only accurate if the sample size corresponds to at least four years of daily observations, i.e., $H \geq 1000$.

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

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

The TR statistic tells risk managers the size of the aggregate tail losses that 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 the results of Berkowitz and O'Brien (2002) and Kerkhof and Melenberg (2004) and transforming the non-normal density into a normal one, then applying a saddlepoint analysis to the transformed variable. However, this transformation complicates the empirical application of the TR statistic. Alternatively, Colletaz, Hurlin, and Pérignon (2013) propose the Risk Map as a new method of validating risk models which accounts for the number and magnitude of extreme losses jointly and summarizes all information about the performance of a risk model graphically. The Risk Map boils down to the intuition that a large loss not only exceeds the regular VaR with level α, but is also likely to exceed a VaR defined with a much lower probability. In order to test a risk model, Colletaz et al. (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, Colletaz, Hurlin, and Tokpavi (2011) for further developments of duration tests in the context of GMM, and Pelletier and Wei (in press) for a very recent proposal. Santos and Alves (2012b) also propose an independence test based on durations. Very recently, Ziggel, Berens, Wei, and Wied (2014) proposed 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

9 The computation of the DQ test has been carried out by implementing the Matlab code of Simone Manganelli, available at http://www.simonemandanelli.org/Simone/Research.html.

enters the joint test of conditional coverage. It is based on iid Bernoulli random variables and, according to the authors, is very intuitive and easy to implement.

The tests described above are based on the LM principle and have a low finite sample power against a variety of model misspecifications; see Christoffersen, Hahn, and Inoue (2001) and Gourieroux and Jasiak (2010a), who point out that the count of exceedances may be misleading as an instrument for VaR control. The problem arises because the binary variables, δ_t^α, are constructed to represent rare events, meaning that, in finite samples, there could be few extreme observations, leading to a lack of the information needed to reject a misspecified model. LM tests sacrifice too much information, implying a reduced 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 specified correctly:

$$R_t = \beta_0(U_t) + \beta_1(U_t)\mathrm{VaR}_t, \tag{16}$$

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:

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

The corresponding test, denoted VaR quantile regression (VQR), uses more information to reject a misspecified model, which makes it able 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 of the 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 ten-step-ahead VaRs are not straightforward.

Finally, Hoogerheide, Ravazzolo, and van Dijk (2012) propose an extension of the forecast rationality and optimally 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 most adequate for forecasting the VaR depend on the particular quantile of the distribution of returns that is being forecast; see, for example, Chen, Gerlach, Lin et al. (2012), Gaglianone et al. (2011), Giacomini and Komunjer (2005), Gerlach et al. (2011), and Kuester et al. (2006). 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. However, a weakness of this test is that models with superior density forecasts will not necessarily meet the requirements of risk managers, who are concerned mainly with forecasting the tails. Consequently, Hurlin and Tokpavi (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 Eq. (14) to the multivariate case by considering the vector of centered hits given by $\delta_t = \left(\delta_t^{\alpha_1}, \ldots, \delta_t^{\alpha_m}\right)'$. The joint conditional accuracy of VaR forecasts implies that $\mathrm{Cov}(\delta_t, \delta_{t-j}) = 0$ for $j \neq 0$. Consequently, Hurlin and Topkavi (2007) propose the following statistic:

$$Q_m(J) = T \sum_{j=1}^{J} \left(\mathrm{vec}\widehat{R}_j\right)' \left(\widehat{R}_0^{-1} \otimes \widehat{R}_0^{-1}\right) \left(\mathrm{vec}\widehat{R}_j\right), \tag{18}$$

where $\widehat{R}_j = D^{1/2}\widehat{C}_j D^{1/2}$, with $D = I_m \odot \widehat{C}_0$, $\odot$ being the element to element multiplication, and $\widehat{C}_j = \frac{1}{H}\sum_{t=1}^{H} \delta_t \delta_{t-j}'$; see Sajjad et al. (2008) for an empirical application of this test. Herwartz (2009) shows that the test in Eq. (18) is oversized when using the asymptotic distribution. By using Monte Carlo critical values, the test in Eq. (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 accompany the failures of the VaR model which Herwartz (2009) shows remain undetected when implementing the statistic in Eq. (18). Furthermore, Leccadito, Boffelli, and Urga (2014) point out that if the coverage rates considered in the test are very close to each other, the matrix $\widehat{R}_0$ is singular, and, as a consequence, the $Q_m(J)$ test cannot be calculated. Gourieroux and Jasiak (2008) also consider multi-level tests.

Alternatively, Pérignon and Smith (2008) propose a multilevel test that extends the unconditional test of Kupiec (1995). However, the authors present only a graphical analysis, rather than a formal testing procedure. Alternatively, Escanciano and Velasco (2010) propose omnibus specification tests that extend the unconditional, conditional and independence tests of Christoffersen (1998) and Kupiec (1995) to test a possible continuum of quantiles. The asymptotic critical values are approximated by subsampling, which accounts for the 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 to those of Hurlin and Tokpavi (2007) and Pérignon and Smith (2008), but do not provide a 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 provides accurate VaR forecasts. However, often, if several accurate estimators are available, one also wants to decide which of the estimators is best. With this goal, Lopez (1999) proposes the selection of the procedure that minimizes $C(m) = \sum_{t=T+1}^{T+H} C_t^{(m)}$, with

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

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, Baronyan, and Mengütürk (2012) for descriptions of alternative functions. Angelidis and Degiannakis (2007) and Sarma, Thomas, and Shah (2003) use the regulatory loss function (RLF), in which $f(R_t, \mathrm{VaR}_t^{(m)}) = \left(R_t - \mathrm{VaR}_t^{(m)}\right)^2$ and $g\left(R_t, \mathrm{VaR}_t^{(m)}\right) = 0$. More recently, Sener et al. (2012) propose a loss function that penalizes 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) proposes using a consistent-scoring Generalized Piecewise Linear (GPL) family of functions. A function is consistent for the α quantile if and only if it is of the following form:

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

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

Table 4 reports the values of $C(m)$ defined with the 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 of the models have similar results, meaning that the identification of a "best" model is not obvious and may not even be possible.

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

$$C_t^{(m)} = \left[\alpha - 1\left(R_t < \mathrm{VaR}_t^{\alpha}\right)\right]\left[R_t - \mathrm{VaR}_t^{\alpha}\right]. \tag{19}$$

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 seem to be slightly higher in periods 1 and 2.

Giacomini and Komunjer (2005) also propose a comparison of two VaR forecasts with 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 determine which of two models is the superior, Sarma et al. (2003) propose testing $H_0: \{\theta = 0\}$ against $H_1: \{\theta < 0\}$, where θ is the median of the distribution of the loss differential between procedures i and 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

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

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 Eq. (20) can be obtained as the t-statistic of the regression of z_t on a constant using the Newey and West (1987) heteroscedasticity and autocorrelation consistent standard errors.

An alternative to the test in Eq. (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

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

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 $h_t = \left(1, \Delta C_t^{(m)}\right)'$. Under the null hypothesis of equal conditional predictive ability, CPA is asymptotically distributed as a χ_q^2.

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

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

where $\widehat{\omega}_m^2$ is a consistent estimator of $\omega_m^2 = \mathrm{var}\left(H^{1/2}\bar{z}_m\right)$ and $\bar{z}_m = H^{-1}\sum_{t=1}^{H}(C_t^{(0)} - C_t^{(m)})$. The estimation of ω_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 using the block selection algorithm proposed by Politis and White (2004, 2009). Recently, Sener et al. (2012) proposed a predictive ability test that does not require a benchmark model, thus allowing for the simultaneous comparison of several procedures.

Table 4
Comparison of VaR forecasts obtained by alternative procedures.

	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
CAViaR(1)	–	–	–	–	0.02	0.17	1.00	0.00
CAViaR(2)	2.22	0.54	0.47	**0.00**	–	–	–	–
CAViaR(3)	9.24	0.32	0.61	0.85	**0.00**	0.06	0.93	0.99
GARCH-*N*	–	–	–	–	0.35	0.06	0.36	1.00
GARCH-*t*	–	–	–	–	0.35	0.06	1.00	**0.00**
GARCH-*S*	4.42	0.33	0.43	0.65	0.03	0.06	0.89	0.95
EGARCH-*N*	8.45	0.33	**0.06**	0.70	1.29	0.06	0.65	0.89
EGARCH-*t*	**0.66**	0.33	**0.04**	**0.00**	0.23	0.07	0.72	**0.01**
EGARCH-*S*	2.25	**0.31**	**0.02**	0.99	0.76	0.07	0.63	0.41
CHS	5.56	0.33	0.38	0.75	0.07	0.06	0.78	0.95
CEVT-BM	2.77	0.45	0.98	**0.00**	0.08	0.10	1.00	**0.00**
CEVT-GPD	4.70	0.33	0.48	0.74	0.01	0.06	0.91	0.71
CEVT-Hill	–	–	–	–	0.02	0.06	0.79	0.72
FHS	–	–	–	–	0.37	0.37	1.00	**0.00**
Bootstrap	5.67	0.33	0.65	0.89	0.04	0.06	0.83	0.98

Notes: 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. Models that are not rejected as inferior at the 10% significance level are shown in bold.

Table 4 reports the *p*-values of the SPA test when, for each out-of-sample period, each model is selected as a benchmark in turn against all of the others. We can see that, in period 1, when the RLF function is used, no individual model is rejected as inferior to 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. Thus, we can conclude that the ranking of the models depends strongly on the period considered and the function used for it.

5.3. *A final remark on backtesting*

The literature on the forecasting and backtesting of VaR largely assumes that the appropriate data are used. However, Frésard, Pérignon, and Wilhelmsson (2011), using information from the annual reports of the 200 largest US and international commercial banks, document that a large fraction of them boost the performances of their models artificially 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, fewer than 6% of the largest commercial banks in the world evaluated their VaR models using the appropriate uncontaminated data. They also show that all of the 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 only 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 therefore significantly reduced regulatory capital.

6. VaR forecasting in practice

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

Although a myriad of procedures are currently available for forecasting and testing VaR, no consensus has been reached as to which procedures are best. As the illustration above shows, for a given returns series, the ranking of VaR forecast procedures depends 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 performances of different procedures. She concludes that VaR is always estimated wrongly if we do not take crisis and non-crisis periods into account, and that market capitalization has a positive effect on the estimation of VaR. There are many papers comparing the performances of alternative VaR forecasts, with very mixed results; see, for example, Angelidis and Degiannakis (2007). In this section, we focus on those that consider the performances 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, the empirical evidence presented by Berkowitz and O'Brien (2002), Pérignon and Smith (2008, 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 et al. (2012). This alternation of over- and underestimation of market risk levels may be due, at least to some extent, to the fact that, typically, VaR measures are calibrated using historical data. Following a period of calm in financial markets, the VaR estimates and the accompanying CR may decline to low levels, but might then underestimate the risk during a future period of stress.

With respect to the comparison of VaR forecasts obtained using different procedures, Chen, Gerlach, Lin et al. (2012) provide an empirical comparison of a range of parametric models, combined with four error distributions, that are implemented on four Asia-Pacific stock markets. They conclude that GARCH models always outperform stochastic volatility models when they are estimated using MCMC methods. Furthermore, asymmetric models were favoured in the period before the crisis, while models with Skewed-Student errors were ranked best during and after the crisis. In this latter period, all of the 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 determine that models with Skewed-Student distributions can do a good job during the crisis period. They also show that the CEVT procedure can perform reasonably well.

Recently, Boucher et al. (2014) conclude that the RiskMetrics and GARCH-based models are among the preferred ones. However, many authors have concluded that methods that assume specific densities may yield poor 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 applied to the 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, Weitao, Chen, and Zhou (2010) conclude that EVT is more appropriate than HS for forecasting the VaR of the Yuan. On the other hand, other authors, such as Sener et al. (2012), for example, rank EVT as having the worst performance in more extensive analyses of series and procedures. Sener et al. (2012) conclude that VaR forecasts based on asymmetric CAViaR and EGARCH models perform the best. The preference for CAViaR is also established by Chen, Gerlach, Hwang et al. (2012) and Chen and Lu (2012). Tolikas (2014) also shows that when EVT is based on the GEV distribution, the corresponding VaR forecasts underestimate the risk. Alternatively, he proposes using the Generalized Logistic distribution for the extremes, and finds that it performs better during the crisis period. Chen and Gerlach (2013) also conclude that the differences between models are dominated by the error distribution.

7. Conclusions

In this paper, we review recent contributions to the forecasting and backtesting of the VaR measure of risk. The obvious benefit of VaR is that it is easy and intuitive for non-specialists to understand. 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 for obtaining VaR forecasts, the results of a particular test could vary depending on the number of out-of-sample observations and the particular period being analyzed. There is no single procedure that clearly outperforms the others, and, with the exception of the EGARCH model with Skewed-Student errors, all of them are rejected by at least one test in at least one out-of-sample period. It seems that rather simple forecasts based on modelling the evolution of the conditional variance using asymmetric GARCH-type models and asymmetric leptokurtic errors are among the most competitive ones. Including the time varying skewness and kurtosis of the distribution of standardized returns does not seem to improve the VaR forecasts. The introduction of bias correction, the uncertainty of VaR forecasts and their validation are important topics that still deserve more research in order to reach more conclusive results on the performances of alternative procedures.

It is important to note that although this paper is devoted to the surveying of VaR forecasting procedures, the recent Basel accords suggest that the expected shortfall (ES) be used 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, Brandtner (2013), Chen (2008), Chen, Gerlach, and Lu (2012), Komunjer (2007), Lönnbark (2013), Stoyanov, Rachev, and Fabozzi (2013), Wu and Xiao (2002), Xu (in press) and Yamai and Yoshiba (2005) for some recent references to the forecasting of ES. However, Gneiting (2011) shows that it is not possible to measure the adequacy of ES using a consistent scoring function.

Acknowledgments

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.

References

Acerbi, C., & Tasche, D. (2002). Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, *31*, 379–388.

Angelidis, T., Benos, A., & Degiannakis, S. (2007). A robust VaR model under different time periods and weighting schemes. *Review of Quantitative Finance and Accounting*, *28*, 187–201.

Angelidis, T., & Degiannakis, S. (2007). Backtesting VaR models: a two-stage procedure. *Journal of Risk Model Validation*, *1*(2), 1–22.

Ardia, D., & Hoogerheide, L. F. (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.

Arroyo, J., González-Rivera, G., Maté, C., & Muñoz San Roque, A. (2011). Smoothing methods for histogram-valued time series. An application to value-at-risk. *Statistical Analysis and Data Mining*, *4*(2), 216–228.

Asai, M., & McAleer, M. (2008). A portfolio index GARCH model. *International Journal of Forecasting*, *24*(3), 449–461.

Azzalini, A., & Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society, Series B*, *65*, 367–389.

Bali, T. G., Mo, H., & Tang, Y. (2008). The role of autoregressive conditional skewness and kurtosis in the estimation of VaR. *Journal of Banking and Finance*, *32*(2), 269–282.

Bali, T. G., & Theodossiou, P. (2007). A conditional-sgt-VaR approach with alternative GARCH models. *Annals of Operations Research*, *151*, 241–267.

Bao, Y., Lee, T.-H., & Saltoglu, B. (2006). Evaluating predictive performance of value-at-risk models in emerging markets: a reality check. *Journal of Forecasting*, *25*, 101–128.
Bao, Y., & Ullah, A. (2004). Bias of value-at-risk estimator. *Finance Research Letters*, *1*, 241–249.
Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (1999). VaR without correlations for portfolios of derivative securities. *Journal of Future Markets*, *19*(5), 583–602.
Barone-Adesi, G., Giannopoulos, K., & Vosper, L. (2002). Backtesting derivative portfolios with filtered historical simulation (FHS). *European Financial Management*, *8*, 31–58.
Berkowitz, J. (2001). Testing density forecast, with applications to risk management. *Journal of Business and Economic Statistics*, *19*, 465–474.
Berkowitz, J., Christoffersen, P., & Pelletier, D. (2011). Evaluating Value-at-Risk models with desk-level data. *Management Science*, *57*(2), 2213–2227.
Berkowitz, J., & O'Brien, J. (2002). How accurate are value-at-risk models at commercial banks? *Journal of Finance*, *57*, 1093–1111.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*, 307–327.
Boucher, C. M., Daníelsson, J., Kouonchou, P. S., & Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking and Finance*, *44*, 72–92.
Brandtner, M. (2013). Conditional value-at-risk, spectral measures and (non)diversification in portfolio selection problems — a comparison with mean–variance analysis. *Journal of Banking and Finance*, *37*(12), 5526–5537.
Broda, S. A., Haas, M., Krause, J., Paolella, M. S., & Steude, S. C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, *172*, 292–306.
Brownlees, C. T., & Gallo, G. M. (2010). Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, *8*, 29–56.
Cai, Z., & Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, *147*, 120–130.
Campbell, S. D. (2007). A review of backtesting and backtesting procedures. *Journal of Risk*, *9*(2), 1–17.
Candelon, B., Colletaz, G., Hurlin, C., & Tokpavi, S. (2011). Backtesting value-at-risk: A GMM duration-based approach. *Journal of Financial Econometrics*, *9*(2), 314–343.
Caporin, M. (2008). Evaluating value-at-risk measures in the presence of long-memory conditional volatility. *Journal of Risk*, *10*, 79–110.
Chan, N. H., Deng, S.-J., Peng, L., & Xia, Z. (2007). Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. *Journal of Econometrics*, *137*(2), 556–576.
Chan, K. F., & Gray, P. (2006). Using extreme value theory to measure value-at-risk for daily electricity spot prices. *International Journal of Forecasting*, *22*(2), 283–300.
Chan, N. H., Li, D., Peng, L., & Zhang, R. (2013). Tail index of an AR(1) model with ARCH(1) errors. *Econometric Theory*, *29*(5), 920–940.
Chan, N. H., Peng, L., & Zhang, R. (2012). Interval estimation of the tail index of a GARCH model. *TEST*, *21*, 546–565.
Chavez-Demoulin, V., Davidson, A., & McNeil, A. (2005). A point process approach to value-at-risk estimation. *Quantitative Finance*, *5*(2), 227–234.
Chavez-Demoulin, V., Embrechts, P., & Sardy, S. (2014). Extreme-quantile tracking for financial time series. *Journal of Econometrics*, *181*(1), 44–52.
Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, *6*, 87–107.
Chen, Q., & Gerlach, R. H. (2013). The two-sided Weibull distribution and forecasting financial tails. *International Journal of Forecasting*, *29*, 527–540.
Chen, C. W. S., Gerlach, R., Hwang, B. B. K., & McAleer, M. (2012). Forecasting value-at-risk using nonlinear regression quantiles and the intra-day range. *International Journal of Forecasting*, *28*(3), 557–574.
Chen, C. W. S., Gerlach, R., Lin, E. M. H., & Lee, W. C. W. (2012). Bayesian forecasting for financial risk management, pre and post the global financial crisis. *Journal of Forecasting*, *31*, 661–687.
Chen, Q., Gerlach, R., & Lu, Z. (2012). Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution. *Computational Statistics and Data Analysis*, *56*, 3498–3516.
Chen, Y., Härdle, W., & Jeong, S.-O. (2008). Nonparametric risk management with generalized hyperbolic distributions. *Journal of the American Statistical Association*, *103*(483), 910–923.
Chen, W.-H., & Hung, J.-C. (2011). Skewness and leptokurtosis in GARCH-type VaR estimation of petroleum and metal asset returns. *Journal of Empirical Finance*, *18*, 160–173.
Chen, Y., & Lu, J. (2012). Value at risk estimation. In J.-C. Duan, W. K. Härdle, & J. E. Gentle (Eds.), *Handbook of computational finance*. Berlin: Springer-Verlag.
Chen, S. X., & Tang, C. Y. (2005). Nonparametric inference of value-at-risk for dependent financial returns. *Journal of Financial Econometrics*, *3*(2), 227–255.
Chiu, C. L., Lee, M. C., & Hung, J. C. (2005). Estimation of value-at-risk under jump dynamics and asymmetric information. *Applied Financial Economics*, *15*, 1095–1106.
Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, *39*, 841–862.
Christoffersen, P. F. (2009). Value-at-risk models. In T. Andersen, R. Davis, J.-P. Kreiss, & T. Mikosch (Eds.), *Handbook of financial time series*. Springer Verlag.
Christoffersen, P. F. (2012). *Elements of financial risk management* (2nd Ed.). Amsterdam: Academic Press.
Christoffersen, P. F., & Goncalves, S. (2005). Estimation risk in financial risk management. *Journal of Risk*, *7*, 1–28.
Christoffersen, P. F., Hahn, J., & Inoue, A. (2001). Testing and comparing value-at-risk measures. *Journal of Empirical Finance*, *8*, 325–342.
Christoffersen, P. F., & Pelletier, D. (2004). Backtesting value-at-risk: a duration-based approach. *Journal of Financial Econometrics*, *2*(1), 84–108.
Clements, M. P., Galvao, A. B., & Kim, J. H. (2008). Quantile forecast of daily exchange returns from forecasts of realized volatility. *Journal of Empirical Finance*, *15*, 729–750.
Clements, M. P., & Taylor, N. (2003). Evaluating interval forecasts of high-frequency financial data. *Journal of Applied Econometrics*, *18*, 445–456.
Colletaz, G., Hurlin, C., & Pérignon, C. (2013). The risk map: a new tool for validating risk models. *Journal of Banking and Finance*, *37*(10), 3843–3856.
Coroneo, L., & Veredas, D. (2012). A simple two-component model for the distribution of intraday returns. *European Journal of Finance*, *18*(9), 775–797.
Danielsson, J. (2011). *Financial risk forecasting*. Wiley.
Danielsson, J., de Haan, L., Peng, L., & de Vries, C. G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, *76*, 226–248.
Danielsson, J., & de Vries, C. G. (2000). Value at risk and extreme returns. *Annales d'Economie et Statistique*, *60*, 239–270.
Danielsson, J., Jorgensen, B. N., Samorodnitsky, G., Sarma, M., & de Vries, C. G. (2013). Fat tails, VaR and subadditivity. *Journal of Econometrics*, *172*, 283–291.
Danielsson, J., Jorgensen, B. N., Sarma, M., & de Vries, C. G. (2006). Comparing downside risk measures for heavy tailed distributions. *Economics Letters*, *92*(2), 202–208.
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), Article 3. http://dx.doi.org/10.2202/1558-3708.1720.
Degiannakis, S. (2004). Volatility forecasting: evidence from a fractionally integrated asymmetric power ARCH skewed-*t* model. *Applied Financial Economics*, *14*, 1333–1342.
de la Pena, V. H., Rivera, R., & Ruiz-Mata, J. (2007). Quality control of risk measures: backtesting VaR models. *Journal of Risk*, *9*(2), 39–54.
De Rossi, G., & Harvey, A. C. (2009). Quantiles, expectiles and splines. *Journal of Econometrics*, *152*, 179–185.
Dias, A. (2013). Market capitalization and value-at-risk. *Journal of Banking and Finance*, *37*(12), 5248–5260.
Diebold, F. X. (1988). *Empirical modeling of exchange rate dynamics*. New York: Springer-Verlag.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.
Diebold, F. X., Schuermann, T., & Stroughair, J. D. (2000). Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance*, *1*(2), 30–35.
Ding, Z., Granger, C., & Engle, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, *1*, 83–106.
Dowd, K. (2006). Using order statistics to estimate confidence intervals for probabilistic risk measures. *Journal of Derivatives*, *14*(2), 77–81.
Dowd, K. (2007). *Measuring market risk* (2nd Ed.). Chichester: Wiley.
Dowd, K., & Blake, D. (2006). After VaR: the theory, estimation, and insurance applications of quantile-based risk measures. *The Journal of Risk and Insurance*, *73*(2), 193–229.
Duan, J. (2004). Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, *4*, 155–167.
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.
Dupuis, D. J. (1999). Exceedances over high thresholds: A guide to threshold selection. *Extremes*, *1*, 251–261.
Dupuis, D. J., Papageorgiou, N., & Rémillard, B. (2015). Robust conditional variance and value-at-risk estimation. *Journal of Financial Econometrics*, *13*(4), 896–921.

Embrechts, P. (2009). Linear correlation and EVT: properties and caveats. *Journal of Financial Econometrics*, *7*(1), 30–39.
Embrechts, P., Klüppelberg, C., & Mikosch, T. (2000). *Modelling extremal events for insurance and finance*. London: Springer.
Engle, R.F. (2003). *Risk and volatility: econometric models and financial practice*, Nobel Lecture.
Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, *22*(4), 367–381.
Engle, R. F., & Russell, J. (1998). Autoregressive conditional duration: a new model for irregularly spaced transaction data. *Econometrica*, *66*, 1127–1162.
Ergen, I. (2015). Two-step methods in VaR prediction and the importance of fat tails. *Quantitative Finance*, *15*(6), 1013–1030.
Escanciano, J. C., & Olmo, J. (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business and Economic Statistics*, *28*(1), 36–51.
Escanciano, J. C., & Olmo, J. (2011). Robust backtesting tests for value-at-risk models. *Journal of Financial Econometrics*, *9*, 132–161.
Escanciano, J. C., & Pei, P. (2012). Pitfalls in backtesting historical simulation VaR models. *Journal of Banking and Finance*, *36*, 2233–2244.
Escanciano, J. C., & Velasco, C. (2010). Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, *159*, 209–221.
Fan, J. (1992). Desing-adaptive nonparametric regression. *Journal of the American Statistical Association*, *87*, 99–1004.
Fan, J., & Gu, J. (2003). Semiparametric estimation of value at risk. *Econometrics Journal*, *6*, 260–289.
Ferreira, A., de Haan, L., & Peng, L. (2003). On optimizing the estimation of high quantiles of a probability distribution. *Statistics*, *37*(5), 401–434.
Ferraty, F., & Quintela-del-Río, A. (2015). Conditional VaR and expected shortfall: A new functional approach. *Econometric Reviews*, in press.
Fleming, J., & Kirby, C. (2003). A closer look at the relation between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics*, *1*, 365–419.
Francq, C., & Zakoïan, J.-M. (2015). Risk-parameters estimation in volatility models. *Journal of Econometrics*, *184*(1), 158–173.
Frésard, L., Pérignon, C., & Wilhelmsson, A. (2011). The pernicious effects of contaminated data in risk management. *Journal of Banking and Finance*, *35*(10), 2569–2583.
Fuertes, A.-M., & Olmo, J. (2013). Optimally harnessing inter-day and intra-day information for Value-at-Risk prediction. *International Journal of Forecasting*, *29*(1), 28–42.
Gaglianone, W. P., Lima, L. R., Linton, O., & Smith, D. R. (2011). Evaluating Value-at-Risk models via quantile regression. *Journal of Business and Economic Statistics*, *29*(1), 150–160.
Gao, F., & Song, F. (2008). Estimation risk in GARCH VaR and ES estimates. *Econometric Theory*, *24*(5), 1404–1424.
Gençay, R., & Selçuk, F. (2004). Extreme value theory and value-at-risk: Relative performance in emerging markets. *International Journal of Forecasting*, *20*(2), 287–303.
Gerlach, R. H., Chen, C. W. S., & Chan, N. Y. C. (2011). Bayesian time-varying quantile forecasting for value-at-risk in financial markets. *Journal of Business and Economic Statistics*, *29*(4), 481–492.
Gerlach, R., Lu, Z., & Huang, H. (2013). Exponentially smoothing the skewed Laplace distribution for value-at-risk forecasting. *Journal of Forecasting*, *32*, 534–550.
Geweke, J., & Amisano, G. (2011). Hierarchical Markov normal mixture with applications to financial asset returns. *Journal of Applied Econometrics*, *26*(1), 1–29.
Giacomini, R., & Komunjer, I. (2005). Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics*, *23*, 416–431.
Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.
Giannopoulos, K., & Tunaru, R. (2005). Coherent risk measures under filtered historical simulation. *Journal of Banking and Finance*, *29*(4), 979–996.
Giot, L., & Laurent, S. (2003). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics*, *18*(6), 641–663.
Giot, L., & Laurent, S. (2004). Modeling daily Value-at-Risk using realized volatility and ARCH-type models. *Journal of Empirical Finance*, *11*, 379–398.
Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, *106*(494), 746–762.
Gomes, M. I., de Haan, L., & Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: Accommodation of bias in weighted log-excesses. *Journal of the Royal Statistical Society, Series B*, *70*(1), 31–52.
Gomes, M. I., Figueiredo, F., Rodrigues, L. H., & Miranda, M. C. (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.
Gomes, M. I., Matins, M. J., & Neves, M. M. (2007). Improving second order reduced-bias tail index estimator. *Review of Statistics*, *5*(2), 177–207.
Gomes, M. I., & Pestana, D. (2007). A sturdy reduced bias extreme quantile (VaR) estimator. *Journal of the American Statistical Association*, *102*(477), 280–292.
Gong, Y., Zhouping, L., & Peng, L. (2010). Empirical likelihood intervals for conditional value-at-risk in ARCH-GARCH models. *Journal of Time Series Analysis*, *31*(2), 65–75.
González-Rivera, G., Lee, T.-H., & Mishra, S. (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.
González-Rivera, G., Lee, T.-H., & Yoldas, E. (2007). Optimality of the RiskMetrics VaR model. *Finance Research Letters*, *4*, 137–145.
Gonzalo, J., & Olmo, J. (2004). Which extreme values are really extreme? *Journal of Financial Econometrics*, *2*, 349–369.
Gourieroux, C., & Jasiak, J. (2008). Dynamic quantile models. *Journal of Econometrics*, *147*, 198–205.
Gourieroux, C., & Jasiak, J. (2010a). Value at risk. In Y. Aït-Sahalia, & L. P. Hansen (Eds.), *Handbook of financial econometrics*. Amsterdam: North-Holland.
Gourieroux, C., & Jasiak, J. (2010b). Local likelihood density estimation and value-at-risk. *Journal of Probability and Statistics*, Article ID: 754851.
Gourieroux, C., & Zakoïan, J.-M. (2013). Estimation-adjusted VaR. *Econometric Theory*, *29*(4), 735–770.
Granger, C. W. J. (2002). Some comments on risk. *Journal of Applied Econometrics*, *17*, 447–456.
Grigoletto, M., & Lisi, F. (2009). Looking for skewness in financial time series. *Econometrics Journal*, *12*(2), 310–323.
Grigoletto, M., & Lisi, F. (2011). Practical implications of higher moments in risk management. *Statistical Methods and Applications*, *20*, 487–506.
Guermat, C., & Harris, R. (2001). Robust conditional variance estimation and value-at-risk. *Journal of Risk*, *4*, 25–41.
Haas, M. (2005). Improved duration-based backtesting of value-at-risk. *Journal of Risk*, *8*(2), 17–38.
Halbleib, R., & Pohlmeier, W. (2012). Improving the value at risk forecasts: Theory and evidence from financial crisis. *Journal of Economic Dynamics and Control*, *36*(8), 1212–1228.
Hall, P., & Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica*, *71*, 285–317.
Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, *35*, 705–730.
Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics*, *23*(4), 365–380.
Hartz, C., Mittnik, S., & Paolella, M. (2006). Accurate value-at-risk forecasting based on the normal-GARCH model. *Computational Statistics and Data Analysis*, *51*(4), 2295–2312.
Herrera, R., & Schipp, B. (2013). Value at risk forecasts by extreme value models in a conditional duration framework. *Journal of Empirical Finance*, *23*, 33–47.
Herwartz, H. (2009). Exact inference in diagnosing Value-at-Risk estimates — A Monte Carlo evidence. *Economics Letters*, *103*, 160–162.
Herwartz, H., & Waichman, I. (2010). A comparison of bootstrap and Monte-Carlo testing approaches to value-at-risk diagnosis. *Computational Statistics*, *25*, 725–732.
Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, *3*(5), 1163–1174.
Hill, J. B. (2010). On tail index estimation for dependent heterogeneous data. *Econometric Theory*, *26*(5), 1398–1436.
Hoogerheide, L., Ravazzolo, F., & van Dijk, H. K. (2012). Comment. *Journal of Business and Economic Statistics*, *30*(1), 30–33.
Hoogerheide, L., & van Dijk, H. K. (2010). Bayesian forecasting of value at risk and expected shortfall using adaptive importance sampling. *International Journal of Forecasting*, *26*, 231–247.
Huang, H., & Lee, T.-H. (2013). Forecasting value-at-risk using high frequency information. *Econometrics*, *1*, 127–140.
Huang, D., Yu, B., Luz, Z., Fabozzi, F., Focardi, S., & Fukushima, M. (2010). Index-exciting CAViaR: a new empirical time-varying risk model. *Studies in Nonlinear Dynamics and Econometrics*, *14*(1), http://dx.doi.org/10.2202/1558-3708.1805.
Huisman, R., Koedijk, K. G., Kool, C. J. M., & Palm, F. (2001). Tail index estimates in small samples. *Journal of Business and Economic Statistics*, *19*, 208–216.
Hull, J., & White, A. (1998). Incorporating volatility updating into the historical simulation method for value at risk. *Journal of Risk*, *1*, 5–19.
Hurlin, C., & Tokpavi, S. (2007). Backtesting value-at-risk accuracy: a simple new test. *Journal of Risk*, *9*(2), 19–37.
Jalal, A., & Rockinger, M. (2008). Predicting tail-related risk measures: The consequences of using GARCH filters for non-GARCH data. *Journal of Empirical Finance*, *15*, 868–877.

Jeon, J., & Taylor, J. W. (2013). Using CAViaR models with implied volatility for value-at-risk estimation. *Journal of Forecasting*, *32*, 62–74.

Jorion, P. (1996). Risk^2: Measuring the risk in value at risk. *Financial Analysts Journal*, *52*, 47–56.

Jorion, P. (2002). How informative are value-at-risk disclosures? *Accounting Review*, *77*(4), 911–931.

Jorion, P. (2006). *Value-at-Risk: the new benchmark for managing financial risk* (3rd Ed.). New York: McGraw-Hill.

Jorion, P. (2009). Risk management lessons from the credit crisis. *European Financial Management*, *15*(5), 923–933.

Juarez, S. F., & Schucany, W. R. (2004). Robust and efficient estimation for generalized Pareto estimation. *Extremes*, *7*, 237–251.

Kerkhof, J., & Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking and Finance*, *28*, 1845–1865.

Koenker, R., & Basset, G. (1978). Regression quantiles. *Econometrica*, *46*, 33–50.

Koenker, R., & Xiao, Z. (2002). Inference on the quantile regression process. *Econometrica*, *70*(4), 1583–1612.

Komunjer, I. (2005). Quasi-maximum likelihood estimation for conditional quantiles. *Journal of Econometrics*, *128*, 137–164.

Komunjer, I. (2007). Asymmetric power distribution: theory and applications to risk measurement. *Journal of Applied Econometrics*, *22*, 891–921.

Krause, A. (2003). Exploring the limitations of value at risk: How good is it in practice?. *The Journal of Risk and Finance*, 19–28. Winter.

Kuester, K., Mittik, S., & Paolella, M. S. (2006). Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, *4*(1), 53–89.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, *2*, 173–184.

Leccadito, A., Boffelli, S., & Urga, G. (2014). Evaluating the accuracy of value-at-risk forecasts: New multilevel tests. *International Journal of Forecasting*, *30*, 206–216.

Lehar, A., Scheicher, M., & Schittenkopf, C. (2002). GARCH vs. stochastic volatility: option pricing and risk management. *Journal of Banking and Finance*, *26*, 323–345.

Li, Z., Gong, Y., & Peng, L. (2011). Empirical likelihood intervals for conditional value-at-risk in heteroscedastic regression models. *Scandinavian Journal of Statistics*, *38*(4), 781–787.

Li, D., Peng, L., & Yan, J. (2010). Bias reduction for high quantiles. *Journal of Statistical Planning and Inference*, *140*(6), 2433–2441.

Lien, D., Yan, X., & Ye, K. (2014). Alternative approximations to value-at-risk. *Communications in Statistics: Simulation and Computation*, *43*(10), 2225–2240.

Liu, S., & Tse, Y.-K. (2015). Intraday value-at-risk: An asymmetric autoregressive conditional duration approach. *Journal of Econometrics*, *189*(2), 437–446.

Longin, F. (2000). From value at risk to stress testing: the extreme value approach. *Journal of Banking and Finance*, *24*, 1097–1130.

Lönnbark, C. (2010). A corrected value-at-risk predictor. *Applied Economic Letters*, *17*(12), 1193–1196.

Lönnbark, C. (2013). On the role of the estimation error in prediction expected shortfall. *Journal of Banking and Finance*, *37*(3), 847–853.

Lopez, A. J. (1999). Regulatory evaluation of value-at-risk models. *Journal of Risk*, *1*, 37–64.

Louzis, D. P., Xanthopoulos-Sisinis, S., & Refenes, A. P. (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.

Maheu, J. M., & McCurdy, T. H. (2004). News arrival, jump dynamics and volatility components for individual stock returns. *Journal of Finance*, *59*, 755–793.

Mancini, L., Ronchetti, E., & Trojani, F. (2005). Optimal conditionally unbiased bounded-influence inference in dynamic location and scale models. *Journal of the American Statistical Association*, *100*, 628–641.

Mancini, L., & Trojani, F. (2011). Robust value-at-risk prediction. *Journal of Financial Econometrics*, *9*(2), 281–313.

Martens, M., van Dijk, D., & Pooter, M. (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.

Martins-Filho, C., & Yao, F. (2006). Estimation of value-at-risk and expected shortfall based on nonlinear models of returns dynamics and extreme value theory. *Studies in Nonlinear Dynamics and Econometrics*, *10*(2), 1–43.

McAleer, M. (2009). The ten commandments for optimizing value-at-risk and daily capital charges. *Journal of Economic Surveys*, *23*(5), 831–849.

McAleer, M., & da Veiga, B. (2008). Single index and portfolio models for forecasting value-at-risk thresholds. *Journal of Forecasting*, *27*, 217–235.

McAleer, M., Jiménez-Martín, J.-A., & Pérez-Amaral, T. (2010). What happened to risk management during the 2008–09 financial crisis?. In R. W. Kolb (Ed.), *Lessons from the financial crisis: causes, consequences, and our economic future*. New York: Wiley.

McAleer, M., Jiménez-Martín, J.-A., & Pérez-Amaral, T. (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.

McAleer, M., Jiménez-Martín, J.-A., & Pérez-Amaral, T. (2013b). International evidence on GFC-robust forecasts for risk management under the Basel Accords. *Journal of Forecasting*, *32*, 267–288.

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, *7*, 271–300.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative risk management: concepts, techniques, tools*. Princeton University Press.

Mincer, J., & Zarnowitz, V. (1969). The evaluation of economic forecasts and expectations. In J. Mincer (Ed.), *Economic forecasts and expectations*. New York: National Bureau of Economic Research.

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.

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

Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, *28*(3), 777–787.

Pascual, L., Ruiz, E., & Romo, J. (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics and Data Analysis*, *50*(9), 2293–2312.

Patton, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review*, *47*, 527–556.

Patton, A., & Timmermann, A. (2012). Forecast rationality tests based on multi-horizon bounds. *Journal of Business and Economics Statistics*, *30*(1), 1–17.

Pelletier, D., & Wei, W. (2015). The Geometric-VaR backtesting method. *Journal of Financial Econometrics*, in press.

Peng, L., & Qi, Y. (2006a). Confidence regions for high quantiles of a heavy distribution. *Annals of Statistics*, *34*(4), 1964–1986.

Peng, L., & Qi, Y. (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.

Pérignon, C., Deng, Z. Y., & Wang, Z. Y. (2008). Do banks overstate their value-at-risk? *Journal of Banking and Finance*, *32*, 783–794.

Pérignon, C., & Smith, D. R. (2008). A new approach to comparing VaR estimation method. *The Journal of Derivatives*, *16*(2), 54–66.

Pérignon, C., & Smith, D. R. (2010a). Diversification and Value-at-Risk. *Journal of Banking and Finance*, *34*, 55–66.

Pérignon, C., & Smith, D. R. (2010b). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking and Finance*, *34*, 362–377.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, *89*(428), 1303–1313.

Politis, D. N., & White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, *23*(1), 53–70.

Politis, D. N., & White, H. (2009). Correction to "Automatic block-length selection for the dependent bootstrap". *Econometric Reviews*, *28*(4), 372–375.

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking and Finance*, *30*(2), 561–582.

Ridder, T. (1998). Basics of statistical VaR-estimation. In G. Bol, G. Nakhaeizadeh, & K.-. H. Vollmer (Eds.), *Risk measurement, econometrics and neural networks*. New York: Physica-Verlag.

Rocco, M. (2014). Extreme value theory in finance: A survey. *Journal of Economic Surveys*, *28*(1), 82–108.

Rodríguez, M. J., & Ruiz, E. (2012). Revisiting several popular GARCH models with leverage effect: differences and similarities. *Journal of Financial Econometrics*, *10*(4), 637–668.

Rubia, A., & Sanchis-Marco, L. (2013). On downside risk predictability through liquidity and trading activity: A dynamic quantile approach. *International Journal of Forecasting*, *29*(1), 202–219.

Ruiz, E., & Pascual, L. (2002). Bootstrapping financial time series. *Journal of Economic Surveys*, *16*, 271–300.

Sajjad, R., Coakley, J., & Nankervis, T. (2008). Markov Switching GARCH modelling of Value at Risk. *Studies in Nonlinear Dynamics and Econometrics*, *12*(3), http://dx.doi.org/10.2202/1558-3708.1522.

Sandström, A. (2011). *Handbook of solvency for actuaries and risk managers: theory and practice*. Boca Raton: Chapman & Hall.

Santos, A. P. P., & Alves, M. F. (2012a). Forecasting value-at-risk with a duration-based POT method. *Mathematics and Computers in Simulation*, *94*, 295–309.

Santos, A. P. P., & Alves, M. F. (2012b). A new class of independence tests for interval forecast evaluation. *Computational Statistics and Data Analysis*, *56*(11), 3366–3380.
Sarma, M., Thomas, S., & Shah, A. (2003). Selection of value at risk models. *Journal of Forecasting*, *22*, 337–358.
Scarrott, C., & MacDonald, A. (2012). A review of extreme value threshold estimation with uncertainty quantification. *REVSTAT, Statistical Journal*, *10*(1), 33–60.
Sener, E., Baronyan, S., & Mengütürk, L. A. (2012). Ranking the predictive performances of value-at-risk estimation methods. *International Journal of Forecasting*, *28*(4), 849–873.
Sherman, M., & Carlstein, E. (2004). Confidence intervals based on estimators with unknown rates of convergence. *Computational Statistics and Data Analysis*, *46*, 123–139.
So, M., & Yu, P. (2006). Empirical analysis of GARCH models in value at risk estimation. *Journal of International Markets, Institutions and Money*, *18*, 180–197.
Spierdijk, L. (2015). Confidence intervals for ARMA-GARCH value-at-risk: The case of heavy tails and skewness. *Computational Statistics and Data Analysis*, in press.
Stoyanov, S. V., Rachev, S. T., & Fabozzi, F. J. (2013). CVaR sensitivity with respect to tail thickness. *Journal of Banking and Finance*, *37*, 977–988.
Taniai, H., & Taniguchi, M. (2008). Statistical estimation errors of VaR under ARCH returns. *Journal of Statistical Planning and Inference*, *139*, 3568–3577.
Taylor, J. W. (2008a). Estimating value at risk and expected shortfall using expectiles. *Journal of Financial Econometrics*, *6*(2), 231–252.
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.
Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Management Science*, *44*, 1650–1661.
Tolikas, K. (2014). Unexpected tail in risk measurement: Some international evidence. *Journal of Banking and Finance*, *40*, 476–493.
Wang, H. J., Li, D., & He, X. (2012). Estimation of high conditional quantiles for heavy-tailed distributions. *Journal of the American Statistical Association*, *107*(500), 1453–1464.
Wang, Z., Weitao, W., Chen, C., & Zhou, Y. (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.
Wilhelmsson, A. (2009). Value at risk with time varying variance, skewness and kurtosis — the NIG-ACD model. *The Econometrics Journal*, *12*, 82–104.
Wong, W. K. (2010). Backtesting value-at-risk based on tail losses. *Journal of Empirical Finance*, *17*(3), 526–538.
Wu, G., & Xiao, Z. (2002). An analysis of risk measures. *Journal of Risk*, *4*, 53–75.
Xiao, Z., & Koenker, R. (2009). Conditional quantile estimation and inference for GARCH models. *Journal of the American Statistical Association*, *104*(488), 1696–1712.
Xu, K.-L. (2013). Nonparametric inference for conditional quantiles of time series. *Econometric Theory*, *29*(4), 673–698.
Xu, K.-L. (2015). Model-free inference for tail risk measures. *Econometric Theory*, in press.
Yamai, Y., & Yoshiba, T. (2005). Value-at-risk versus expected shortfall: a practical perspective. *Journal of Banking and Finance*, *29*, 997–1015.
Yi, Y., Feng, X., & Huang, Z. (2014). Estimation of extreme value-at-risk: An EVT approach for quantile GARCH model. *Economics Letters*, *124*(3), 378–381.
Yu, K., & Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, *93*(441), 228–237.
Yu, P. L. H., Li, W. K., & Jin, S. (2010). On some models for value-at-risk. *Econometric Reviews*, *29*(5–6), 622–641.
Ziggel, D., Berens, T., Wei, G. N. F., & Wied, D. (2014). A new set of improved value-at-risk backtests. *Journal of Banking and Finance*, *14*, 29–41.
Zikovic, S., & Aktan, B. (2011). Decay factor optimization in time weighted simulation — evaluating VaR performance. *International Journal of Forecasting*, *27*(4), 1147–1159.

Maria Rosa Nieto is Lecturer and Researcher at the Universidad Anáhuac in Mexico since 2012. She got her Ph.D. in Business Administration and Quantitative Methods at the Statistics Department of the Universidad Carlos III de Madrid in 2010.

Esther Ruiz is Full Professor of Econometrics at Universidad Carlos III de Madrid where she has been since 1992. She obtained her Master and Ph.D. degrees in Statistics from the London School of Economics (LSE) in 1992 and after that she was visiting Professor at LSE for one year. Prof. Ruiz publications have appeared in journals as the *International Journal of Forecasting, Journal of Econometrics, Journal of Time Series Analysis, Review of Economic Studies, Statistica Sinica, Journal of Financial Econometrics* and *Computational Statistics & Data Analysis* among others. She works mainly on methodological and empirical issues related with modelling and forecasting financial time series. One of her papers on stochastic volatility has been path breaking, being one of the most cited papers in the area. Prof. Ruiz is co-Editor of the *International Journal of Forecasting*.