

Auer, Benjamin R.; Mögel, Benjamin

Working Paper

How Accurate are Modern Value-at-Risk Estimators Derived from Extreme Value Theory?

CESifo Working Paper, No. 6288

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Auer, Benjamin R.; Mögel, Benjamin (2016) : How Accurate are Modern Value-at-Risk Estimators Derived from Extreme Value Theory?, CESifo Working Paper, No. 6288, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:

<http://hdl.handle.net/10419/155530>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Working Papers

www.cesifo.org/wp

How Accurate are Modern Value-at-Risk Estimators Derived from Extreme Value Theory?

Benjamin R. Auer
Benjamin Mögel

CESIFO WORKING PAPER NO. 6288

CATEGORY 7: MONETARY POLICY AND INTERNATIONAL FINANCE

DECEMBER 2016

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

ISSN 2364-1428

How Accurate are Modern Value-at-Risk Estimators Derived from Extreme Value Theory?

Abstract

In this study, we compare the out-of-sample forecasting performance of several modern Value-at-Risk (VaR) estimators derived from extreme value theory (EVT). Specifically, in a multi-asset study covering 30 years of stock, bond, commodity and currency market data, we analyse the accuracy of the classic generalised Pareto peak over threshold approach and three recently proposed methods based on the Box-Cox transformation, L-moment estimation and the Johnson system of distributions. We find that, in their unconditional form, some of the estimators are acceptable under current regulatory assessment rules but none of them can continuously pass more advanced tests of forecasting accuracy. In their conditional forms, forecasting power is significantly increased and the Box-Cox method proves to be the most promising estimator. However, it is also important to stress that the traditional historical simulation approach, which is currently the most frequently used VaR estimator in commercial banks, can not only keep up with the EVT-based methods but occasionally even outperforms them (depending on the setting: unconditional vs. conditional). Thus, recent claims to generally replace this simple method by theoretically more advanced EVT-based methods may be premature.

JEL-Codes: G100, G110, G170.

Keywords: value-at-risk, extreme value theory, historical simulation, backtest, financial crisis.

*Benjamin R. Auer**
University of Leipzig
Department of Finance
Grimmaische Str. 12
Germany – 04109 Leipzig
auer@wifa.uni-leipzig.de

Benjamin Mögel
University of Leipzig
Department of Finance
Grimmaische Str. 12
Germany – 04109 Leipzig
moegel@wifa.uni-leipzig.de

*corresponding author

December 22, 2016

1 Introduction

Over the past decades, stock market crashes, bond market collapses and foreign exchange crises have forced financial institutions and regulators to recognise the growing importance of risk management. To protect single institutions and the entire financial system against such extraordinary events, the determination of capital requirements to meet catastrophic market risk has become an important topic, which in turn has stimulated the development of various methods to measure the risks that financial institutions face (see Christoffersen, 2003; Pflug and Römisch, 2007).

A particularly popular tool for financial risk assessment is the Value-at-Risk (VaR), which is defined as the maximum loss expected on a portfolio over a certain holding period at a given confidence level (see Berkowitz et al., 2011), i.e., for a portfolio with random return X and cumulative distribution function $F(x)$, the VaR with a confidence level of $0 < \alpha < 1$ (typically set to 95% or 99%) is simply the $(1 - \alpha)$ -quantile associated with $F(x)$. Regardless of some criticisms,¹ risk management is now heavily geared toward the VaR. Besides risk reporting to senior managers and shareholders, the VaR is applied for allocating financial resources and for evaluating risk-adjusted performance (see Jorion, 2007). Furthermore, with the advent of the internal model approach of the Basel Committee on Banking Supervision (1995, 1996a, 2009, 2011a), banks in the main financial jurisdictions may now use their in-house VaR models for the calculation of regulatory capital requirements (see Brooks et al., 2005). In light of this high practical relevance of the VaR concept, the need for reliable VaR estimation and prediction strategies has arisen.

Accurate estimation of the VaR depends on the ability to estimate the tails of the probability density function $f(x)$ associated with $F(x)$. Conceptually, this can be accomplished in two distinct ways: (i) direct estimation of $f(x)$, or (ii) indirectly through a suitably defined (parametric) model for the tails of $f(x)$. Early implementations of the VaR have followed the first approach by assuming that the return-generating process for the portfolio in question is normally distributed with moments that can be estimated using historical data and are time-invariant (see Danielsson and de Vries, 2000; Neftci, 2000). Even though this simplifies VaR computation considerably, the normality-assumption is inconsistent with empirical evidence of skewed and fat-tailed asset returns (see Cont, 2001). Extreme events are much more likely to occur in practice than would be predicted by the symmetric thinner-tailed normal distribution, which makes normality-based VaR numbers unsuitable measures of the true risk (see Lee and Su, 2012). Similar problems arise for other distributional assumptions because, unless estimation is based on a correct specification of $f(x)$, direct estimation will most likely provide a poor fit for its tails since most observed data will probably take values away from the tail region of $f(x)$ (see Martins-Filho and Yao, 2006). To compensate for these drawbacks, a series of indirect estimation methods based on extreme value theory (EVT) has recently emerged, including Embrechts et al. (1999), Longin (2000), McNeil and Frey (2000), Brooks et al. (2005), Martins-Filho and Yao (2006) and Bali (2007). These methods follow the idea of approximating only the tails of $f(x)$ by an appropriately defined parametric density function in order to resolve the the problem of underpredicting both the size and the frequency of extreme market movements.

Empirical evidence indicates that EVT approaches may be considered superior to most direct methods (see Danielsson and Morimoto, 2000; Kuester et al., 2006; Abad et al., 2014) and suggests that historical simulation, which is the most frequently used VaR estimation method in commercial banks (see Pérignon and Smith, 2010), may have to be replaced by these more advanced methods. While this result is important for choosing the adequate VaR estimation methodology in practical applications, the question of which EVT approach should be preferred, largely remains unanswered. In the literature, we can detect a tendency that whenever a new EVT approach is proposed it is

¹The VaR is not a coherent risk measure (see Artzner et al., 1999). It can lead to Pareto-inferior allocations if agents are risk averse. In addition, the VaR can fail to account appropriately for portfolio risk diversification (see Yamai and Yoshida, 2002, 2005).

usually compared either to standard direct approaches or to very early EVT methods.² To the best of our knowledge, a structured empirical comparison of more recent EVT methods does not yet exist in the literature.

The purpose of our study is to fill this gap by comparing several modern EVT approaches for univariate VaR prediction and to provide some guidance for choosing the appropriate estimation strategy in practice. In principle, various estimators could be considered in such an analysis (see Abad et al., 2014). Our choice of alternatives was motivated not by the desire to be exhaustive but rather to represent what is commonly used both in the empirical finance literature and in practice. Specifically, we compare the peak over threshold estimator of McNeil and Frey (2000), the Box-Cox method of Bali (2007), the L-moment technique of Martins-Filho and Yao (2006) and the simulated Johnson system approach of Brooks et al. (2005). Furthermore, to assess their performance relative to classic direct methods, we also include the normal, Student-t and historical simulation approaches in our set of VaR estimators. In general, we use all VaR estimators in the specifications which have become standard because this way we can judge the errors of the approaches as they would actually occur in practice. Given the serial correlation and heteroskedasticity of most financial data (see Cont, 2001), we implement all approaches in an unconditional and a conditional form. This is important because unconditional implementations may weaken the empirical performance of VaR estimators because they do not reflect the current mean-volatility background.

We examine VaR forecasting performance for portfolios that are long in different asset classes because this way we can address the critique of Bali (2007) that most studies focus on only one asset class (mainly stocks) and thus their results may not hold for other assets with different return characteristics. Specifically, we cover stock, commodity, bond and currency markets from 1986 to 2016 in order to obtain a more general picture. For each asset class, we obtain VaR predictions by using a rolling window out-of-sample approach because, in practice, portfolio managers obtain VaR estimates based on the past data and subsequently use these estimates to assess the risks associated with current and future movements of their portfolios' value. The quality of the VaR forecasts is then evaluated by means of standard tests as well as some recently proposed (more advanced) techniques.

Following the literature standard (see Kuester et al., 2006; Basel Committee on Banking Supervision, 2011b; Cheng and Hung, 2011), the core of our performance evaluation is the Christoffersen (1998) framework because it specifies several properties a predictor needs to satisfy in order to be considered accurate. That is, for example, that the nominal error probability $1 - \alpha$ used in the VaR estimation should equal the frequency of violations (i.e., cases where the actual losses exceed the estimated VaR). The statistical test for this property (the unconditional coverage test), which is based on the binomial distribution, is currently indirectly embodied in the 1996 Market Risk Amendment to the Basel Capital Accord (see Basel Committee on Banking Supervision, 1996b; Ziggel et al., 2014) which requires that commercial banks with significant trading activities set aside capital to cover the market risk exposure by their trading accounts. The market risk capital requirements are to be based on the VaR estimates generated by the banks' own risk management models provided that they pass this test.

While the unconditional coverage test is the current regulatory standard, it leaves important additional properties of an appropriate VaR predictor unexamined. Thus, banks may identify an approach that fulfils the regulatory minimum standard but still provides inadequate forecasts which is not desirable from a rational perspective. To avoid such misjudgement, we apply two additional tests which have become the most frequently used tests in the empirical literature (see Huang and Lin, 2004; Bali, 2007; Marimoutou et al., 2009; Gaglianone et al., 2011). First, we use

²For example, McNeil and Frey (2000) compare their peak over threshold estimator to historical simulation and Brooks et al. (2005) evaluate the performance of their simulation-based EVT approach relative to classic tail estimators (e.g., the Hill estimator).

the standard conditional coverage which takes into account that violations should be independent over time. Second, because the information set of this test is restricted to past violations, we also implement the dynamic quantile test of Engle and Manganelli (2004) which uses a more general information set.

Our empirical analysis is supported by a variety of sensitivity checks that ensure the robustness of our results. First, we use different types of models in our filter approaches for conditional VaR estimators. Second, we analyse the impact of different sample sizes and differences in results when looking at short positions instead of long positions. Third, we apply some recently proposed simulation-based variants of unconditional and conditional coverage tests which are robust to small sample issues and additionally take into account that the violations of an ideal predictor not only need to be independent but also have to be identically distributed. Finally, in a last exercise, we briefly document the behaviour of the VaR estimators in times of financial turmoil, i.e., in the time spans of the dotcom crash of 2000 and the global financial crisis of 2007 to 2009.

The remainder of the article is organised as follows. Section 2 starts with a discussion of our unconditional VaR estimators (traditional and EVT-based) followed by a description of filter processes producing conditional VaR estimates. Section 3 presents the main characteristics of our dataset and some preliminary results on the forecasting accuracy of the different VaR estimators. Section 4 describes the statistical tests we use for analysing the quality of VaR forecasts. Section 5 presents the test results and summarises the outcome of several robustness checks. Finally, Section 6 concludes and outlines some directions for future research.

2 Value-at-Risk estimation

2.1 Traditional approaches

2.1.1 Normal and Student-t method

Probably the best-known way to estimate the VaR is to assume normally distributed (ND) asset returns. For a confidence level α , the VaR forecast for time $t + 1$ is then

$$\text{VaR}_{t+1,\alpha}^{ND} = \mu + \sigma z_{1-\alpha} \quad (1)$$

where μ and σ denote the mean and the standard deviations of a moving window of T returns up to time t (estimated via maximum-likelihood), respectively, and $z_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of the standard normal distribution (see Linsmeier and Pearson, 2000).

Because returns are typically characterised by substantial kurtosis, a straightforward approach to generating better VaR forecasts is to assume that returns are sufficiently well described by a Student-t (ST) distribution which can model fat-tails. In this case, we can estimate the VaR via

$$\text{VaR}_{t+1,\alpha}^{ST} = \mu + \sigma \sqrt{\frac{v-2}{v}} t_{1-\alpha}^v, \quad (2)$$

where μ and σ are defined as stated above and $t_{1-\alpha}^v$ is the $(1 - \alpha)$ -quantile of the Student-t distribution with v degrees of freedom (see Alexander, 2008). In empirical applications, fat-tailed data is typically modelled by $v = 5$ (see Campbell et al., 2001) or by determining the optimal value of v via maximum-likelihood fit (see Lange et al., 1989).

2.1.2 Historical simulation method

In contrast to the two previous approaches, historical simulation, which is currently the most popular VaR estimation technique in investment practice (see Pritsker, 2006; Pérignon and Smith, 2010), is not bound to a prespecified theoretical distribution. Instead, it assumes that the distri-

bution of past returns is a good and complete representation of expected future returns. This way, it can not only account for kurtosis but additionally for skewness in empirical data.

For the confidence level α , historical simulation forecasts the VaR in $t + 1$ via the empirical $(1 - \alpha)$ -quantile, i.e.,

$$\text{VaR}_{t+1,\alpha}^{HS} = \text{quantile}_{1-\alpha}(x_t, x_{t-1}, \dots, x_{t-T+1}), \quad (3)$$

where x_t is the return in t . This means, for example, that the 95% VaR for a moving window of length 1,000 is simply the 50th value in a series of returns sorted in ascending order.

2.2 EVT approaches

2.2.1 Method of block maxima

Recent research has directed considerable attention towards deriving new VaR estimators from EVT because it is concerned with the asymptotic behaviour of extremes and thus represents a natural candidate for VaR modeling. In the early literature, we can identify two fairly simple approaches. While one makes use of the limiting distribution of block maxima and the other is based on the asymptotic theory of exceedances over given thresholds (see Gilli and K ellezi, 2006; Ferreira and De Haan, 2015). In this section, we cover the former.

Consider a sequence of independently and identically distributed (iid) random variables X_1, X_2, \dots with the unknown distribution function $F(x) := P(X_t \leq x)$ and define the maximum of a block of n observations to be $M_n := \max(X_1, \dots, X_n)$. A central result of EVT states that the generalised extreme value distribution (GEVD) is the only possible limiting distribution for block maxima as the block size increases. More formally, the limit law for the block maxima M_n is given by the following theorem dating back to Fisher and Tippett (1928) and Gnedenko (1943).

Theorem 1 (Fisher-Tippett-Gnedenko Theorem). *If there are constants $a_n > 0$, $b_n \in \mathbb{R}$ as well as some non-degenerate distribution function H such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow{d} H$$

then H belongs to one of the three standard extreme value distributions:

$$\text{Fr chet:} \quad \Phi_{\vartheta}(x) = \begin{cases} 0 & \text{if } x \leq 0 \text{ and } \vartheta > 0 \\ \exp(-x^{-\vartheta}) & \text{if } x > 0 \text{ and } \vartheta > 0 \end{cases}$$

$$\text{Weibull:} \quad \Psi_{\vartheta}(x) = \begin{cases} \exp(-(-x)^{\vartheta}) & \text{if } x \leq 0 \text{ and } \vartheta > 0 \\ 1 & \text{if } x > 0 \text{ and } \vartheta > 0 \end{cases}$$

$$\text{Gumbel:} \quad \Lambda(x) = \exp(-\exp(-x)) \text{ if } x \in \mathbb{R}.$$

This theorem illustrates that the normalised maxima M_n either follows a Fr chet, a Weibull or a Gumbel distribution. However, in general, we (i) do not know in advance the type of limiting distribution and (ii) do not know the norming constants. To address (i), we can use a one-parameter distribution proposed by Jenkinson (1955) and von Mises (1954) which includes all three standard extreme value distributions as special cases. This distribution is the GEVD, which has the form

$$H_{\xi}(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0 \\ \exp(-\exp(-x)) & \text{if } \xi = 0 \end{cases} \quad (4)$$

with x such that $1 + \xi x > 0$. It is obtained by setting $\xi = \vartheta^{-1} > 0$ ($\xi = -\vartheta^{-1}$) for the Fr chet (Weibull) distribution and by interpreting the Gumbel distribution as the limit case for $\xi = 0$. To additionally address (ii), the GEVD can be generalised to a three-parameter specification

$$H_{\xi,\gamma,\delta}(x) = \begin{cases} \exp\left(-\left(1 + \xi\frac{x-\gamma}{\delta}\right)^{-\frac{1}{\xi}}\right) & \text{if } \xi \neq 0 \\ \exp(-\exp(-\frac{x-\gamma}{\delta})) & \text{if } \xi = 0 \end{cases}, \quad (5)$$

for $1 + \xi\frac{x-\gamma}{\delta} > 0$, where γ and δ are the location and scale parameters representing the unknown norming constants (see Embrechts et al., 1999). This is the limiting distribution of the unnormalised maxima. It is particularly useful when maximum likelihood estimates have to be computed. Integrated into the Fisher-Tippett-Gnedenko Theorem it delivers an analogous position with respect to the study of maxima as the famous central limit theorem holds for the study of sums or averages of random variables.

Based on these results, the VaR of the block maxima (BM) method can be obtained by inverting (5) for a given confidence level α , i.e., by fitting the GEVD via maximum likelihood and plugging the resulting parameter estimates into

$$\text{VaR}_{t+1,\alpha}^{\text{GEVD}} = \begin{cases} \gamma - \frac{\delta}{\xi}(1 - (-\ln(1 - \alpha))^{-\xi}) & \text{if } \xi \neq 0 \\ \gamma - \delta \ln(-\ln(1 - \alpha)) & \text{if } \xi = 0 \end{cases}. \quad (6)$$

However, with few exceptions (see Ho et al., 2000; Longin, 2000), this approach is usually not utilized in practice because it is subject to several drawbacks. For example, it ignores other potentially important extremes within a block (see Marimoutou et al., 2009). Furthermore, there are no rules for optimal block-length selection (see Tsay, 2005) such that subsamples are typically based on calendar frequencies which introduces potential bias via seasonal effects (see Gilli and K ellezi, 2006). However, as we will see in the following sections, the Fisher-Tippett-Gnedenko Theorem is used in other forms of extreme value approaches.

2.2.2 Peak over threshold method

The peak over threshold (POT) method is one the most popular extreme value approaches in practice because, in contrast to the BM method, it makes more efficient use of the often limited data on extreme values (see Gen ay and Sel uk, 2004). Instead of focusing on block maxima, this method models the distribution of excesses $Y = X - u$ of X over a high threshold u . This distribution is defined as

$$F_u(y) = P(X - u \leq y | X > u), \quad (7)$$

for $0 \leq y < x_0 - u$, where $x_0 \leq \infty$ is the right endpoint of F (see McNeil, 1999).³ It represents the probability that X exceeds the threshold u by at most an amount y , given the information that it exceeds the threshold.

The following limit theorem of Pickands (1975) and Balkema and de Haan (1974) gives a key property of (7) and lays the cornerstone for obtaining the VaR based on the POT method.

Theorem 2 (Pickands-Balkema-de-Haan Theorem). *For a large class of underlying distribution functions,⁴ the conditional excess distribution function $F_u(y)$, for large u , is well approximated by the generalised Pareto distribution (GPD)*

³With an infinite right endpoint, we would allow the possibility of unreasonably large outcomes. Also note that the form of (7) requires to multiply the empirical long-position returns by -1 in order to model the correct tail.

⁴Of course this is not mathematically complete because we do not exactly say what we mean by ‘a large class’. For this article, it is sufficient to know that the class contains all the common continuous distributions of statistics and actuarial science (normal, lognormal, χ^2 , Student-t, F, gamma, exponential, uniform, beta, etc.).

$$G_{\xi,\delta}(y) = \begin{cases} 1 - (1 + \frac{\xi}{\delta}y)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - \exp(-\frac{y}{\delta}) & \text{if } \xi = 0 \end{cases}, \quad (8)$$

with $y \in [0, x_0 - u]$ if $\xi \geq 0$ and $y \in [0, -\frac{\delta}{\xi}]$ if $\xi < 0$,⁵ because we can find a positive measurable function $\delta(u)$ such that

$$\lim_{u \rightarrow x_0} \sup_{y \in (0, x_0 - u)} |F_u(y) - G_{\xi,\delta(u)}(y)| = 0. \quad (9)$$

Because we can rewrite (7) as $F_u(y) = [F(u+y) - F(u)]/[1 - F(u)] = [F(x) - F(u)]/[1 - F(u)]$, we can isolate $F(x)$ in the form $F(x) = (1 - F(u))F_u(y) + F(u)$ for $x > u$. Thus, we may move easily to an interpretation of the model in terms of the tail of the underlying distribution $F(x)$ for $x > u$. Replacing $F_u(y)$ by the GPD and $F(u)$ by the estimate $(T - T_u)/T$, where T is the total number of observations and T_u is the number of observations above the threshold u , we obtain

$$F(x) = \begin{cases} 1 - q (1 + \xi \frac{x-u}{\delta})^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - q \exp(-\frac{x-u}{\delta}) & \text{if } \xi = 0 \end{cases}, \quad (10)$$

where $q = T_u/T$ is the proportion of observations above the threshold (see Bali and Neftci, 2003). Given a confidence level α with $\alpha > 1 - q$ and a threshold u , we can invert this formula to get

$$\text{VaR}_{t+1,\alpha}^{\text{POT}} = \begin{cases} u + \frac{\delta}{\xi} ((\frac{1-\alpha}{q})^{-\xi} - 1) & \text{if } \xi \neq 0 \\ u + \delta \ln(\frac{q}{1-\alpha}) & \text{if } \xi = 0 \end{cases}. \quad (11)$$

To implement the POT method in an empirical application, we simply have to set u , fit the GPD to the excesses by means of maximum likelihood and plug the resulting parameter estimates for the GPD into (11). Typically, researchers set u such that $q = 10\%$ because McNeil and Frey (2000) show that for this setting, the potential errors caused by an insufficient number of observations in the GPD fit (too high u) and by a not fully achieved GPD convergence in the sample (too low u) are reasonably balanced.⁶ In a simulation using Student-t returns, these authors also show that the POT approach is more efficient than HS.

2.2.3 Box-Cox method

In a recent article, Bali (2007) proposes a novel VaR estimation approach based on a more general extreme value distribution and shows that it performs well in capturing both the rate of occurrence and the extent of extreme events in financial markets. He also highlights that the new method yields more precise VaR estimates than the normal and skewed-t distributions.

The new distribution is obtained following Bali (2003) by applying the one-parameter Box and Cox (1964) transformation to the GEVD of Jenkinson (1955). The resulting Box-Cox (BC) extreme value distribution has the cumulative distribution function

$$BC_{\xi,\gamma,\delta,\lambda}(x) = \left(\frac{\left(\exp \left(- \left(1 + \xi \left(\frac{x-\gamma}{\delta} \right) \right)^{-1/\xi} \right) \right)^\lambda - 1}{\lambda} \right) + 1, \quad (12)$$

⁵Similar to the GEVD, the GPD nests other distributions. The special cases $\xi = 0$ and $\xi = 1$ yield, respectively, the exponential distribution with mean δ and the uniform distribution on $[0, \delta]$. Classic Pareto distributions are obtained when $\xi < 0$ (see Hosking and Wallis, 1987).

⁶Alternative methods for threshold determination are described in El-Aroui and Diebold (2002) and Scarrott and MacDonald (2012).

where the nature of the transformation ensures that $BC_{\xi,\gamma,\delta,\lambda}(x) \rightarrow 1$ for $x \rightarrow \infty$. The Box-Cox distribution nests the GEVD and the GPD (see Rocco, 2014). For $\lambda = 1$, we immediately recover the GEVD, being $BC_{\xi,\gamma,\delta,1}(x) = H_{\xi,\gamma,\delta}(x)$. For $\lambda \rightarrow 0$, we obtain the GPD,⁷ being $BC_{\xi,\gamma,\delta,\lambda}(x) \rightarrow G_{\xi,\gamma,\delta}(x) = G_{\xi,\delta}(x - \gamma)$, where $x - \gamma$ are the excesses over the threshold γ .

In analogy to the POT method of Section 2.2.2, we can derive a quantile function which directly yields the Box-Cox VaR as

$$\text{VaR}_{t+1,\alpha}^{\text{BC}} = \gamma + \frac{\delta}{\xi} \left(\left(-\frac{1}{\lambda} \ln \left(1 - \frac{(1-\alpha)\lambda}{q} \right) \right)^{-\xi} - 1 \right) \quad (13)$$

for $\lambda \in [0, 1]$, where q is the ratio of the number of extremes to the total number of data points. Bali (2007) uses this VaR technique for (i) extremes defined as excesses over a threshold given by $q = 10\%$, (ii) extremes over several trading days (i.e., 23 days or 1 month) and (iii) extremes derived from the ‘mean excess function’ method described in Embrechts and Mikosch (1997). We follow this approach, however, we report only the results for variant (i) because, as in Bali (2007), the qualitative results turn out to be very similar for the alternatives.⁸

Because, in our application, the maximum likelihood method does not generally converge for the Box-Cox distribution, we use a non-linear regression technique to estimate its parameters. This approach is based on the idea of a Q-Q plot and is implemented as follows (see Tsay, 2005). First, we sort the extremes in ascending order $x_1 \leq \dots \leq x_i \leq \dots \leq x_n$. Using an important property of order statistics (see Gumbel, 1958; Cox and Hinkley, 1974), we have $E(BC_{\xi,\gamma,\delta,\lambda}(x_i)) = i/(n+1)$ for $i = 1, \dots, n$. Furthermore, we have $BC_{\xi,\gamma,\delta,\lambda}(x_i)$ from (12). Consequently, approximating expectation by an observed value, we can equate both expressions:

$$\frac{i}{n+1} = \left(\frac{\left(\exp \left(- \left(1 + \xi \left(\frac{x_i - \gamma}{\delta} \right) \right)^{-1/\xi} \right) \right)^\lambda - 1}{\lambda} \right) + 1. \quad (14)$$

After some algebra, adding the deviation ω_i between the previous two quantities and assuming that ω_i is not serially correlated, we can come up with a regression setup

$$\ln \left(\left(-\frac{1}{\lambda} \right) \ln \left(1 + \lambda \left(\frac{i}{n+1} - 1 \right) \right) \right) = -\frac{1}{\xi} \ln \left(1 + \xi \left(\frac{x_i - \gamma}{\delta} \right) \right) + \omega_i. \quad (15)$$

The least squares estimates of γ , δ and ξ can be obtained by minimizing the sum of squares of ω_i . Intuitively, this means that we seek to minimise the distance between the diagonal in the Q-Q plot and the quantile points plotted there because the closer the points are to the diagonal, the more accurately the theoretical distribution can describe the empirical data. This procedure produces consistent but less efficient estimates than the maximum likelihood method.

2.2.4 L-moment method

Given the results of Smith (1984, 1987) estimation of the GPD parameters in the POT method is typically done by maximum likelihood. As most of the useful properties of maximum likelihood only hold asymptotically and thus its effectiveness can be limited in small samples, Martins-Filho and Yao (2006) propose to estimate ξ and δ based on the L-moment theory (LMT) of Hosking (1990) and Hosking and Wallis (1997). As we will see, L-moment estimators exhibit a number of desirable properties which make them preferable to maximum likelihood in some applications.

⁷This is because $\lim_{\lambda \rightarrow 0} \frac{x^\lambda - 1}{\lambda} = \ln x$.

⁸The detailed results for the variants (ii) and (iii) are available upon request.

To illustrate L-moment estimation, recall that, for a random variable Y with quantile $Q(\alpha)$, the L-moment of order $r = 1, 2, \dots$ is defined as

$$\varphi_r = \int_0^1 Q(\alpha) P_{r-1}(\alpha) d\alpha, \quad (16)$$

where $P_r(\alpha) = \sum_{k=0}^r p_{r,k} \alpha^k$ and $p_{r,k} = ((-1)^{r-k} (r+k)!)/((k!)^2 (r-k)!)$, which contrasts with the traditional raw moments $\phi_r = \int_0^1 Q(\alpha)^r d\alpha$ involving powers of the quantile function.

Hosking (1990) shows that a distribution can be characterised by its L-moments even if raw moments of order greater than one do not exist and, most importantly, this characterisation is unique, which is not true for raw moments. Because $\varphi_1 = \phi_1$, the first L-moment (when it exists) provides the traditionally used measure of location of a distribution. Up to a scalar, φ_2 is the expectation of Gini's mean difference statistic, therefore representing a measure of scale that differs from the traditional variance by placing smaller weights on differences between realisations of the random variable (see Hosking and Wallis, 1997). If ϕ_1 exists, then $\tau_3 = \frac{\varphi_3}{\varphi_2}$ with $-1 < \tau_3 < 1$ and $\tau_3 = 0$ for symmetric distributions gives a measure of skewness that is less sensitive to the extreme tails of the distribution than the traditional measure of skewness (see Hosking, 1989). Similarly, $\tau_4 = \frac{\varphi_4}{\varphi_2}$ with $-1 < \tau_4 < 1$ is as a robust measure of kurtosis (see Oja, 1981).

L-moments can be used to estimate a finite number of parameters θ that identify a member of a family of distributions (see Martins-Filho and Yao, 2006). Suppose $F(\theta)$ is a family of distributions known up to θ . A sample $y_t, t = 1, \dots, T$, is available and the objective is to estimate θ . Since $\varphi_r, r = 1, 2, \dots$, uniquely characterises F , θ may be expressed as a function of φ_r . Hence, if estimators $\hat{\varphi}_r$ are available, we may obtain $\hat{\theta}(\hat{\varphi}_1, \hat{\varphi}_2, \dots)$.

From (16) we get $\varphi_{r+1} = \sum_{k=0}^r p_{r,k} \nu_k$, where $\nu_k = \int_0^1 Q(\alpha) \alpha^k d\alpha$ for $r = 0, 1, 2, \dots$. Given the sample, we define $y_{k,T}$ to be the k th smallest element of the sample, such that $y_{1,T} \leq y_{2,T} \leq \dots \leq y_{T,T}$. As shown by Landwehr et al. (1979) and Wang (1990), an unbiased estimator of ν_k is

$$\hat{\nu}_k = T^{-1} \sum_{j=k+1}^T \frac{(j-1)(j-2)\dots(j-k)}{(T-1)(T-2)\dots(T-k)} y_{j,T}, \quad (17)$$

which can be used to obtain the estimate $\hat{\varphi}_{r+1} = \sum_{k=0}^r p_{r,k} \hat{\nu}_k$ for $r = 0, 1, \dots, T-1$. For the GPD with $\theta = (\xi, \delta)$ it can be shown that $\delta = (1 - \xi)\varphi_1$ and $\xi = 2 - \varphi_1/\varphi_2$. Thus, their L-moment estimators are $\hat{\delta} = (1 - \hat{\xi})\hat{\varphi}_1$ and $\hat{\xi} = 2 - \hat{\varphi}_1/\hat{\varphi}_2$.

Similar to maximum likelihood estimators, L-moment parameter estimators are asymptotically normally distributed. However, they are easier and faster to implement because no numerical optimisation or iterative procedure is required. Furthermore, in finite samples, L-moment estimators outperform maximum likelihood methods in terms of mean squared error (see Hosking et al., 1985; Vogel and Fennessey, 1993).

2.2.5 Johnson method

Finally, in a last VaR estimation method, Brooks et al. (2005) suggest that, in contrast to most EVT approaches focussing exclusively on the tails, a valid approach should make use of the information from both the tails and the centre of the distribution, but treat each separately because extremes are structurally different from the return generating process under normal market conditions. They show that their approach is superior to three tail index estimation methods (the large sample estimators of Hill, 1975; Pickands, 1975; De Haan and Resnick, 1980) used in the early EVT literature and a more recent one (the small sample bias-corrected estimator of Huisman et al., 2001) and that it is considerably closer to the nominal probability of violations than competing

approaches which fit a single model for the whole distribution (bootstrapped historical simulation and bootstrapped GARCH(1,1)).⁹

The general idea of the Brooks et al. (2005) approach is to (i) use a structured Monte Carlo simulation which differentiates between the center and the tails of the empirical distribution, (ii) fit an adequate member of the Johnson (1949) system of distributions (JSD) to the simulated data and then (iii) obtain a VaR estimate based on the theoretical characteristics of this distribution.

Specifically, the simulation is performed as follows. In a first step, a GPD is fitted to each of the tails $x_t > x_u$ and $x_t < x_l$ of a given empirical distribution, where x_u and x_l are the thresholds for the upper and lower tails, respectively. In a second step, the POT method is used to determine the quantiles Q_u and Q_l for given probabilities α and $1 - \alpha$. Finally, the simulation is performed by Efron (1982) bootstrapping from both the two fitted tails and from the empirical distribution function derived from the returns. Specifically, one return path is simulated as follows. (i) Draw a sample of size T , with replacement, from the $x_t, t = 1, \dots, T$ observations of the empirical distribution. Then, within the newly generated, data do the following. (ii) If $x_t < Q_l$ then draw from the GPD fitted to the lower tail. (iii) If $x_t > Q_u$, then draw from the GPD fitted to the upper tail.¹⁰ (iv) However, if x_t falls in the middle of the empirical distribution, i.e., $Q_l \leq x_t \leq Q_u$, then x_t is retained. This simulation can be repeated 10,000 times to come up with 10,000 series $\{x_1^l, \dots, x_{n_l}^l, x_1^u, \dots, x_{n_u}^u\}$ with $n_l + n_u = T$.

Based on the simulated data, one might think of estimating the VaR by fitting the JSD to each simulated series, obtaining a VaR estimate for each series and generating a final forecast $VaR_{t+1, \alpha}^{JSD}$ by averaging the VaR estimates. Instead, motivated by the work of Hsieh (1993), Brooks et al. (2005) determine the lowest log price ratio $x_L = \ln(P_{min}/P_0)$ for each return path and model the resulting distribution, i.e., the sampling distribution of the maximum loss. Because its specific distributional form is unknown, they use the fact that theoretically any non-normal continuous distribution can be transformed into a standard normal distribution (see Baker, 1934). One way to do this is via the JSD.

Johnson (1949) shows that a continuous random variable X with an unknown distribution can be transformed to a standard normal variable Z by

$$Z = \varpi + \varsigma \tilde{f} \left(\frac{X - \gamma}{\delta} \right), \quad (18)$$

where $\tilde{f}(\cdot)$ is a transformation function. ϖ and $\varsigma > 0$ determine the shape of the distribution of X . γ is the location parameter and $\delta > 0$ is the scale parameter. A collection of suitable transformation functions forms the JSD as

$$\begin{aligned} S_{log} : Z &= \varpi + \varsigma \ln(X - \gamma) && \text{for } X > \gamma \\ S_{unb} : Z &= \varpi + \varsigma \sinh^{-1} \left(\frac{X - \gamma}{\delta} \right) && \text{for } -\infty < X < \infty, \\ S_{bou} : Z &= \varpi + \varsigma \ln \left(\frac{X - \gamma}{\gamma + \delta - X} \right) && \text{for } \gamma < X < \gamma + \delta \end{aligned}$$

where S refers to a sub-system and the subscript to the nature of the range. That is, the system S_{log} covers the lognormal family. S_{unb} covers the unbounded system of distributions (e.g., the Student-t distribution) and S_{bou} covers bounded distributions (e.g., the gamma or beta distribution).

If we have a statistic whose first four moments are known either exactly or approximately, then the JSD density with equivalent first four moments (mean, variance, skewness, kurtosis) may be used as an approximant (see Bowman and Shenton, 2004). This means that we can use

⁹Random GARCH returns have been produced by estimating the model, drawing randomly (with replacement) from the sample standardised residuals and then using the GARCH equations to construct a simulated return path.

¹⁰The reason why the GPD distribution is used for the tails rather than the empirical distribution throughout is that the number of observations in the tails may be insufficient to obtain accurate results without using an appropriate fitted distribution.

$$x_{1-\alpha} = \begin{cases} \gamma + \exp\left(\frac{z_{1-\alpha}-\varpi}{\zeta}\right) & \text{for } S_{log} \\ \gamma + \delta \sinh\left(\frac{z_{1-\alpha}-\varpi}{\zeta}\right) & \text{for } S_{unb} \\ \frac{\gamma+(\gamma+\delta)\exp\left(\frac{z_{1-\alpha}-\varpi}{\zeta}\right)}{1+\exp\left(\frac{z_{1-\alpha}-\varpi}{\zeta}\right)} & \text{for } S_{bou} \end{cases} \quad (19)$$

where $z_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of the standard normal distribution, to obtain four-moment approximations for the quantile $x_{1-\alpha}$. The method of moments is a natural candidate to estimate the JSD parameters for each sub-system (see Florence, 2011).

Brooks et al. (2005) indicate that, in empirical data, the distribution of x_L is best matched by the unbounded system. For this case, they also show that the maximum loss, given a confidence level α , can be obtained as

$$VaR_{t+1,\alpha}^{JSD,S_{unb}} = 1 - \exp((\pm x_{L,1-\alpha} \cdot \sigma_L) + \mu_L), \quad (20)$$

where $x_{L,1-\alpha}$ is calculated as in (19) and μ_L and σ_L are the mean and the standard deviation of the simulated distribution of x_L , respectively.

2.3 Filter processes

2.3.1 GARCH filtering

Even though the VaR estimation techniques in Sections 2.1 and 2.2 are based on sound statistical theory, a critical point remains. They all assume that returns are iid which is in sharp contrast to the empirical observations of serial correlation and conditional heteroskedasticity in most financial data (see Campbell et al., 1993; Engle and Patton, 2001). Thus, the unconditional VaR estimates produced by these methods may not provide an accurate characterisation of the actual VaR.

This problem is partially attenuated by the results of Leadbetter et al. (1983), Resnick (1987) and Castillo (1988). They show that unconditional VaR approaches based on EVT are not necessarily distorted because even if the underlying series exhibits a dependence structure, i.e., follows a moving average (MA), autoregressive (AR) or ARMA process, the limit distribution of extremes belongs to the domain of attraction of standard EVT distributions. Thus, as long as the underlying series is stationary and follows an MA, AR or ARMA process, then the conventional procedures developed for the iid data perform adequately on data with dependence structures.

As these findings do not fully resolve the potential problems with unconditional approaches, it has become standard in the empirical literature to carry out extremal analysis on returns which have been standardised by a conditional mean-volatility model and are thus approximately iid if the model is correctly specified. As outlined by Diebold et al. (2000), this approach has two steps. In the first stage, we fit a Bollerslev (1986) type generalised autoregressive conditional heteroskedasticity (GARCH) model to the returns via maximum likelihood. Most commonly, researchers and practitioners use a parsimonious AR(1)-GARCH(1,1) specification of the form

$$x_t = \kappa_0 + \kappa_1 x_{t-1} + z_t \sigma_t = \mu_t + z_t \sigma_t \quad (21)$$

$$\sigma_t^2 = \tau_0 + \tau_1 z_{t-1}^2 \sigma_{t-1}^2 + \tau_2 \sigma_{t-1}^2 \quad (22)$$

where z_t is a normally-distributed iid random variable with zero mean and unit variance, μ_t is the conditional mean, and σ_t^2 is the conditional variance of returns based on the information set up to time $t - 1$. With the estimated model parameters, we can (i) obtain the model residuals or standardised returns $s_t = (x_t - \mu_t)/\sigma_t$ and (ii) generate out-of-sample forecasts of the conditional mean μ_{t+1} and the conditional volatility σ_{t+1} .

In the second step, we calculate the unconditional VaR of the standardised returns $VaR_\alpha^{uc}(s_t)$. Depending on which VaR approach we are focussing on, this computation is different. For example,

for the historical simulation, the peak-over-threshold method and the Box-Cox approach, we use (3), (11) and (13), respectively (see McNeil and Frey, 2000; Bali, 2007; Marimoutou et al., 2009).¹¹ Together with the forecasts of the first step, we can plug the unconditional VaR estimate into

$$\text{VaR}_{t+1,\alpha} = \mu_{t+1} + \sigma_{t+1} \text{VaR}_{\alpha}^{uc}(s_t) \quad (23)$$

to generate a conditional VaR estimate which is consistent with the current state of the market. Various studies have shown that such conditional forecasts are more precise than their unconditional counterparts (see Barone-Adesi and Giannopoulos, 2001; McNeil and Frey, 2000).

With respect to this specification, several issues are worth noting. First, the low GARCH orders are typically motivated by the observation that higher orders usually do not significantly increase the explanatory power of the model (see Bollerslev et al., 1992). Second, several studies have shown that filters like (21) and more complex specifications (other variants of variance equations and distributional assumptions) significantly diminish intertemporal dependencies (see Bali, 2007). However, they also show that, for many financial series, we still observe significant Ljung and Box (1978) Q-statistics for the fitted models, indicating that the standardised residuals of a GARCH process are not fully iid (see Bollerslev et al., 1994; Bali and Theodossiou, 2007). Nonetheless, researchers continue to use these models because of a lack of alternatives that can reliably provide iid series. As our objective is not to find the ideal filter but to analyse the forecasting power of methods as they are typically used, we focus on the specification described above.¹²

2.3.2 Nonlinear filtering

While most of our approaches use a GARCH filter in their conditional versions, the method of Martins-Filho and Yao (2006) differs with respect to the filter model used for the returns. They propose a nonparametric Markov chain model (see Härdle and Tsybakov, 1997) as well as an improved nonparametric estimation procedure for the conditional volatility of returns (see Fan and Yao, 1998). The idea is to come up with a model that is general enough to accommodate nonlinearities that have been regularly verified in empirical work (see Tauchen, 2001; Patton, 2004) and thus generates a sequence of residuals adequate for EVT approaches. The nonlinear model is more difficult to estimate than classic parametric models but there can be substantial inferential gains if the alternative parametric models are misspecified or unduly restrictive.

Martins-Filho and Yao (2006) use the following model. The two dimensional strictly stationary process $\{(X_t, X_{t-1})'\}$ is described by a Markov chain of order one, i.e.,

$$X_t = \mu(X_{t-1}) + \sigma(X_{t-1})\epsilon_t, \quad (24)$$

where $\mu(x)$ and $\sigma^2(x)$ are the conditional mean $E(X_t|X_{t-1} = x)$ and variance function $E((X_t - \mu(x))^2|X_{t-1} = x)$ and ϵ_t is iid with zero mean, unit variance and unknown marginal distribution. This is the conditional heteroskedastic autoregressive nonlinear (CHARN) model of Diebold and Guégan (1993), Härdle and Tsybakov (1997) and Hafner (1998). It is more flexible than the popular GARCH(1,1) model in that $\mu(x)$ is a nonparametric function and, most importantly, $\sigma^2(x)$ is not a linear function of X_{t-1}^2 . However, note that its Markov property restricts its ability to effectively model longer memory that is commonly observed in return processes.

To estimate $\mu(x)$ and $\sigma(x)$ in (24), Martins-Filho and Yao (2006) suggest to use a procedure proposed by Fan and Yao (1998),¹³ which consists of two steps. First, $\mu(x)$ is estimated using the

¹¹In the case of the traditional approaches based on specific distributional assumptions, we use the quantile of the theoretical distribution (standard normal, Student-t) instead of (1) and (2).

¹²In our robustness checks in Section 5.2, we summarise the results for more general GARCH variants.

¹³This is because the approach suggested by Härdle and Tsybakov (1997) has several undesirable properties. For example, the procedure for estimating conditional variance suffers from significant bias and does not produce estimates that are constrained to be positive. Furthermore, it is sensitive to how well $\mu(x)$ is estimated.

local estimator of Fan (1992). Let $A(x)$ and $B(x)$ be Gaussian kernel functions, x in the support of the conditional density of X , and $\tilde{a}(T), \tilde{b}(T)$ be sequences of positive real numbers (bandwidths) such that $\tilde{a}(T), \tilde{b}(T) \rightarrow 0$ as the sample size $T \rightarrow \infty$. Then with

$$(\hat{\varrho}, \hat{\rho})' = \arg \min_{\varrho, \rho} \sum_{t=2}^T (x_t - \varrho - \rho(x_{t-1} - x))^2 A\left(\frac{x_{t-1} - x}{\tilde{a}(T)}\right) \quad (25)$$

the local estimator of $\mu(x)$ is $\hat{\varrho}(x)$. In the second step, let $\bar{x}_t = (x_t - \hat{\varrho}(x_{t-1}))^2$ and define

$$(\hat{\eta}, \hat{\zeta})' = \arg \min_{\eta, \zeta} \sum_{t=2}^T (\bar{x}_t - \eta - \zeta(x_{t-1} - x))^2 B\left(\frac{x_{t-1} - x}{\tilde{b}(T)}\right), \quad (26)$$

such that the local estimator of $\sigma^2(x)$ is $\hat{\eta}(x)$. The bandwidths are selected using the data driven method of Ruppert and Wand (1995) which is a consistent estimator of the (optimal) bandwidth sequence that minimises the asymptotic mean integrated squared error of $\hat{\varrho}(x)$ and $\hat{\eta}(x)$.

Thus, we can estimate the conditional VaR similar to (23) with the only difference that the residuals and the forecasts for conditional mean and volatility come from a new model. Martins-Filho and Yao (2006) show that using the alternative filter model and the L-moment approach of Section 2.2.4 outperforms the approach of McNeil and Frey (2000), which uses a GARCH filter and the classic maximum likelihood method for the GPD fit, in terms of mean squared error in simulation and backtesting experiments for stock market series.¹⁴

3 Data and preliminary results

To examine the accuracy of different VaR forecasting models, it has become standard to compare their predictions with actual return realisations following the prediction. To this end, most studies focus on one or more representative time series for the stock market. For example, Kuester et al. (2006), Martins-Filho and Yao (2006) and Bali et al. (2008) use the NASDAQ Composite Index, the S&P 500 and the CRSP Value-Weighted Index, respectively. Because one may argue that the results for one specific asset class cannot be generalised (a good performance for stock returns may not imply comparable performance for commodities, bonds or currencies), we combine the ideas outlined in Brooks et al. (2005), Bali (2007) and Cheng and Hung (2011), that is, we focus on several asset classes by conducting a multi-asset study. Specifically, we use the CRSP Value-Weighted Index and the S&P 500 Index to represent the stock market. Furthermore, we have a look at the Goldman Sachs Commodity Index, which represents a futures-based total return investment in a rich commodity basket, and a spot investment in the gold market by using London Bullion Market quotations. Finally, we include the Datastream US Government Bond Index and the US-Dollar to Great Britain Pound exchange rate to cover bond and currency markets.

Our dataset consists of daily closing prices (or index levels) from January 1, 1986 to February 29, 2016, which leads to a sample size of 7,869 observations for each series. The CRSP index was obtained from the data library of Kenneth French,¹⁵ whereas the other data was collected from Thomson Reuters Datastream.¹⁶ The series are used to calculate percentage log returns $x_t = [\ln(P_t) - \ln(P_{t-1})] \cdot 100$, where P_t is the closing price (or index level) at time t .

Table 1 presents the descriptive statistics for our return sample. Besides minimum, maximum, mean, standard deviation, skewness and kurtosis of returns, we also report the Q-statistics to test the null hypothesis that there is no autocorrelation in the returns and squared returns. For

¹⁴In the simulation study, they use an interesting stochastic process with Hansen (1994) skewed-t errors, for which the true VaR can be directly calculated. Thus, in repeated sampling, they can answer the question of how close the estimates of different VaR methods are to the true VaR.

¹⁵See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

¹⁶The codes of the series are S&PCOMP, GSCITOT, GOLDBLN, BMUS10Y and BOECGBP.

our sample period, we can observe the highest mean returns in the stock indices followed by the commodity investments and the bond index. The $\$/\pounds$ exchange rate shows a negative mean return. Volatilities are highest for the commodity index and lowest for the bond index. As far as skewness is concerned, we obtain negative values for all series. While the most negative values occur in the stock market, the bond market does not deviate strongly from a normal distribution in terms of skewness. Similarly, the stock market shows crucial kurtosis which is significantly higher than for all other series. In general, the fat-tail property is more dominant than skewness in the sample.¹⁷

The Q-statistics for the cumulative effect of up to twelfth-order autocorrelation are highly significant and indicate the presence of intertemporal dependencies in the first and second moments of the return distributions.¹⁸ This implies that the returns are not iid and thus using filtered returns may significantly improve the forecasting performance of our VaR estimators.

	CRSP	SP500	GSCI	GOLD	USGBI	$\$/\pounds$
Min	-17.413	-22.900	-18.432	-10.162	-2.876	-3.960
Max	11.354	10.957	7.600	7.382	4.061	4.429
Mean	0.044	0.028	0.011	0.017	0.005	-0.001
Std Dev	1.132	1.147	1.290	0.984	0.463	0.590
Skewness	-0.648	-1.280	-0.568	-0.358	-0.044	-0.136
Kurtosis	18.350	31.294	11.849	10.601	6.437	6.517
Q(12)	42.289	44.198	13.308	30.724	23.547	38.930
Q ² (12)	2,988.763	1,345.942	1,114.469	919.475	769.190	1,598.932

For the period from January 1, 1986 to February 29, 2016 and our six time series, this table reports the minimum, maximum, mean, standard deviation, skewness and kurtosis of daily percentage log returns. Furthermore, we display the Ljung-Box Q-statistics with 12 lags for the returns and the squared returns. It tests the null hypothesis that there is no autocorrelation up to lag 12. The time series are abbreviated as follows: CRSP = CRSP Value-Weighted Index, SP500 = S&P 500 Index, GSCI = Goldman Sachs Commodity Index (futures-based total return version), GOLD = London Bullion Market gold spot price, USGBI = Datastream US Government Bond Index, $\$/\pounds$ = US-Dollar to Great Britain Pound exchange rate.

Table 1: Descriptive statistics

To get a first impression of the accuracy of our VaR forecasting models, we have a descriptive look at their out-of-sample performance. In contrast to in-sample performance, it provides a more practically relevant perspective on forecasting abilities because risk managers by definition obtain VaR estimates in real time and hence must use parameters obtained from an already observed sample to evaluate risks associated with current and future movements in risk factors. Hence, the true quality of a VaR approach can only be revealed by its performance outside the sample, i.e., in data not used in the estimation of the underlying parameters (see Angelidis et al., 2004).

To capture this idea of performance evaluation, we perform a rolling window analysis from which we obtain violation rates (see Christoffersen, 1998; Berkowitz and O'Brien, 2002). We define a rolling window of 1,000 daily returns (roughly four years of trading data) which we move in steps of one day from the beginning to the end of our sample. This implies that we allow the corresponding parameters to change over time. The choice of window length is motivated by prior studies on EVT-based VaR methods (see McNeil and Frey, 2000) and by the fact that this number is a minimum requirement for adequate filter performance (see Hwang and Vallis Pereira, 2006). Within each window, we calculate our VaR estimators (as specified in Section 2) for confidence levels of 95%, 97.5% und 99%, i.e., nominal error probabilities of 5%, 2.5% and 1%, and compare the VaR estimates to the actual return on the day following the window. That is, as in Kuester

¹⁷The fact that heavy-tailed distributions may not possess low-order moments implies that usual significance tests for skewness and kurtosis are most likely unreliable and are not worth reporting (see Paoletta, 2001).

¹⁸Note that the statistics take high values partially because of the large sample size (see Bali, 2007).

et al. (2006) and Bali (2007), we are interested in the performance of one-step-ahead forecasts for the VaR.¹⁹ More formally, we calculate a violation indicator

$$I_t = \begin{cases} 1 & \text{if } x_t < VaR_{t,\alpha} \\ 0 & \text{else} \end{cases} \quad (27)$$

which takes the value 1 if the realised return exceeds the VaR forecast and is 0 otherwise. To obtain the violation rate \hat{p} , we divide the number of ones (over all windows) by the total number of windows in our sample.²⁰ Then \hat{p} can be compared to the nominal error probability $\pi = (1 - \alpha)$ used in VaR estimation. For example, if the error probability is set to be $\pi = 5\%$, then $\hat{p} = 7\%$ ($\hat{p} = 3\%$) implies that our VaR estimator predicts that the calculated VaR will only be exceeded with a probability of 5% but, in fact, it is exceeded more (less) frequently. While a good VaR model would be one that generated a \hat{p} close to π , a situation $\hat{p} > \pi$, where the true risk is underestimated, is likely to be far more serious for a firm (because risk management activities based on the VaR estimates may lead to insufficient hedges) than having too much capital reserves in the case of $\hat{p} < \pi$ indicating an overestimated VaR (see Brooks et al., 2005; Ziggel et al., 2014).^{21,22}

Table 2 reports the violation rates for the unconditional versions of our EVT-based VaR estimators and, for comparison, also for selected traditional approaches. These rates are based on $7,869 - 1,000 = 6,869$ VaR forecasts. Starting with a look at the 5% error probability and the traditional methods, we can observe violation rates above 5% for all series when the Student-t method is used. With the exception of the currency market, the normal and the historical simulation methods also have a tendency to underestimate risk.²³ Turning to the EVT approaches, the peak over threshold method and the L-moment method tend to deliver higher (lower) violation rates than historical simulation in stock and bond markets (currency markets). The Box-Cox method is, with the exception of the exchange rate, superior to these two EVT approaches and to the traditional methods because its violation rates are closer to the nominal error probability. The Johnson method provides violation rates below 5% for all time series and is thus the most conservative approach. For the least volatile series (bond and currency market), the difference from 5% is most substantial.

As far as the 2.5% error probability is concerned, the picture becomes clearer. The violation rates are highest for the normal and Student-t approaches and they are consequently above 2.5% for all series. Interestingly, despite the conceptual differences of the methods, the violation rates for the simple historical simulation and the more complex peak over threshold method are very similar. Both methods and also the L-moment approach deliver violation rates above 2.5% for all series. Finally, the Box-Cox method overestimates risk for some series (both stock indices and gold) and the Johnson method overestimates for all series.

Finally, for the 1% level, we arrive at conclusions similar to the 2.5% case. However, now we can detect fewer instances of overestimation for the Box-Cox approach (one vs. three) and the Johnson method (five vs. six).

Table 3 allows a more compact comparison of the unconditional VaR estimators because it summarises the violations across all error probabilities and time series by reporting the mean

¹⁹We leave the performance of multi-step-ahead forecasts for future research because problems with the square-root-of-time scaling rule and related techniques must to be resolved first (see McNeil and Frey, 2000).

²⁰Brooks et al. (2005) use only one out-of-sample period of 250 days and calculate the percentage of days for which the VaRs were exceeded by actual trading losses. Bali (2007) defines a 10-year rolling sample (in one-year increments) to estimate parameters and sets a one-year holdout sample (subsequent to the estimation) to evaluate performance. Kuuster et al. (2006) use our approach.

²¹Of course, in a regime of negative interest rates, this point of view can change because then reserves are subject to capital depreciation.

²²Underpredictions have potentially serious solvency implications in the context of futures margin systems because margin setting is known to be sensitive to the occurrence of large price changes (see Brooks et al., 2005).

²³This is consistent with the results of Kuuster et al. (2006) for the NASDAQ Composite index.

	CRSP	SP500	GSCI	GOLD	USGBI	\$/£
<i>5% level</i>						
Normal	5.38	5.24	5.80	5.13	5.24	4.89
Student-t	5.91	5.61	6.46	5.58	5.88	5.40
Historical simulation	5.58	5.45	5.66	5.40	5.08	4.82
Peak over threshold	5.68	5.58	5.85	5.33	5.20	4.79
L-moment	5.81	5.65	5.78	5.50	5.23	4.70
Box-Cox	5.03	5.10	5.71	4.85	5.05	4.76
Johnson	4.38	4.72	4.32	4.35	3.64	3.83
<i>2.5% level</i>						
Normal	3.71	3.60	3.77	3.33	3.29	3.16
Student-t	3.56	3.34	3.62	3.23	3.16	3.06
Historical simulation	3.05	2.85	3.16	2.68	2.80	2.55
Peak over threshold	3.04	2.88	3.26	2.66	2.77	2.65
L-moment	3.10	2.92	3.25	2.75	2.72	2.62
Box-Cox	2.28	2.15	2.93	2.20	2.59	2.55
Johnson	2.23	2.32	2.36	2.21	2.01	1.89
<i>1% level</i>						
Normal	2.23	2.21	2.37	2.22	2.17	1.86
Student-t	1.70	1.67	1.57	1.54	1.38	1.15
Historical simulation	1.41	1.35	1.54	1.14	1.35	1.14
Peak over threshold	1.35	1.28	1.53	1.21	1.25	1.11
L-moment	1.40	1.33	1.50	1.18	1.30	1.08
Box-Cox	1.08	1.00	1.21	0.82	1.06	1.03
Johnson	0.79	0.92	1.05	0.67	0.77	0.55

For our unconditional VaR estimators (i.e., the approaches of Sections 2.1 and 2.2), three error probabilities and six time series, this table reports the percentage violation rates obtained from a rolling window approach in which the VaR estimates based on 1,000 day windows are compared to actual return realisations on the day following the windows. Series abbreviations are used as in Table 1.

Table 2: Violation rates for unconditional VaR estimators

	ND	ST	HS	POT	LM	BC	JO
MAPE	54.36	33.14	17.98	17.80	18.82	7.73	17.19
Rank	7	6	4	3	5	1	2
MPE	54.11	33.14	17.58	17.33	18.16	1.00	-16.63
Rank	7	6	4	3	5	2	1

For each of our unconditional VaR estimators, this table reports the mean absolute percentage errors (MAPE) and mean percentage errors (MPE) across the three error probabilities and six time series in Table 2. Furthermore, it presents rankings of the VaR estimators, where rank 1 (rank 7) is the best (worst) methodology in terms of MAPE or MPE. The methodologies are abbreviated as follows: ND = normal distribution, ST = Student-t, HS = historical simulation, POT = peak over threshold, LM = L-moments, BC = Box-Cox, JO = Johnson.

Table 3: Violation summary for unconditional VaR estimators

absolute percentage error $MAPE = \frac{1}{18} \sum_{j=1}^6 \sum_{k=1}^3 |(\pi_k - \hat{p}_{jk})/\pi_k|$ and the mean percentage error $MPE = \frac{1}{18} \sum_{j=1}^6 \sum_{k=1}^3 (\pi_k - \hat{p}_{jk})/\pi_k$, where j (k) is an index for our six (three) time series (error probabilities). While MAPE reflects the overall deviations regardless of their direction, the sign of the MPE gives additional information on whether a VaR model has a tendency to underestimate (positive sign) or to overestimate (negative sign) risk. Besides MAPE and MPE, Table 3 contains a ranking of the VaR estimators based on MAPE and MPE from best (rank 1) to worst (rank 7), where in the case of MPE we assume that overestimation is preferable to underestimation.

Focussing on the MAPE first, the Box-Cox method appears to be superior in terms of absolute error. It is followed by the Johnson, peak over threshold, historical simulation and L-moments

techniques, which show similar error levels. Finally, the normal and Student-t approaches rank lowest. Turning to the MPE, we detect a similar ranking. However, now we can see that the Johnson method scores highest because it is the only estimator with an (on average) negative error. Furthermore, the Box-Cox method ranks second with a very low positive MPE. The MPE values for all other estimators crucially differ from these two methods.

So far, we have concentrated on unconditional approaches. If we repeat our violation analysis for the conditional versions of the VaR estimators, we obtain the results of Tables 4 and 5. We can gain several important insights. First, with a few exceptions, using filtered returns improves forecasting accuracy (i.e., reduces violation rates) over all error probabilities and time series. Second, historical simulation can keep up with the peak over threshold method and the L-moment method by delivering largely similar errors or partially even lower errors. Finally, while the L-moment method now appears to be preferable in terms of MAPE, the Box-Cox method and the Johnson method (i.e., their MPE) again show a tendency to overestimation and can thus be interpreted as the most conservative estimators.

	CRSP	SP500	GSCI	GOLD	USGBI	\$/£
<i>5% level</i>						
Normal	5.87	5.58	5.65	4.89	5.02	5.36
Student-t	6.69	6.51	6.44	5.46	5.81	6.09
Historical simulation	5.29	5.14	5.46	4.97	4.92	4.99
Peak over threshold	5.41	5.13	5.49	4.99	4.95	5.02
L-moment	5.33	5.11	5.45	5.05	5.04	4.98
Box-Cox	4.86	4.63	5.33	4.71	4.88	5.01
Johnson	3.47	3.73	3.52	3.73	3.36	3.57
<i>2.5% level</i>						
Normal	3.66	3.68	3.28	3.15	3.04	3.16
Student-t	3.47	3.54	3.07	3.03	2.84	3.01
Historical simulation	2.58	2.56	2.68	2.53	2.46	2.50
Peak over threshold	2.81	2.77	2.69	2.53	2.36	2.64
L-moment	2.78	2.81	2.71	2.52	2.43	2.55
Box-Cox	2.34	2.15	2.45	2.14	2.18	2.52
Johnson	1.58	1.79	1.67	1.82	1.50	1.56
<i>1% level</i>						
Normal	2.08	2.15	1.86	1.91	1.57	1.64
Student-t	1.38	1.43	1.09	1.28	1.09	0.98
Historical simulation	1.12	1.12	1.22	0.98	1.09	0.93
Peak over threshold	1.06	1.08	1.18	1.00	1.11	1.06
L-moment	1.00	1.05	1.16	0.99	1.14	0.96
Box-Cox	0.71	0.73	0.99	0.83	1.00	0.96
Johnson	0.67	0.61	0.70	0.61	0.66	0.41

For our conditional VaR estimators (i.e., the approaches of Sections 2.1 and 2.2 combined with the filters of Section 2.3), three error probabilities and six time series, this table reports the percentage violation rates obtained from a rolling window approach in which the VaR estimates based on 1,000 day windows are compared to actual return realisations on the day following the windows. Series abbreviations are used as in Table 1.

Table 4: Violation rates for conditional VaR estimators

4 Evaluating VaR estimators

4.1 General framework

To statistically back up and extend descriptive analyses like the one in Section 3, it has become standard to rely on the Christoffersen (1998) framework which is designed to evaluate out-of-sample VaR forecasts by focusing on ideal properties of the violation indicator (27) (see Kuuster et al.,

	ND	ST	HS	POT	LM	BC	JO
MAPE	42.87	23.74	5.54	6.36	5.42	8.53	33.87
Rank	7	5	2	3	1	4	6
MPE	42.62	23.52	4.10	5.60	4.51	-7.69	-33.87
Rank	7	6	3	5	4	2	1

For each of our conditional VaR estimators, this table reports the mean absolute percentage errors (MAPE) and mean percentage errors (MPE) across the three error probabilities and six time series in Table 4. Furthermore, it presents rankings of the VaR estimators, where rank 1 (rank 7) is the best (worst) methodology in terms of MAPE or MPE. Method abbreviations are used as in Table 3.

Table 5: Violation summary for conditional VaR estimators

2006; Bali et al., 2008; Candelon et al., 2011; Ziggel et al., 2014). In this framework, a sequence of VaR forecasts with specified nominal error probability π is said to be efficient with respect to the information set Ω_{t-1} if

$$E(I_t|\Omega_{t-1}) = \pi. \quad (28)$$

This implies that I_t has to be uncorrelated with any function of a variable in the information set available at $t - 1$. If the efficiency condition holds, then VaR violations will occur with the correct probability, and neither the forecast for $VaR_{t,\alpha}$, nor that for I_t could be improved.

Although a general test is desirable, we follow Kuester et al. (2006) by using intermediate statistics for testing specific implications of the general hypothesis (28), such that particular inadequacies of a model can be revealed.

4.2 Indicator information sets

To make the criterion (28) operational, Christoffersen (1998) considers an information set $\Omega_{t-1} = \{I_{t-1}, I_{t-2}, \dots, I_1\}$, which consists of past realisations for the indicator sequence, and shows that, in this case, testing $E(I_t|\Omega_{t-1}) = E(I_t|I_{t-1}, I_{t-2}, \dots, I_1) = \pi$, for all t , is equivalent to testing that the sequence $\{I_t\}$ is iid Bernoulli with parameter π . He calls a sequence of interval forecasts with $\{I_t\} \stackrel{iid}{\sim} Bern(\pi)$, for all t , to have correct conditional coverage and proposes a likelihood ratio testing framework to test this property. This framework consists of a test of unconditional coverage and a test of independence which can be combined to a test of conditional coverage. These tests are most frequently used in practice (see Basel Committee on Banking Supervision, 2011b).

Unconditional coverage test. To test for unconditional coverage, the hypothesis $H_0 : E(I_t) = \pi$ should be tested against the alternative $H_A : E(I_t) \neq \pi$ given independence (see Kupiec, 1995). As shown by Christoffersen (1998), the likelihood under this null hypothesis is simply $L(\pi; I_1, I_2, \dots, I_w) = \pi^{w_1}(1 - \pi)^{w_0}$ and under the alternative $L(p; I_1, I_2, \dots, I_w) = p^{w_1}(1 - p)^{w_0}$. Thus, testing for unconditional coverage can be performed by a standard likelihood ratio test

$$LR_{uc} = -2 \ln (L(\pi; I_1, I_2, \dots, I_w) / L(\hat{p}; I_1, I_2, \dots, I_w)) \stackrel{asy}{\sim} \chi^2(1), \quad (29)$$

where $\hat{p} = w_1 / (w_1 + w_0)$ is the maximum likelihood estimator of p and w_1 (w_0) is the number of ones (zeros) in the indicator sequence of length $w = w_1 + w_0$. As the test is a two-sided test, a model will be rejected if it generates too many or too few violations.

Independence test. VaR estimators that do not take temporal volatility dependence into account may be correct on unconditional coverage but will produce violation clusters, which cause incorrect conditional coverage (see Bali et al., 2008). As the test of unconditional coverage has no power against such an alternative, an additional test – a test of independence of violations – is required to fully judge the adequacy of VaR forecasts. For this purpose, various standard tests could

be used (see Campbell et al., 1997). Christoffersen (1998) proposes to test the null hypothesis of independence, i.e., that a violation today has no influence on the probability of violation tomorrow, against an explicit alternative, i.e., that $\{I_t\}$ follows a binary first-order Markov chain with the transition probability matrix

$$\Pi = \begin{pmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{pmatrix},$$

where $\pi_{ij} = P(I_t = j | I_{t-1} = i)$.²⁴ The approximate joint likelihood for this process, conditional on the first observation, is

$$L(\Pi; I_2, \dots, I_w | I_1) = (1 - \pi_{01})^{w_{00}} \pi_{01}^{w_{01}} (1 - \pi_{11})^{w_{10}} \pi_{11}^{w_{11}},$$

where w_{ij} is the number of indicator observations with value i followed by j and the maximum-likelihood estimators for the parameters are $\hat{\pi}_{01} = w_{01}/(w_{00} + w_{01})$ and $\hat{\pi}_{11} = w_{11}/(w_{10} + w_{11})$.

Under the null hypothesis of independence, we have $\pi_{01} = \pi_{11} =: \pi_1$, from which the conditional binomial joint likelihood follows as

$$L(\pi_1; I_2, \dots, I_w | I_1) = (1 - \pi_1)^{w_{00} + w_{10}} \pi_1^{w_{01} + w_{11}}$$

and the maximum-likelihood estimate is $\hat{\pi}_1 = (w_{01} + w_{11})/(w_{00} + w_{10} + w_{01} + w_{11})$. Based on the results of Hoel (1954), the likelihood ratio statistic is then given by

$$LR_{ind} = -2 \ln(L(\hat{\pi}_1; I_2, \dots, I_w | I_1) / L(\hat{\Pi}; I_2, \dots, I_w | I_1)) \stackrel{asy}{\sim} \chi^2(1). \quad (30)$$

Note that, because $\hat{\pi}_1$ is unconstrained, this test statistic tests only for independence and does not take correct coverage into account.

Conditional coverage test. An accurate VaR assessor should have the property to examine both, independence and coverage. Therefore, Christoffersen (1998) suggests to combine (29) and (30) to

$$LR_{cc} = -2 \ln(L(\pi; I_2, \dots, I_w | I_1) / L(\hat{\Pi}; I_2, \dots, I_w | I_1)) \stackrel{asy}{\sim} \chi^2(2) \quad (31)$$

to test for correct conditional coverage.²⁵ By conditioning on the first observation in (29), we have $LR_{cc} = LR_{uc} + LR_{ind}$, which provides a means to check in which regard the violation series $\{I_t\}$ fails the correct conditional coverage property. Thus, if the null hypothesis is rejected, we can look at the components of L_{cc} to answer the question of whether the rejection is caused by inaccurate coverage and/or by clustered violations. As argued by Campbell (2007) even in the case of not rejecting the null we should have a look at the components because the test result may disguise a violation of one of the properties. For example, consider the test statistics $L_{uc} = 7.4$ and $L_{ind} = 1.5$ such that $L_{cc} = 8.9$. At the 1% level, the critical values are $\chi^2(1) = 6.63$ and $\chi^2(2) = 9.21$. Thus, the conditional coverage null hypothesis wouldn't be rejected although the unconditional coverage null hypothesis is rejected.

As with all asymptotically motivated inferential procedures, the actual size of the tests for finite samples can deviate from their nominal sizes. Lopez (1997) examines the size of unconditional and conditional coverage tests via simulation, as well as their power against various model misspecifications. For a sample size of 500 observations, he finds both tests to be adequately sized. Furthermore, even for such a small sample, power appears to be reasonable. However, as Berkowitz (2001) and Berkowitz et al. (2011) provide evidence of low power in small samples for an extended

²⁴Berkowitz et al. (2011) summarises tests which focus on the duration between violations because, under the null that VaR forecasts are correctly specified, this duration should be completely unpredictable.

²⁵In effect, the null hypothesis of the unconditional coverage test will be tested against the alternative of the independence test.

set of alternatives, we follow Bali (2007) and minimise such potential problems by resorting to a large sample of daily data.

4.3 Extended information set

Because the conditional coverage test only considers past violations in the information set, it only tests a special case of conditional coverage, whereas the condition (28) suggests that any variable in Ω_{t-1} be uncorrelated with I_t . To address this limitation, Engle and Manganelli (2004) propose to add a judicious choice of additional variables to the information set. Specifically, they operationalise (28) by regressing I_t on past violations and the VaR estimate for period t . That is, they run the regression

$$I_t = \tilde{\pi} + \sum_{i=1}^m \beta_i I_{t-i} + \beta_{m+1} VaR_{t,\alpha} + \epsilon_t, \quad (32)$$

where, under the null hypothesis, $\tilde{\pi} = \pi$ and $\beta_i = 0$, $i = 1, \dots, m + 1$. In vector notation, we have

$$\mathbf{I} - \pi \boldsymbol{\iota} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (33)$$

where $\beta_0 = \tilde{\pi} - \pi$ and $\boldsymbol{\iota}$ is a conformable vector of ones. Under the null hypothesis (28), the regressors should not have explanatory power, that is, $H_0 : \boldsymbol{\beta} = \mathbf{0}$ (see Chen and Lu, 2012). Because the regressors are not correlated with the dependent variable under the null hypothesis, invoking a suitable central limit theorem yields

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \pi\boldsymbol{\iota}) \stackrel{asy}{\sim} N(\mathbf{0}, (\mathbf{X}'\mathbf{X})^{-1}\pi(\mathbf{1} - \pi)) \quad (34)$$

from which Engle and Manganelli (2004) deduce the test statistic

$$DQ = \frac{\hat{\boldsymbol{\beta}}'_{LS}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{LS}}{\pi(1 - \pi)} \stackrel{asy}{\sim} \chi^2(m + 2).$$

In our study, we follow Kuester et al. (2006) and Bali et al. (2008) by using a pure violation-based and a mixed specification of the dynamic quantile test. In the first, denoted DQ_{Hit} , the regressor matrix contains a constant and three lagged indicators I_{t-1}, \dots, I_{t-3} . For the second one, denoted by DQ_{VaR} , we additionally include the contemporaneous VaR estimate.²⁶

5 Empirical analysis

5.1 Main results

5.1.1 Unconditional approaches

Tables 6 to 8 report the test results for the unconditional VaR estimators, where each table concentrates on two of our six time series. We present the test statistics of the unconditional coverage test, the independence test, the unconditional coverage test and the two variants of the dynamic quantile test. Cases in which the null hypothesis cannot be rejected at the 1% level are marked in bold type because insignificance of these tests is typically interpreted as implying that a VaR approach provides a good assessment of the risk exposure of a given asset or portfolio (see Bali, 2007; Bali et al., 2008). Furthermore, we calculate some additional statistics. First, for each test and error probability, we rank the VaR estimators based on the test statistics from lowest (rank 1) to highest (rank 7). For each estimator and error probability, we then compute the

²⁶Thus, at a 1% level, the tests require the critical values $\chi^2(4) = 13.28$ and $\chi^2(5) = 15.09$, respectively (see Domitrescu et al., 2012).

average rank across all tests in order to be able to judge the relative performance of the estimators even in cases where they are rejected (or not rejected) by almost all tests. Here, lower ranks signal more favorable VaR estimators. Second, we obtain the loss measure of Lopez (1999). That is, we (i) define a loss function L_t which takes the value $1 + (r_t - VaR_{t,\alpha})^2$ if $r_t < VaR_{t,\alpha}$ and is zero otherwise and (ii) calculate the average of its realisations over time. For a good VaR model, this average should be close to zero. It also allows judging the relative performance but implicitly rewards VaR estimators for being conservative. Finally, as a last statistic, we report the (absolute) average VaR estimate for each methodology because based on this value we can directly assess differences in the VaR levels produced by the different estimators.

Starting with a look at the results for the CRSP Value-Weighted Index in Panel A of Table 6, we find that only the Box-Cox and the Johnson method pass the unconditional coverage test for all error probabilities. Because this test and a general nominal error level of 1% are implicitly incorporated in the framework for backtesting internal models proposed by the Basel Committee on Banking Supervision (1996b), our results imply that, from a regulatory perspective, only these two methods would be acceptably accurate and thus be used for obtaining the VaR estimates that have to be reported to the regulators who observe when actual portfolio losses exceed these estimates.²⁷ Of course this also means that, for all other approaches, the rejection of the null of unconditional coverage indicates that the computed VaR estimates are not accurate enough.

As discussed in Section 4, the quality of a VaR model does not depend only on hitting the given confidence level. It should also be able to produce violations independent of both past violations and other potential predictive variables. A look at the additional tests analysing these properties reveals that their null hypotheses are rejected for all VaR estimators and all nominal error probabilities. Thus, the use of the Box-Cox and the Johnson approaches is justified under the Basel framework but they cannot be considered accurate VaR estimation techniques. However, even though they are rejected by most tests, our average ranks and losses indicate that they are still more accurate than the alternative VaR estimators. In line with our overestimation tendency identified in Section 3, we can also observe higher VaR estimates on average in comparison to the other approaches.

The results for the S&P 500 Index in Panel B of Table 6 show that, while all VaR estimators pass the unconditional coverage test for the 5% error probability, only the peak over threshold, Box-Cox and Johnson methods pass this test when we switch to the 1% error probability. Interestingly, while, for the 1% error level, almost all VaR approaches are rejected based on the independence and conditional coverage test, this is not the case for the Box-Cox method. Thus, given a violation today there is no higher probability of a violation tomorrow. However, despite this promising result, the dynamic quantile test rejects all VaR estimators including the Box-Cox method.

Because the 1% error probability is the focus of interest in the Basel framework, we concentrate on this error level in the following discussion of our results for the commodity, bond and currency series (and also in the conditional analysis in Section 5.1.2). Turning to the commodity index in Panel A of Table 7, we can document results similar to the CRSP Value-Weighted Index. That is, (i) only the Box-Cox and Johnson approaches pass the unconditional coverage test, (ii) all VaR estimators fail the more advanced tests and (iii) the overall performance of the Box-Cox and Johnson method appears to be superior. For the gold investment in Panel B of Table 7, the results are slightly different. Historical simulation and all EVT-based methods (except for the Johnson method) pass the unconditional coverage test. The Box-Cox and the L-moment method

²⁷The Basel three-zone framework deems a VaR model acceptable (green zone) if the number of violations of the 99%-VaR is below the 1%-binomial 95% quantile. A model is disputable (yellow zone) up to the 99.99% quantile and is deemed seriously flawed (red zone) whenever more violations occur (see Kuester et al., 2006; Campbell, 2007). Thus, with the decision rule 'reject the null hypothesis of a valid VaR model whenever the model scores red', the procedure can be interpreted as a significance test, i.e., basically as a one-sided version of the unconditional coverage test (see Ziggel et al., 2014). Not-rejecting the two-sided version directly implies not-rejecting the one-sided version.

additionally pass the independence and conditional coverage test. However, all approaches are rejected based on the dynamic quantile test. Based on our testing ranks, the L-moment method scores first, the Box-Cox method second and historical simulation third.

Finally, while the results for the US-Dollar to Great Britain Pound exchange rate in Panel B of Table 8 also indicate that all methods fail the more advanced backtests but highlight superiority of the Box-Cox and Johnson method similar to our stock market results, the numbers in Panel A of Table 8 for the bond market are quite interesting. Here, we find several cases where the VaR methods are not rejected based on more advanced tests. Specifically, we see that the L-moment and the Box-Cox method pass all tests (except for the second dynamic quantile specification). Furthermore, the Johnson method cannot be rejected by any of the tests indicating a far better performance of this method in bond markets than in stock, commodity and currency markets.

We can conclude that none of the unconditional VaR models is acceptable for all asset classes and target probabilities. However, we find some evidence that (i) EVT-based approaches are superior to traditional ones and (ii) the Box-Cox and Johnson method are the most promising VaR estimators in the EVT-based class.

5.1.2 Conditional approaches

Given the relatively poor performance of most approaches in their unconditional versions, this section analyses the accuracy of their conditional counterparts which take into account the dynamic behaviour of asset returns over time. In a structure similar to Section 5.1.1, Tables 9 to 11 present our results. As to be expected from the apparent serial correlation and volatility clustering in our time series (see Table 1), introducing filtering almost uniformly improves VaR prediction performance. Across all error levels and all series considered, we can observe much more cases where the VaR estimators are not rejected by our statistical tests.

Again with a focus on the 1% nominal error probability, we start by having a more detailed look at the results for the CRSP Value-Weighted Index in Panel A of Table 9. We can observe that the conditional normal approach performs quite poorly even though some studies (see, for example, Danielsson and Morimoto, 2000) show that it might have some merit (especially for larger values of the nominal error probability).²⁸ However, while it does not perform satisfactorily in terms of unconditional coverage, it appears to be able to produce cluster-free violations. Turning to the alternative approaches, we detect several instances hinting at more efficient VaR forecasts. Historical simulation and all EVT-based approaches (except for the Johnson method) cannot be rejected based on the unconditional coverage, independence and conditional coverage tests. Interestingly, now the peak over threshold method is the most promising EVT-based approach in terms of average test ranks. It is followed by the Johnson method which again is characterised by the lowest average value of the loss function confirming its conservatism. However, it is important to note that the historical simulation also performs quite well with the lowest average testing rank over all approaches (traditional and EVT-based).

For the S&P 500 Index in Panel B of Table 9, we find that the peak over threshold and the L-moment estimators cannot be rejected based on almost all tests but the second dynamic quantile specification. Historical simulation passes all tests and scores first based on the average testing rank, again followed by the peak over threshold method.

As far as the commodity market is concerned, Panel A of Table 10 for the Goldman Sachs Commodity Index reveals that the peak over threshold, L-moment, Box-Cox and historical simulation approaches and even the Student-t method pass all tests. The Box-Cox method provides the highest average testing rank. When focusing on the gold spot investment in Panel B of Table

²⁸Our results indicate that, at least for our dataset, the 5% quantile is still not large enough for the normal assumption to be adequate.

	Test results					Additional statistics		
	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	\overline{Rank}	\overline{Loss}	\overline{VaR}
<i>Panel A: CRSP Value-Weighted Index</i>								
<i>5% level</i>								
Normal	1.95 (2)	23.83 (5)	25.78 (3)	151.94 (2)	175.27 (3)	3.0	0.15	1.74
Student-t	10.91 (7)	17.62 (3)	28.53 (4)	192.08 (5)	213.42 (6)	5.0	0.17	1.65
Historical simulation	4.46 (3)	25.49 (7)	29.95 (6)	170.40 (4)	187.22 (4)	4.8	0.16	1.70
Peak over threshold	6.22 (5)	23.37 (4)	29.59 (5)	194.50 (6)	209.74 (5)	5.0	0.16	1.69
L-moment	8.56 (6)	24.56 (6)	33.12 (7)	199.58 (7)	215.27 (7)	6.6	0.16	1.67
Box-Cox	0.01 (1)	14.22 (2)	14.23 (1)	152.66 (3)	171.24 (2)	1.8	0.15	1.78
Johnson	5.60 (4)	13.90 (1)	19.51 (2)	139.35 (1)	156.92 (1)	1.8	0.14	1.88
<i>2.5% level</i>								
Normal	34.39 (6)	10.88 (1)	45.26 (6)	229.45 (6)	264.46 (6)	5.0	0.11	2.09
Student-t	34.39 (6)	10.88 (1)	45.26 (6)	229.45 (6)	264.47 (7)	5.2	0.11	2.09
Historical simulation	7.77 (4)	16.49 (4)	24.26 (3)	155.14 (4)	167.46 (4)	3.8	0.10	2.22
Peak over threshold	7.36 (3)	19.25 (7)	26.61 (4)	146.67 (3)	162.29 (3)	4.0	0.10	2.21
L-moment	9.05 (5)	18.13 (6)	27.18 (5)	162.32 (5)	176.79 (5)	5.2	0.10	2.19
Box-Cox	1.35 (1)	16.85 (5)	18.20 (2)	121.44 (2)	140.34 (2)	2.4	0.08	2.54
Johnson	1.98 (2)	12.01 (3)	13.99 (1)	93.56 (1)	107.23 (1)	1.6	0.08	2.48
<i>1% level</i>								
Normal	74.94 (6)	9.48 (2)	84.42 (6)	337.95 (6)	376.75 (6)	5.2	0.08	2.49
Student-t	74.94 (6)	9.48 (2)	84.42 (6)	337.95 (6)	376.75 (6)	5.2	0.08	2.49
Historical simulation	10.07 (5)	12.77 (5)	22.83 (5)	122.45 (3)	143.66 (4)	4.4	0.05	2.98
Peak over threshold	7.44 (3)	13.84 (7)	21.28 (3)	127.54 (5)	148.04 (5)	4.6	0.05	2.97
L-moment	9.38 (4)	13.03 (6)	22.41 (4)	123.63 (4)	142.87 (3)	4.2	0.05	2.96
Box-Cox	0.41 (1)	10.81 (4)	11.22 (2)	81.19 (1)	100.04 (2)	2.0	0.03	3.75
Johnson	3.15 (2)	7.01 (1)	10.16 (1)	93.16 (2)	97.55 (1)	1.4	0.03	3.48
<i>Panel B: S&P 500 Index</i>								
<i>5% level</i>								
Normal	0.83 (2)	14.04 (6)	14.87 (4)	134.11 (1)	159.35 (1)	2.8	0.16	1.77
Student-t	5.11 (6)	13.06 (5)	18.17 (6)	171.44 (7)	196.21 (7)	6.2	0.17	1.68
Historical simulation	2.79 (4)	11.18 (3)	13.98 (3)	152.02 (4)	172.20 (4)	3.6	0.17	1.72
Peak over threshold	4.64 (5)	14.92 (7)	19.56 (7)	165.66 (5)	186.15 (5)	5.8	0.17	1.70
L-moment	5.86 (7)	11.13 (2)	16.99 (5)	167.38 (6)	186.28 (6)	5.2	0.17	1.68
Box-Cox	0.13 (1)	9.02 (1)	9.16 (1)	148.12 (3)	169.94 (3)	1.8	0.16	1.78
Johnson	1.17 (3)	12.49 (4)	13.66 (2)	136.02 (2)	159.84 (2)	2.6	0.15	1.87
<i>2.5% level</i>								
Normal	29.89 (7)	7.91 (2)	37.80 (7)	232.41 (7)	269.36 (7)	6.0	0.12	2.11
Student-t	18.38 (6)	7.43 (1)	25.82 (6)	193.53 (6)	226.85 (6)	5.0	0.11	2.15
Historical simulation	3.38 (2)	9.64 (3)	13.01 (2)	135.36 (4)	150.67 (4)	3.0	0.10	2.23
Peak over threshold	3.94 (4)	11.32 (6)	15.26 (3)	132.41 (3)	150.24 (3)	3.8	0.10	2.25
L-moment	4.87 (5)	10.72 (5)	15.59 (4)	139.87 (5)	154.81 (5)	4.8	0.10	2.22
Box-Cox	3.52 (3)	12.48 (7)	16.00 (5)	120.86 (2)	137.10 (2)	3.8	0.08	2.51
Johnson	0.99 (1)	10.15 (4)	11.14 (1)	93.71 (1)	113.92 (1)	1.6	0.09	2.44
<i>1% level</i>								
Normal	75.89 (7)	11.60 (5)	87.49 (7)	326.53 (7)	378.56 (7)	6.6	0.08	2.51
Student-t	26.24 (6)	14.52 (7)	40.75 (6)	250.90 (6)	278.89 (6)	6.2	0.07	2.82
Historical simulation	7.83 (5)	9.75 (2)	17.58 (5)	96.49 (2)	112.24 (2)	3.2	0.05	3.02
Peak over threshold	5.04 (3)	10.87 (4)	15.92 (3)	114.95 (5)	130.29 (5)	4.0	0.05	3.04
L-moment	6.65 (4)	10.19 (3)	16.84 (4)	111.01 (4)	125.07 (4)	3.8	0.05	3.02
Box-Cox	0.00 (1)	4.33 (1)	4.34 (1)	77.88 (1)	96.96 (1)	1.0	0.04	3.67
Johnson	0.49 (2)	13.38 (6)	13.87 (2)	103.70 (3)	113.04 (3)	3.2	0.04	3.52

For our unconditional VaR estimators and three nominal error probabilities, this table reports the test statistics of the unconditional coverage test (L_{uc}), the independence test (L_{ind}), the conditional coverage test (L_{cc}) and the dynamic quantile test (DQ) of VaR forecasts for the CRSP Value-Weighted Index and the S&P 500 Index. DQ is applied in two specifications: one that contains only a constant and three lagged violations (DQ_{Hit}) and one that additionally considers the contemporaneous VaR estimate (DQ_{VaR}). Bold type entries are not significant at the 1% level. Besides these test statistics we report the ranks of each estimator based on each test (from best = 1 to worst = 7). The last three columns present the average rank over all tests (\overline{Rank}), the average value of the loss function (\overline{Loss}) and the (absolute) average of the VaR estimates (\overline{VaR}).

Table 6: VaR prediction performance of unconditional models for CRSP and SP500

	Test results					Additional statistics		
	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	\overline{Rank}	\overline{Loss}	\overline{VaR}
<i>Panel A: Goldman Sachs Commodity Index</i>								
<i>5% level</i>								
Normal	8.71 (5)	22.26 (3)	30.97 (3)	77.99 (3)	88.49 (2)	3.2	0.20	2.07
Student-t	28.52 (7)	20.90 (2)	49.42 (7)	95.86 (7)	115.06 (7)	6.0	0.22	1.96
Historical simulation	6.21 (1)	19.69 (1)	25.81 (1)	63.02 (1)	73.18 (1)	1.0	0.21	2.03
Peak over threshold	10.00 (6)	22.83 (4)	32.83 (5)	80.47 (5)	89.33 (4)	4.8	0.21	2.02
L-moment	8.40 (4)	24.27 (6)	32.67 (4)	78.57 (4)	88.91 (3)	4.2	0.21	2.02
Box-Cox	6.94 (3)	23.97 (5)	30.91 (2)	72.29 (2)	89.89 (5)	3.4	0.20	2.08
Johnson	6.90 (2)	42.04 (7)	48.95 (6)	88.50 (6)	103.75 (6)	5.4	0.17	2.32
<i>2.5% level</i>								
Normal	39.48 (7)	41.31 (1)	80.79 (7)	184.80 (7)	203.45(7)	5.8	0.15	2.47
Student-t	31.40 (6)	42.65 (2)	74.05 (6)	176.13 (6)	194.85 (6)	5.2	0.14	2.51
Historical simulation	11.33 (3)	51.12 (7)	62.44 (3)	169.56 (3)	192.19 (4)	4.0	0.13	2.67
Peak over threshold	14.93 (5)	47.68 (4)	62.61 (5)	171.09 (5)	190.93 (3)	4.4	0.13	2.62
L-moment	14.39 (4)	48.16 (5)	62.55 (4)	170.25 (4)	192.63 (5)	4.4	0.14	2.62
Box-Cox	4.87 (2)	48.42 (6)	53.29 (2)	148.42 (2)	182.47 (2)	2.8	0.12	2.82
Johnson	0.57 (1)	45.06 (3)	45.63 (1)	122.47 (1)	136.74 (1)	1.4	0.11	2.99
<i>1% level</i>								
Normal	94.43 (7)	36.87 (7)	131.31 (7)	384.40 (7)	402.71 (7)	7.0	0.11	2.93
Student-t	19.37 (6)	32.82 (2)	52.19 (6)	205.84 (6)	211.95 (6)	5.2	0.08	3.29
Historical simulation	17.57 (5)	33.75 (3)	51.32 (5)	181.83 (5)	196.93 (5)	4.6	0.08	3.43
Peak over threshold	16.70 (4)	34.22 (4)	50.92 (4)	173.49 (3)	185.22 (3)	3.6	0.08	3.44
L-moment	15.02 (3)	35.18 (5)	50.20 (3)	176.20 (4)	188.71 (4)	3.8	0.08	3.44
Box-Cox	2.83 (2)	20.46 (1)	23.29 (1)	78.50 (1)	97.82 (1)	1.2	0.07	3.93
Johnson	0.16 (1)	35.71 (6)	35.87 (2)	142.46 (2)	146.90 (2)	2.6	0.07	3.93
<i>Panel B: London Bullion Market gold</i>								
<i>5% level</i>								
Normal	0.23 (1)	11.33 (1)	11.55 (1)	27.49 (1)	28.15 (1)	1.0	0.14	1.53
Student-t	4.64 (6)	13.48 (2)	18.12 (5)	37.07 (2)	37.50 (2)	3.4	0.15	1.45
Historical simulation	2.28 (4)	14.62 (4)	16.90 (4)	38.62 (4)	40.16 (4)	4.0	0.14	1.51
Peak over threshold	1.53 (3)	14.23 (3)	15.76 (3)	37.09 (3)	38.49 (3)	3.0	0.14	1.50
L-moment	3.56 (5)	16.06 (6)	19.62 (6)	40.81 (6)	42.18 (6)	5.8	0.14	1.49
Box-Cox	0.33 (2)	15.34 (5)	15.67 (2)	38.78 (5)	40.60 (5)	3.8	0.13	1.58
Johnson	6.31 (7)	22.23 (7)	28.54 (7)	47.71 (7)	53.97 (7)	7.0	0.12	1.67
<i>2.5% level</i>								
Normal	17.78 (7)	13.12 (5)	30.90 (7)	67.28 (6)	68.07 (6)	6.2	0.10	1.83
Student-t	13.86 (6)	14.70 (7)	28.55 (6)	68.30 (7)	69.18 (7)	6.6	0.09	1.86
Historical simulation	0.88 (2)	9.82 (2)	10.70 (1)	34.77 (2)	39.95 (2)	1.8	0.08	2.00
Peak over threshold	0.75 (1)	10.01 (3)	10.76 (2)	39.35 (3)	44.58 (3)	2.4	0.08	2.00
L-moment	1.73 (3)	13.25 (6)	14.99 (5)	40.86 (5)	44.60 (4)	4.6	0.08	1.98
Box-Cox	2.67 (5)	11.81 (4)	14.48 (4)	40.31 (4)	56.59 (5)	4.4	0.07	2.27
Johnson	2.41 (4)	9.12 (1)	11.53 (3)	30.89 (1)	38.66 (1)	2.0	0.07	2.22
<i>1% level</i>								
Normal	77.51 (7)	11.38 (6)	88.89 (7)	183.65 (7)	193.69 (7)	6.8	0.07	2.17
Student-t	17.57 (6)	10.19 (5)	27.76 (6)	59.69 (6)	61.60 (6)	5.8	0.05	2.44
Historical simulation	1.22 (1)	9.53 (4)	10.75 (3)	29.34 (3)	34.71 (2)	2.6	0.04	2.66
Peak over threshold	2.83 (4)	8.47 (3)	11.30 (4)	33.91 (4)	41.10 (3)	3.6	0.04	2.70
L-moment	2.11 (2)	5.60 (2)	7.71 (2)	25.67 (1)	31.78 (1)	1.6	0.04	2.68
Box-Cox	2.52 (3)	0.49 (1)	3.02 (1)	26.23 (2)	44.43 (4)	2.2	0.03	3.33
Johnson	8.56 (5)	13.74 (7)	22.30 (5)	45.44 (5)	46.04 (5)	5.4	0.03	3.15

For our unconditional VaR estimators and three nominal error probabilities, this table reports the test statistics of the unconditional coverage test (L_{uc}), the independence test (L_{ind}), the conditional coverage test (L_{cc}) and the dynamic quantile test (DQ) of VaR forecasts for the Goldman Sachs Commodity Index and London Bullion Market gold. DQ is applied in two specifications: one that contains only a constant and three lagged violations (DQ_{Hit}) and one that additionally considers the contemporaneous VaR estimate (DQ_{VaR}). Bold type entries are not significant at the 1% level. Besides these test statistics we report the ranks of each estimator based on each test (from best = 1 to worst = 7). The last three columns present the average rank over all tests (\overline{Rank}), the average value of the loss function (\overline{Loss}) and the (absolute) average of the VaR estimates (\overline{VaR}).

Table 7: VaR prediction performance of unconditional models for GSCI and GOLD

	Test results					Additional statistics		
	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	\overline{Rank}	\overline{Loss}	\overline{VaR}
<i>Panel A: Datastream US Government Bond Index</i>								
<i>5% level</i>								
Normal	0.83 (5)	12.57 (7)	13.40 (5)	24.36 (4)	51.87 (2)	4.6	0.06	0.74
Student-t	10.68 (6)	9.44 (4)	20.12 (6)	36.95 (7)	73.90 (7)	6.0	0.07	0.70
Historical simulation	0.10 (2)	7.95 (2)	8.05 (1)	18.64 (1)	45.00 (1)	1.4	0.06	0.75
Peak over threshold	0.56 (3)	9.09 (3)	9.65 (2)	21.45 (2)	54.33 (3)	2.6	0.06	0.75
L-moment	0.74 (4)	10.02 (5)	10.76 (3)	25.20 (5)	62.93 (5)	4.4	0.06	0.74
Box-Cox	0.04 (1)	10.89 (6)	10.93 (4)	23.91 (3)	57.35 (4)	3.6	0.06	0.75
Johnson	29.41 (7)	3.49 (1)	32.90 (7)	36.40 (6)	64.13 (6)	5.4	0.04	0.85
<i>2.5% level</i>								
Normal	16.04 (7)	3.68 (7)	19.73 (7)	31.01 (7)	60.95 (7)	7.0	0.04	0.89
Student-t	11.33 (6)	1.36 (2)	12.68 (6)	19.28 (6)	47.38 (6)	5.2	0.04	0.90
Historical simulation	2.37 (4)	1.20 (1)	3.57 (3)	8.25 (2)	36.60 (4)	2.8	0.03	0.95
Peak over threshold	1.93 (3)	2.35 (5)	4.28 (4)	8.25 (2)	27.53 (1)	3.0	0.03	0.95
L-moment	1.36 (2)	1.51 (3)	2.87 (2)	8.45 (4)	37.26 (5)	3.2	0.03	0.95
Box-Cox	0.23 (1)	2.17 (4)	2.40 (1)	5.47 (1)	30.06 (2)	1.8	0.03	0.98
Johnson	7.26 (5)	2.96 (6)	10.23 (5)	14.28 (5)	35.41 (3)	4.8	0.02	1.09
<i>1% level</i>								
Normal	71.11 (7)	0.87 (6)	71.98 (7)	101.32 (7)	126.92 (7)	6.8	0.03	1.05
Student-t	9.10 (5)	0.32 (3)	9.42 (5)	11.71 (5)	28.74 (3)	4.2	0.02	1.18
Historical simulation	7.83 (4)	0.38 (4)	8.21 (4)	10.02 (4)	53.46 (5)	4.2	0.02	1.24
Peak over threshold	67.99 (6)	0.97 (7)	68.96 (6)	97.37 (6)	123.39 (6)	6.2	0.02	1.22
L-moment	5.55 (3)	0.02 (1)	5.57 (3)	6.18 (3)	40.14 (4)	2.8	0.02	1.22
Box-Cox	0.27 (1)	0.06 (2)	0.33 (1)	0.48 (1)	18.63 (2)	1.4	0.01	1.30
Johnson	3.92 (2)	0.62 (5)	4.54 (2)	5.30 (2)	7.21 (1)	2.4	0.01	1.40
<i>Panel B: US-Dollar to Great Britain Pound exchange rate</i>								
<i>5% level</i>								
Normal	0.17 (1)	34.22 (3)	34.39 (3)	64.40 (5)	78.95 (5)	3.4	0.07	0.95
Student-t	2.28 (6)	28.56 (1)	30.84 (1)	66.44 (6)	85.77 (6)	4.0	0.07	0.90
Historical simulation	0.48 (2)	33.75 (2)	34.23 (2)	55.62 (1)	67.18 (1)	1.6	0.06	0.96
Peak over threshold	0.64 (3)	36.83 (5)	37.48 (5)	62.58 (3)	73.28 (3)	3.8	0.06	0.96
L-moment	1.30 (5)	34.35 (4)	35.65 (4)	58.39 (2)	71.18 (2)	3.4	0.06	0.97
Box-Cox	0.84 (4)	37.60 (6)	38.44 (6)	64.19 (4)	77.45 (4)	4.8	0.06	0.97
Johnson	21.48 (7)	42.51 (7)	63.99 (7)	82.51 (7)	94.08 (7)	7.0	0.05	1.07
<i>2.5% level</i>								
Normal	11.33 (7)	31.67 (1)	43.00 (6)	88.29 (5)	101.88 (7)	5.2	0.04	1.14
Student-t	8.19 (5)	34.35 (4)	42.54 (5)	89.43 (6)	101.12 (6)	5.2	0.04	1.15
Historical simulation	0.06 (1)	31.90 (2)	31.96 (1)	70.08 (1)	85.98 (1)	1.2	0.03	1.25
Peak over threshold	0.62 (4)	35.91 (6)	36.53 (4)	78.49 (4)	87.75 (3)	4.2	0.04	1.22
L-moment	0.41 (3)	33.31 (3)	33.72 (2)	72.82 (2)	86.98 (2)	2.4	0.04	1.24
Box-Cox	0.06 (1)	35.36 (5)	35.42 (3)	77.33 (3)	90.88 (4)	3.2	0.03	1.27
Johnson	11.32 (6)	40.11 (7)	51.44 (7)	96.94 (7)	101.00 (5)	6.4	0.03	1.37
<i>1% level</i>								
Normal	41.26 (7)	32.57 (7)	73.83 (7)	242.44 (7)	247.89 (7)	7.0	0.03	1.34
Student-t	1.49 (5)	32.00 (6)	33.49 (6)	172.26 (6)	173.15 (6)	5.8	0.02	1.51
Historical simulation	1.22 (4)	17.71 (3)	18.94 (2)	136.91 (5)	140.83 (5)	3.8	0.02	1.59
Peak over threshold	0.76 (3)	18.39 (4)	19.15 (3)	121.93 (3)	123.53 (3)	3.2	0.02	1.56
L-moment	0.41 (2)	19.09 (5)	19.50 (4)	126.17 (4)	127.97 (4)	3.8	0.01	1.58
Box-Cox	0.08 (1)	15.49 (2)	15.57 (1)	119.11 (2)	123.45 (2)	1.6	0.01	1.68
Johnson	16.52 (6)	10.79 (1)	27.31 (5)	58.81 (1)	59.21 (1)	2.8	0.01	1.82

For our unconditional VaR estimators and three nominal error probabilities, this table reports the test statistics of the unconditional coverage test (L_{uc}), the independence test (L_{ind}), the conditional coverage test (L_{cc}) and the dynamic quantile test (DQ) of VaR forecasts for the Datastream US Government Bond Index and the US-Dollar to Great Britain Pound exchange rate. DQ is applied in two specifications: one that contains only a constant and three lagged violations (DQ_{Hit}) and one that additionally considers the contemporaneous VaR estimate (DQ_{VaR}). Bold type entries are not significant at the 1% level. Besides these test statistics we report the ranks of each estimator based on each test (from best = 1 to worst = 7). The last three columns present the average rank over all tests (\overline{Rank}), the average value of the loss function (\overline{Loss}) and the (absolute) average of the VaR estimates (\overline{VaR}).

Table 8: VaR prediction performance of unconditional models for USGBI and \$/£

	Test results					Additional statistics		
	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	\overline{Rank}	\overline{Loss}	\overline{VaR}
<i>Panel A: CRSP Value-Weighted Index</i>								
<i>5% level</i>								
Normal	9.87 (5)	0.36 (6)	10.23 (5)	24.50 (5)	26.33 (5)	5.2	0.11	1.57
Student-t	35.81 (6)	0.01 (3)	35.82 (6)	53.93 (7)	54.97 (7)	5.8	0.12	1.49
Historical simulation	1.13 (2)	0.01 (3)	1.14 (2)	17.90 (3)	23.07 (3)	2.6	0.09	1.67
Peak over threshold	2.27 (4)	0.00 (1)	2.28 (4)	17.62 (2)	21.71 (2)	2.6	0.09	1.65
L-moment	1.52 (3)	0.00 (1)	1.52 (3)	19.75 (4)	24.75 (4)	3.0	0.09	1.66
Box-Cox	0.26 (1)	0.14 (5)	0.40 (1)	13.04 (1)	15.94 (1)	1.8	0.09	1.71
Johnson	36.40 (7)	0.55 (7)	36.95 (7)	47.85 (6)	49.13 (6)	6.6	0.06	1.92
<i>2.5% level</i>								
Normal	31.99 (7)	0.16 (1)	32.15 (7)	59.49 (7)	61.74 (7)	5.8	0.07	1.89
Student-t	22.49 (5)	0.55 (2)	23.04 (5)	45.75 (6)	47.68 (6)	4.8	0.06	1.92
Historical simulation	0.19 (1)	2.54 (7)	2.72 (2)	11.63 (1)	16.90 (1)	2.4	0.04	2.12
Peak over threshold	2.52 (4)	1.38 (3)	3.90 (4)	15.27 (3)	18.63 (2)	3.2	0.05	2.08
L-moment	2.06 (3)	1.51 (5)	3.57 (3)	15.34 (4)	19.12 (3)	3.6	0.05	2.10
Box-Cox	0.70 (2)	1.40 (4)	2.10 (1)	13.44 (2)	20.18 (4)	2.6	0.04	2.26
Johnson	26.20 (6)	2.51 (6)	28.71 (6)	40.93 (5)	43.22 (5)	5.6	0.03	2.44
<i>1% level</i>								
Normal	59.32 (7)	1.39 (6)	60.71 (7)	110.32 (7)	119.42 (7)	6.8	0.04	2.26
Student-t	8.71 (6)	1.80 (7)	10.51 (6)	38.34 (6)	44.20 (6)	6.2	0.02	2.54
Historical simulation	0.99 (3)	0.03 (1)	1.02 (2)	15.92 (2)	21.10 (2)	2.0	0.02	2.62
Peak over threshold	0.27 (2)	0.08 (2)	0.35 (1)	16.68 (3)	22.40 (3)	2.2	0.02	2.67
L-moment	0.00 (1)	1.33 (5)	1.34 (3)	18.39 (4)	23.50 (4)	3.4	0.02	2.67
Box-Cox	6.03 (4)	0.67 (4)	6.70 (4)	21.19 (5)	25.81 (5)	4.4	0.01	3.05
Johnson	8.26 (5)	0.59 (3)	8.85 (5)	14.36 (1)	17.07 (1)	3.0	0.01	3.03
<i>Panel B: S&P 500 Index</i>								
<i>5% level</i>								
Normal	4.64 (5)	0.62 (6)	5.26 (5)	18.18 (4)	20.02 (4)	4.8	0.11	1.58
Student-t	30.17 (7)	0.03 (1)	30.20 (7)	49.36 (7)	50.41 (7)	5.8	0.13	1.50
Historical simulation	0.28 (3)	0.29 (5)	0.57 (3)	16.40 (2)	18.46 (3)	3.2	0.10	1.66
Peak over threshold	0.22 (2)	0.07 (3)	0.29 (2)	14.63 (1)	16.18 (1)	1.8	0.10	1.66
L-moment	0.17 (1)	0.06 (2)	0.23 (1)	16.57 (3)	18.16 (2)	1.8	0.10	1.66
Box-Cox	2.02 (4)	0.12 (4)	2.14 (4)	18.91 (5)	21.52 (5)	4.4	0.09	1.72
Johnson	25.58 (6)	0.63 (7)	26.22 (6)	35.68 (6)	36.81 (6)	6.2	0.07	1.88
<i>2.5% level</i>								
Normal	34.54 (7)	0.77 (1)	35.30 (7)	58.55 (7)	60.79 (7)	5.8	0.07	1.90
Student-t	26.96 (6)	1.30 (2)	28.26 (6)	49.21 (6)	50.62 (6)	5.2	0.07	1.93
Historical simulation	0.11 (1)	5.30 (7)	5.41 (1)	18.90 (3)	23.32 (3)	3.0	0.05	2.14
Peak over threshold	1.93 (2)	5.12 (6)	7.06 (3)	17.73 (2)	19.90 (1)	2.8	0.05	2.11
L-moment	2.61 (3)	3.31 (4)	5.92 (2)	19.51 (4)	21.16 (2)	3.0	0.05	2.10
Box-Cox	3.52 (4)	3.59 (5)	7.10 (4)	17.32 (1)	25.31 (4)	3.6	0.04	2.32
Johnson	15.70 (5)	2.73 (3)	18.43 (5)	25.08 (5)	30.81 (5)	4.6	0.03	2.41
<i>1% level</i>								
Normal	69.54 (7)	3.59 (6)	73.13 (7)	127.06 (7)	139.11 (7)	6.8	0.04	2.26
Student-t	11.17 (5)	0.24 (2)	11.41 (5)	37.37 (6)	45.81 (6)	4.8	0.03	2.54
Historical simulation	0.98 (3)	0.02 (1)	1.00 (1)	8.11 (1)	13.44 (1)	1.4	0.02	2.67
Peak over threshold	0.41 (2)	1.31 (4)	1.72 (3)	10.03 (2)	16.52 (3)	2.8	0.02	2.72
L-moment	0.16 (1)	1.45 (5)	1.61 (2)	10.35 (3)	16.37 (2)	2.6	0.02	2.71
Box-Cox	5.67 (4)	3.65 (7)	9.32 (4)	16.11 (4)	22.47 (5)	4.8	0.02	3.18
Johnson	12.15 (6)	0.51 (3)	12.67 (6)	18.08 (5)	18.09 (4)	4.8	0.01	4.01

For our conditional VaR estimators and three nominal error probabilities, this table reports the test statistics of the unconditional coverage test (L_{uc}), the independence test (L_{ind}), the conditional coverage test (L_{cc}) and the dynamic quantile test (DQ) of VaR forecasts for the CRSP Value-Weighted Index and the S&P 500 Index. DQ is applied in two specifications: one that contains only a constant and three lagged violations (DQ_{Hit}) and one that additionally considers the contemporaneous VaR estimate (DQ_{VaR}). Bold type entries are not significant at the 1% level. Besides these test statistics we report the ranks of each estimator based on each test (from best = 1 to worst = 7). The last three columns present the average rank over all tests (\overline{Rank}), the average value of the loss function (\overline{Loss}) and the (absolute) average of the VaR estimates (\overline{VaR}).

Table 9: VaR prediction performance of conditional models for CRSP and SP500

	Test results					Additional statistics		
	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	\overline{Rank}	\overline{Loss}	\overline{VaR}
<i>Panel A: Goldman Sachs Commodity Index</i>								
<i>5% level</i>								
Normal	5.86 (5)	2.36 (7)	8.22 (5)	12.23 (5)	12.52 (5)	5.4	0.15	1.97
Student-t	27.43 (6)	2.13 (3)	29.57 (6)	35.50 (7)	35.52 (7)	5.8	0.17	1.87
Historical simulation	2.98 (3)	2.13 (3)	5.11 (3)	9.60 (4)	9.68 (4)	3.4	0.14	2.02
Peak over threshold	3.36 (4)	1.98 (2)	5.34 (4)	9.25 (2)	9.26 (2)	2.8	0.15	2.01
L-moment	2.79 (2)	2.21 (5)	5.01 (2)	9.37 (3)	9.39 (3)	3.0	0.14	2.00
Box-Cox	1.53 (1)	2.21 (5)	3.74 (1)	7.91 (1)	8.14 (1)	1.8	0.14	2.02
Johnson	34.99 (7)	1.36 (1)	36.35 (7)	32.95 (6)	33.15 (6)	5.4	0.10	2.31
<i>2.5% level</i>								
Normal	15.48 (6)	0.91 (2)	16.39 (6)	18.71 (6)	19.39 (6)	5.2	0.09	2.36
Student-t	8.61 (5)	1.76 (6)	10.37 (5)	12.24 (5)	13.45 (5)	5.2	0.09	2.40
Historical simulation	0.88 (2)	1.71 (5)	2.59 (1)	3.30 (1)	4.75 (1)	2.0	0.08	2.51
Peak over threshold	1.03 (3)	1.64 (4)	2.67 (2)	3.39 (2)	5.30 (3)	2.8	0.08	2.50
L-moment	1.19 (4)	1.57 (3)	2.76 (3)	3.49 (3)	5.67 (4)	3.4	0.08	2.50
Box-Cox	0.08 (1)	3.06 (7)	3.14 (4)	4.51 (4)	5.15 (2)	3.6	0.08	2.57
Johnson	21.69 (7)	0.53 (1)	22.22 (7)	19.99 (7)	20.37 (7)	5.8	0.06	2.86
<i>1% level</i>								
Normal	41.26 (7)	0.15 (4)	41.41 (7)	54.50 (7)	55.21 (7)	6.4	0.06	2.81
Student-t	0.57 (2)	1.25 (7)	1.82 (3)	8.87 (6)	9.05 (6)	4.8	0.04	3.15
Historical simulation	3.22 (5)	0.74 (6)	3.96 (5)	4.62 (2)	4.65 (2)	4.0	0.05	3.10
Peak over threshold	2.11 (4)	0.00 (1)	2.11 (4)	7.52 (4)	7.67 (4)	3.4	0.04	3.13
L-moment	1.79 (3)	0.01 (2)	1.80 (2)	7.37 (3)	7.43 (3)	2.6	0.04	3.12
Box-Cox	0.01 (1)	0.14 (3)	0.15 (1)	2.95 (1)	2.98 (1)	1.4	0.04	3.31
Johnson	7.03 (6)	0.67 (5)	7.70 (6)	7.67 (5)	8.18 (5)	5.4	0.03	3.49
<i>Panel B: London Bullion Market gold</i>								
<i>5% level</i>								
Normal	0.17 (4)	1.90 (3)	2.07 (3)	2.65 (3)	2.69 (2)	3.0	0.11	1.50
Student-t	2.96 (6)	1.05 (1)	4.03 (5)	5.55 (6)	5.57 (6)	4.8	0.13	1.42
Historical simulation	0.02 (2)	2.93 (7)	2.94 (4)	3.88 (4)	3.97 (4)	4.2	0.12	1.48
Peak over threshold	0.00 (1)	2.03 (5)	2.03 (2)	2.62 (2)	2.77 (3)	2.6	0.12	1.48
L-moment	0.04 (3)	1.18 (2)	1.22 (1)	1.49 (1)	2.18 (1)	1.6	0.12	1.48
Box-Cox	1.17 (5)	2.89 (6)	4.06 (6)	4.45 (5)	4.71 (5)	5.4	0.11	1.52
Johnson	25.58 (7)	1.98 (4)	27.56 (7)	25.29 (7)	27.27 (7)	6.4	0.09	1.68
<i>2.5% level</i>								
Normal	10.85 (6)	1.42 (2)	12.27 (6)	14.73 (6)	14.99 (5)	5.0	0.08	1.79
Student-t	7.38 (5)	1.09 (1)	8.47 (4)	10.05 (4)	10.27 (4)	3.6	0.08	1.82
Historical simulation	0.03 (2)	2.51 (3)	2.54 (1)	5.21 (2)	7.63 (1)	1.8	0.07	1.94
Peak over threshold	0.03 (2)	2.51 (3)	2.54 (1)	6.17 (3)	8.52 (3)	2.4	0.07	1.93
L-moment	0.01 (1)	2.59 (5)	2.60 (3)	3.95 (1)	7.93 (2)	2.4	0.07	1.93
Box-Cox	3.83 (4)	5.56 (6)	9.39 (5)	13.28 (5)	15.37 (6)	5.2	0.06	2.07
Johnson	14.37 (7)	6.66 (7)	21.02 (7)	20.54 (7)	25.00 (7)	7.0	0.05	2.17
<i>1% level</i>								
Normal	45.12 (7)	8.10 (6)	53.22 (7)	81.56 (7)	87.75 (7)	6.8	0.05	2.12
Student-t	5.04 (5)	10.87 (7)	15.91 (6)	34.70 (6)	36.47 (6)	6.0	0.04	2.38
Historical simulation	0.04 (3)	1.84 (2)	1.88 (1)	3.63 (1)	5.92 (1)	1.6	0.03	2.64
Peak over threshold	0.00 (1)	7.73 (4)	7.74 (2)	16.96 (4)	18.43 (4)	3.0	0.03	2.56
L-moment	0.01 (2)	7.94 (5)	7.95 (3)	17.38 (5)	19.57 (5)	4.0	0.03	2.57
Box-Cox	2.13 (4)	6.26 (3)	8.39 (4)	14.19 (3)	16.58 (3)	3.4	0.03	2.88
Johnson	12.15 (6)	1.26 (1)	13.41 (5)	12.11 (2)	12.78 (2)	3.2	0.02	3.02

For our conditional VaR estimators and three nominal error probabilities, this table reports the test statistics of the unconditional coverage test (L_{uc}), the independence test (L_{ind}), the conditional coverage test (L_{cc}) and the dynamic quantile test (DQ) of VaR forecasts for the Goldman Sachs Commodity Index and London Bullion Market gold. DQ is applied in two specifications: one that contains only a constant and three lagged violations (DQ_{Hit}) and one that additionally considers the contemporaneous VaR estimate (DQ_{VaR}). Bold type entries are not significant at the 1% level. Besides these test statistics we report the ranks of each estimator based on each test (from best = 1 to worst = 7). The last three columns present the average rank over all tests (\overline{Rank}), the average value of the loss function (\overline{Loss}) and the (absolute) average of the VaR estimates (\overline{VaR}).

Table 10: VaR prediction performance of conditional models for GSCI and GOLD

	Test results					Additional statistics		
	LR_{uc}	LR_{ind}	LR_{cc}	DQ_{Hit}	DQ_{VaR}	\overline{Rank}	\overline{Loss}	\overline{VaR}
<i>Panel A: Datastream US Government Bond Index</i>								
<i>5% level</i>								
Normal	0.01 (1)	5.30 (5)	5.31 (4)	6.51 (4)	9.89 (5)	3.8	0.06	0.73
Student-t	9.03 (6)	5.89 (7)	14.92 (6)	18.72 (6)	19.05 (6)	6.2	0.07	0.69
Historical simulation	0.09 (4)	2.37 (2)	2.46 (1)	3.09 (1)	4.77 (1)	1.8	0.06	0.73
Peak over threshold	0.04 (3)	3.55 (3)	3.58 (2)	4.69 (2)	6.08 (2)	2.4	0.06	0.73
L-moment	0.02 (2)	5.41 (6)	5.43 (5)	7.49 (5)	8.24 (4)	4.4	0.06	0.73
Box-Cox	0.22 (5)	4.29 (4)	4.51 (3)	6.47 (3)	7.74 (3)	3.6	0.06	0.74
Johnson	43.55 (7)	2.36 (1)	45.91 (7)	40.13 (7)	41.19 (7)	5.8	0.04	0.84
<i>2.5% level</i>								
Normal	7.78 (6)	2.33 (7)	10.11 (6)	11.06 (6)	14.18 (6)	6.2	0.04	0.87
Student-t	3.11 (5)	1.47 (6)	4.58 (5)	5.27 (5)	8.16 (5)	5.2	0.03	0.88
Historical simulation	0.04 (1)	1.45 (5)	1.49 (2)	1.56 (2)	5.08 (3)	2.6	0.03	0.91
Peak over threshold	0.57 (3)	1.10 (3)	1.67 (3)	2.30 (3)	4.45 (2)	2.8	0.03	0.92
L-moment	0.13 (2)	1.34 (4)	1.47 (1)	1.50 (1)	3.75 (1)	1.8	0.03	0.91
Box-Cox	2.94 (4)	0.60 (2)	3.54 (4)	4.85 (4)	7.07 (4)	3.6	0.03	0.94
Johnson	32.83 (7)	0.23 (1)	33.06 (7)	29.31 (7)	30.02 (7)	5.8	0.02	1.04
<i>1% level</i>								
Normal	19.37 (7)	0.35 (2)	19.71 (7)	26.08 (7)	28.00 (7)	6.0	0.02	1.03
Student-t	0.57 (2)	1.65 (5)	2.22 (3)	2.48 (1)	3.33 (1)	2.4	0.01	1.15
Historical simulation	0.57 (2)	0.04 (1)	0.61 (1)	3.46 (2)	7.08 (5)	2.2	0.01	1.18
Peak over threshold	0.76 (4)	1.70 (6)	2.46 (4)	4.47 (4)	5.79 (2)	4.0	0.01	1.16
L-moment	1.22 (5)	1.79 (7)	3.01 (5)	4.93 (5)	6.52 (4)	5.2	0.01	1.16
Box-Cox	0.00 (1)	1.40 (4)	1.40 (2)	3.91 (3)	6.18 (3)	2.6	0.01	1.21
Johnson	9.39 (6)	0.59 (3)	9.98 (6)	9.74 (6)	10.78 (6)	5.4	0.01	1.31
<i>Panel B: US-Dollar to Great Britain Pound exchange rate</i>								
<i>5% level</i>								
Normal	1.81 (5)	3.48 (6)	5.29 (5)	7.77 (5)	8.36 (5)	5.2	0.06	0.89
Student-t	16.00 (6)	2.99 (5)	18.99 (6)	23.23 (6)	23.66 (6)	5.8	0.07	0.85
Historical simulation	0.00 (1)	1.42 (2)	1.42 (2)	2.16 (1)	3.84 (1)	1.4	0.06	0.92
Peak over threshold	0.01 (3)	1.29 (1)	1.30 (1)	3.22 (2)	5.67 (2)	1.8	0.06	0.92
L-moment	0.01 (3)	2.11 (4)	2.12 (4)	4.01 (3)	6.80 (3)	3.4	0.06	0.92
Box-Cox	0.00 (1)	1.96 (3)	1.96 (3)	4.01 (3)	6.92 (4)	2.8	0.06	0.92
Johnson	32.83 (7)	5.29 (7)	38.12 (7)	34.81 (7)	35.73 (7)	7.0	0.04	1.03
<i>2.5% level</i>								
Normal	11.33 (6)	6.20 (7)	17.53 (6)	23.75 (6)	24.01 (6)	6.2	0.04	1.07
Student-t	6.99 (5)	6.03 (6)	13.02 (5)	17.47 (5)	17.61 (5)	5.2	0.04	1.08
Historical simulation	0.00 (1)	0.63 (1)	0.63 (1)	0.85 (1)	4.20 (1)	1.0	0.03	1.16
Peak over threshold	0.51 (4)	1.93 (5)	2.44 (4)	3.48 (4)	6.22 (4)	4.2	0.03	1.14
L-moment	0.06 (3)	1.31 (3)	1.37 (2)	1.75 (2)	4.55 (2)	2.4	0.03	1.15
Box-Cox	0.01 (2)	1.43 (4)	1.44 (3)	1.85 (3)	4.61 (3)	3.0	0.03	1.15
Johnson	28.81 (7)	0.89 (2)	29.71 (7)	26.54 (7)	28.53 (7)	6.0	0.02	1.29
<i>1% level</i>								
Normal	24.18 (6)	3.79 (7)	27.97 (6)	40.01 (7)	40.10 (7)	6.6	0.02	1.27
Student-t	0.04 (1)	1.32 (5)	1.36 (1)	3.65 (4)	3.67 (3)	2.8	0.01	1.42
Historical simulation	0.33 (5)	1.20 (2)	1.53 (4)	4.31 (5)	5.05 (5)	4.2	0.01	1.44
Peak over threshold	0.27 (4)	1.57 (6)	1.84 (5)	3.30 (3)	4.37 (4)	4.4	0.01	1.41
L-moment	0.11 (2)	1.28 (3)	1.39 (2)	1.14 (1)	2.25 (1)	1.8	0.01	1.43
Box-Cox	0.11 (2)	1.28 (3)	1.39 (2)	1.14 (1)	2.46 (2)	2.0	0.01	1.44
Johnson	31.36 (7)	0.23 (1)	31.59 (7)	30.09 (6)	30.09 (6)	5.4	0.00	1.66

For our conditional VaR estimators and three nominal error probabilities, this table reports the test statistics of the unconditional coverage test (L_{uc}), the independence test (L_{ind}), the conditional coverage test (L_{cc}) and the dynamic quantile test (DQ) of VaR forecasts for the Datastream US Government Bond Index and the US-Dollar to Great Britain Pound exchange rate. DQ is applied in two specifications: one that contains only a constant and three lagged violations (DQ_{Hit}) and one that additionally considers the contemporaneous VaR estimate (DQ_{VaR}). Bold type entries are not significant at the 1% level. Besides these test statistics we report the ranks of each estimator based on each test (from best = 1 to worst = 7). The last three columns present the average rank over all tests (\overline{Rank}), the average value of the loss function (\overline{Loss}) and the (absolute) average of the VaR estimates (\overline{VaR}).

Table 11: VaR prediction performance of conditional models for USGBI and \$/£

10 this picture is significantly altered. Here, only historical simulation passes all tests and reaches the highest rank.

Interestingly, for the bond market (Panel A of Table 11) and the currency market (Panel B of Table 11) our findings are similar to the ones for the commodity index. That is, all methods except for the normal and the Johnson approach pass all tests. As for the commodity index, the L-moment and the Box-Cox method are the highest ranked approaches for the exchange rate, whereas historical simulation scores first and is followed by the Student-t approach in the case of the bond index.

5.1.3 Compact testing rank summary

In order to provide a compact summary of our findings allowing the identification of the most suitable approaches in both the unconditional and the conditional setup, Table 12 summarises our testing ranks over all time series. While Panel A covers all nominal error probabilities (5%, 2.5% and 1%) and is thus more relevant for non-financial institutions, Panel B is geared towards the 1% level required by the Basel framework and thus is of particular interest for financial institutions (see Gilli and K ellezi, 2006).

As we can see, with a focus on all error probabilities, the Box-Cox method is the most promising unconditional approach directly followed by historical simulation. However, in the conditional setting, historical simulation takes the lead before the peak over threshold method indicating that return filtering has a stronger positive effect on historical simulation than on the EVT-based approaches.²⁹ Thus, considering the generally better test results for the conditional approaches and putting ourselves in the perspective of a risk manager who is interested in ‘higher-order’ quantiles, our results suggest that, at least for our data, EVT-based approaches are not necessarily superior to simple historical simulation.

When looking only at the 1% nominal error probability, we find that again the Box-Cox method shows the best unconditional performance. However, this time the Johnson method ranks second instead of historical simulation. For the conditional versions, we find the most promising results for the historical simulation, followed by the Box-Cox method.

	ND	ST	HS	POT	LM	BC	JO
<i>Panel A: All error probabilities</i>							
Unconditional	5.1	5.3	3.1	4.0	4.1	2.6	3.6
Conditional	5.6	5.0	2.5	2.9	3.0	3.3	5.5
<i>Panel B: 1% error probability</i>							
Unconditional	6.6	5.4	3.8	4.2	3.3	1.6	3.0
Conditional	6.5	4.5	2.6	3.3	3.3	3.1	4.5

For each VaR estimation method, Panel A of this table averages the testing ranks reported in Tables 6 to 11 over all analysed time series and nominal error probabilities. Similarly, Panel B provides a summary over all time series but for the 1% level only. Methodology abbreviations are used as in Table 3.

Table 12: Overall testing ranks

²⁹A detailed investigation of this issue is beyond the scope of this article but might be an interesting topic for future research.

5.2 Robustness checks

5.2.1 General specification tests

To ensure that our main results are not driven by some specific characteristics of our research setup, we have performed a variety of sensitivity checks. For the sake of brevity, this section concentrates on describing their design and main outcomes.³⁰

Alternative GARCH filters. So far, we have implemented the GARCH filters for our conditional approaches based on typical standard settings commonly used in practice and academia. That is, we have (i) used only a very parsimonious number of parameters, (ii) focused on normally distributed errors and (iii) specified a standard GARCH equation of the Bollerslev (1986) type. However, as highlighted by Escanciano and Olmo (2010), model misspecification can seriously bias VaR estimators and related procedures. Therefore, to shed some light on the impact of different (potentially more adequate) filter settings on our results, we perform three robustness checks. First, we follow Bali (2007) and Cheng and Hung (2011) by also considering Hansen (1994) skewed-t distributed errors. Second, we implement the set of alternative GARCH specifications used by Bali et al. (2008), i.e., AGARCH, EGARCH, GJR-GARCH, IGARCH, NGARCH, QGARCH, TGARCH and TS-GARCH. Finally, instead of using the simplest specification, we determine the GARCH orders by means of the Akaike information criterion, as done in, for example, Silvennoinen and Teräsvirta (2009). Interestingly, our results on the relative performance of traditional and EVT-based VaR estimators are not significantly altered, indicating that the choice of GARCH filter does not crucially matter. This is partially because, in our data, (i) the Akaike criterion identifies low GARCH orders and (ii) skewness is not critical such that asymmetric GARCH models do not add significant value over our standard model.

Alternative nonparametric filters. We also analyse how modifications of our nonparametric filter affect our main conclusions. Besides using the method of Fan (1992) we also implement the volatility estimator of Carroll et al. (2002) in our filter procedure for the L-moment method. Using their least squares method with a rule-of-thumb bandwidth selection for the bivariate marginal integration estimation (as in Linton and Nielsen, 1995), we find that, in line with the simulation evidence of Martins-Filho and Yao (2006), our results on the performance of the conditional L-moment method do not change.

Direct extreme modeling. Motivated by the results of Bali and Neftci (2003) and Bali and Weinbaum (2007) showing significant time-series variation in the conditional mean and standard deviation of the excesses over high thresholds, Bali (2007) presents an exercise which directly models serial correlation and volatility clustering in the exceedances over high thresholds. That is, he specifies a different kind of conditional approach in which the parameters of the Box-Cox distribution are time-varying.³¹ He parameterises the location parameter γ_t to be a function of the last period's extrema and the scale parameter δ_t as a function of the last period's unexpected information shock (absolute deviation of the extrema from the predicted extrema) and the last period's scale parameter. Furthermore, the shape parameter ξ_t is modelled as a function of the last period's information shock and the last period's shape parameter.³² These functions are estimated based on a rolling time interval and the resulting time-varying Box-Cox parameter values are plugged into the formula (13) to obtain a conditional VaR estimate. We implement this alternative approach and find that, for our time series, its performance is similar to applying the Box-Cox approach to GARCH filtered data.

³⁰Detailed results are available upon request.

³¹This approach is comparable to Bali et al. (2008) where the parameters of the skewed generalised t distribution (which nests the Hansen (1994) skewed-t distribution) are modelled to be time-varying.

³²Thus, not only the means and the dispersion of extremes are time-varying but also the tail index which measures the fatness of the distribution (or the weight of the tails).

Alternative window sizes. Motivated by previous research, our main analysis uses a moving window of 1,000 observations to generate VaR forecasts. However, as it can be decisive to know the extent to which the results are compatible when using different window lengths, we (i) also follow Bali et al. (2008) by using a 10-year sample to obtain more extremes and (ii) repeat all calculations using 500 and 250 observations, as in Kuester et al. (2006).³³ In this analysis, we can gain the following insights. First, less parameterised conditional models (such as the conditional historical simulation) have an advantage as the window size decreases. In other words, we can observe, for example, that the difference in average ranks between the historical simulation and the Box-Cox method increases with shrinking window size.³⁴ Second, also in the conditional setup, we can partially observe higher values of the backtest statistics for smaller window sizes. This may be because the advantage of using only more recent information could compensate for the disadvantage of lower sample sizes in the GARCH parameter and distribution parameter estimation. Finally, our findings for the 10-year sample are quite consistent with the results (rankings) for our standard setting of 1,000 observations.

Short positions. Our main analysis focuses on the risk of long positions. However, as investors may also be interested in evaluating the risk of short positions, it is instructive to have a look at their VaR as well. Because the gains of the long position are the losses of the short position, we can obtain the VaR of short positions in our assets by simply multiplying our time series by -1 and repeating the calculations of Sections 3 and 5.1 (see Marimoutou et al., 2009). Again the general picture indicates that the Box-Cox method tends to be the best EVT approach and that the historical simulation is a serious competitor.

Alternative VaR backtests. Even though we have selected a large sample to limit small sample issues with the unconditional and conditional coverage tests, we additionally apply two more advanced versions of these tests recently proposed by Ziggel et al. (2014).³⁵ These tests are simulation-based and thus more reliable in small samples. Furthermore, they take into account that conditional coverage also means that the violations are identically distributed – a feature not considered by the standard conditional coverage test. Besides using these tests, we employ a recent extension of the dynamic quantile test proposed by Gaglianone et al. (2011). In both cases, our main results are confirmed, indicating that our sample size is sufficient to avoid known small sample problems of classic backtesting procedures.

5.2.2 Crisis subsamples

In a final robustness check, we have a brief look at how our different VaR estimators behave in times of financial turmoil. This is important because a VaR estimator or hedging method is particularly valuable when it performs well in situations where reliable performance is needed most (see Baur and Lucey, 2010; Baur and McDermott, 2010). To shed light on this issue, we select the two most pronounced financial crises in our sample period: the dotcom crash of 2000 (related to the preceding excessive demand for internet stocks) and the global financial crisis of 2007 to 2009 (caused by the US subprime meltdown).³⁶ The dotcom crash is typically dated from March 2000 to December 2000 (see Ofek and Richardson, 2003) and the global financial crisis from October

³³Note that we have also experimented with alternative threshold quotas q . However, both increasing and decreasing our initial value of 10% negatively influences VaR estimator performance. This indicates that the 10% suggestion of McNeil and Frey (2000) is a quite good guideline for a variety of different time series.

³⁴Kuester et al. (2006) also show that conditional historical simulation is quite robust to the choice of window length but also that there are some non-EVT parametric methods that outperform this method even as the sample size shrinks.

³⁵We have also implemented the bootstrap-based procedures of Escanciano and Olmo (2010) which are designed to address the issue that classic coverage tests are affected by model misspecification in conditional VaR models.

³⁶For an nice summary of the past 40 years in financial crises and a review of literature that attempts to identify, classify and explain such episodes, see Anderson (2013) and Claessens and Kose (2013), respectively.

2007 to March 2009 (see Dwyer, 2009).³⁷ Because these crisis subsamples are too small to conduct reliable statistical inference based on our main testing approaches of Section 4, we concentrate on an analysis of violation rates and their summary in terms of MAPE and MPE.³⁸

Similar to Tables 2 and 4, Tables 13 and 14 present the violation rates which occurred in the two crisis periods.³⁹ A closer look at the results reveals several important aspects. First, in comparison to our results for the entire sample period (containing non-crisis and crisis periods), the violation rates in the crisis periods tend to be significantly higher. Second, the violation rates in the global financial crisis are higher than in the period of the dotcom crash reflecting the stronger impact of the former on financial markets. Finally, for the dotcom crash, we can detect several instances of zero violation rates for gold and bonds; a result partially related to the fact that these two assets typically act as hedges (negative correlation to stocks) in times of stock market turbulences and thus show different risk characteristics (see Ciner et al., 2013).

To judge the relative performance of the VaR estimators in times of crisis, Table 15 summarises the results of Tables 13 and 14 by computing the MAPE and the MPE similar to Tables 3 and 5. The reported results confirm our previous finding of high errors in times of crisis (reflected by three-digit percentage errors). Furthermore, they highlight that, even though EVT-based methods show better performance, historical simulation is a serious competitor.

6 Conclusion

In recent decades, the VaR has become one of the most important risk management tools for market regulators, portfolio managers and internal risk control. This growing importance is reflected by (i) the Bank of International Settlements' capital adequacy requirements on financial institutions that are based on the VaR and by (ii) the significant proliferation of the VaR among non-financial firms (see Martins-Filho and Yao, 2006; Jorion, 2007; Berkowitz et al., 2011).

Traditional VaR models make specific assumptions about the distribution for all returns and thus estimate the maximum loss of an asset under normal market conditions. Thus, the VaR measures may provide inaccurate estimates of actual losses during highly volatile periods corresponding to financial crises (see Martins-Filho and Yao, 2006). In contrast, a recent strand of literature focuses on the distributions of extreme returns instead of the distribution of all returns such that the corresponding VaR estimates have higher potential to provide good predictions of catastrophic market risk during extraordinary periods (see Abad et al., 2014).

Given this development, it is crucial to have a body of research that compares different EVT-based approaches to computing VaR. In this article, we provide the first structured empirical comparison of the predictive performance of the most popular EVT-based VaR estimators with special attention to the peak over threshold estimator of McNeil and Frey (2000), the Box-Cox method of Bali (2007), the L-moment technique of Martins-Filho and Yao (2006) and the simulated Johnson system approach of Brooks et al. (2005). In response to the critique that previous studies analysing VaR estimator performance typically focus on one specific asset class (mainly stocks), we provide a multi-asset study covering stocks, commodities, bonds and currencies. Furthermore, we do not only analyse whether the VaR estimators can pass the current regulatory standard tests but also whether they satisfy additional important properties of an efficient VaR estimator.

Our analysis generates several important insights. First, when applied in their unconditional versions (not considering time-varying means and volatilities of asset returns), we detect that none

³⁷Reinhart and Rogoff (2008), Bartram and Bodnar (2009) and Bordo and Landon-Lane (2010) provide comparisons of the global financial crisis to other crises. Ben-David et al. (2012), Fratzscher (2012) and Flannery et al. (2013) discuss its consequences for hedge fund stock trading, international capital flows and bank opaqueness, respectively.

³⁸Nevertheless, an application of the tests of Section 4 (and 5.2.1) confirms the results of our descriptive analysis.

³⁹Tables A1 and A2 of the Appendix report the corresponding (absolute) averages of daily VaR estimates produced by our different methods.

	CRSP	SP500	GSCI	GOLD	USGBI	\$/£
<i>Panel A: Unconditional estimators</i>						
<i>5% level</i>						
Normal	12.70	7.80	9.17	0.92	2.75	6.88
Student t	13.89	9.17	11.01	2.29	3.67	8.26
Historical simulation	11.11	7.34	9.17	3.67	2.29	7.80
Peak over threshold	12.70	8.26	9.63	3.67	2.75	7.34
L-moment	13.89	10.09	9.63	4.13	2.75	7.34
Box-Cox	9.13	5.50	9.63	1.83	2.75	7.34
Johnson	7.54	4.59	8.26	0.92	1.83	5.96
<i>2.5% level</i>						
Normal	6.75	3.21	7.34	0.00	0.92	4.59
Student t	6.75	3.21	7.34	0.00	0.92	4.13
Historical simulation	6.75	3.21	7.34	0.00	0.92	4.59
Peak over threshold	5.56	2.75	7.34	0.00	0.46	4.59
L-moment	6.75	3.21	7.34	0.00	0.46	4.59
Box-Cox	2.38	0.92	6.88	0.00	0.00	4.59
Johnson	3.57	1.38	6.42	0.00	0.46	3.21
<i>1% level</i>						
Normal	3.57	1.38	5.50	0.00	0.00	2.29
Student t	1.98	0.92	2.75	0.00	0.00	0.92
Historical simulation	2.38	0.92	3.67	0.00	0.00	1.38
Peak over threshold	1.19	0.46	3.21	0.00	0.00	1.38
L-moment	1.59	0.92	3.67	0.00	0.00	1.38
Box-Cox	0.79	0.46	2.29	0.00	0.00	1.38
Johnson	0.79	0.46	1.83	0.00	0.00	0.46
<i>Panel B: Conditional estimators</i>						
<i>5% level</i>						
Normal	7.54	5.96	8.26	3.67	3.67	6.42
Student t	9.52	6.88	8.72	4.13	4.13	7.80
Historical simulation	6.75	5.05	8.26	3.67	3.67	7.80
Peak over threshold	5.95	5.50	8.26	4.13	3.67	7.80
L-moment	6.35	5.50	8.26	4.13	3.67	7.80
Box-Cox	5.16	4.13	8.26	3.67	3.67	7.80
Johnson	3.57	3.21	5.50	2.29	2.29	5.50
<i>2.5% level</i>						
Normal	4.37	3.21	5.05	1.38	2.29	3.21
Student t	4.37	3.21	4.59	0.92	2.29	3.21
Historical simulation	2.78	2.29	3.67	0.92	0.92	3.21
Peak over threshold	2.38	2.29	3.67	0.92	0.92	3.67
L-moment	2.78	2.29	3.67	0.92	0.92	3.21
Box-Cox	1.98	0.92	3.67	0.46	0.00	3.21
Johnson	0.79	0.46	1.83	0.00	0.00	2.75
<i>1% level</i>						
Normal	1.98	0.46	2.29	0.46	0.00	2.29
Student t	0.79	0.46	0.92	0.00	0.00	1.38
Historical simulation	0.79	0.46	1.83	0.00	0.00	1.38
Peak over threshold	0.79	0.46	1.38	0.00	0.00	1.38
L-moment	0.79	0.46	1.38	0.00	0.00	1.38
Box-Cox	0.79	0.46	1.38	0.00	0.00	1.38
Johnson	0.79	0.46	0.92	0.00	0.00	0.00

For the period from March 2000 to December 2000 (the dotcom crash), this table reports the violation rates of our set of unconditional and conditional VaR estimators. Similar to Tables 2 and 4 we focus on three nominal error probabilities and six time series representing several asset classes.

Table 13: Violation rates – Dotcom crash

of the EVT-based VaR estimators is accurate for all time series and confidence levels considered. While there are some approaches that pass as acceptable based on the conditional coverage test, they are rejected by more informative tests, which indicates that regulatory forecasting assessment is flawed and decision makers are directed towards inaccurate approaches for calculating capital requirements. However, despite the poor performance of the models, a ranking of the methods based on their test statistics reveals that the Box-Cox method appears to be the most promising

	CRSP	SP500	GSCI	GOLD	USGBI	\$/£
<i>Panel A: Unconditional estimators</i>						
<i>5% level</i>						
Normal	20.73	20.49	14.29	12.94	14.82	15.63
Student t	21.85	21.29	14.82	14.02	16.17	15.90
Historical simulation	20.73	21.56	13.75	11.59	14.82	14.82
Peak over threshold	21.01	21.29	14.02	11.86	15.36	15.09
L-moment	21.01	20.49	14.02	12.40	15.90	14.82
Box-Cox	20.73	20.49	14.02	11.86	15.36	15.09
Johnson	19.89	21.02	11.05	10.24	12.67	12.67
<i>2.5% level</i>						
Normal	17.93	17.79	10.51	8.89	11.59	11.59
Student t	17.37	16.98	10.24	8.89	11.05	11.32
Historical simulation	15.69	14.29	9.97	6.20	10.78	10.24
Peak over threshold	14.29	14.29	10.24	6.47	10.24	10.78
L-moment	14.01	13.48	9.97	6.47	10.51	10.51
Box-Cox	14.29	12.67	9.70	6.20	9.97	10.78
Johnson	12.61	13.21	9.16	5.66	8.63	8.36
<i>1% level</i>						
Normal	13.45	12.40	8.63	6.20	8.09	9.43
Student t	11.20	11.59	7.28	4.58	6.47	7.55
Historical simulation	8.68	8.36	6.74	2.16	6.20	7.01
Peak over threshold	9.52	9.16	7.55	2.70	5.66	7.01
L-moment	9.52	9.16	7.82	2.70	6.47	7.01
Box-Cox	8.96	8.09	6.20	2.43	5.39	6.74
Johnson	6.44	7.28	6.47	2.16	3.77	5.39
<i>Panel B: Conditional estimators</i>						
<i>5% level</i>						
Normal	10.64	9.97	8.89	6.74	7.01	8.89
Student t	11.76	11.59	9.43	7.28	7.82	10.78
Historical simulation	8.68	8.89	8.63	7.01	6.20	8.09
Peak over threshold	8.96	8.89	8.63	6.47	6.20	8.09
L-moment	8.40	8.89	9.16	5.93	6.47	8.09
Box-Cox	8.12	8.09	8.63	6.20	6.20	8.09
Johnson	6.72	7.28	4.85	4.31	3.77	5.93
<i>2.5% level</i>						
Normal	6.72	7.28	4.58	3.77	2.96	5.66
Student t	6.72	7.28	4.58	3.77	2.70	5.39
Historical simulation	4.76	4.85	4.58	2.43	2.70	4.31
Peak over threshold	5.88	5.66	4.58	2.70	2.70	4.31
L-moment	5.32	5.66	4.58	2.43	2.70	4.31
Box-Cox	5.04	4.58	4.58	2.43	2.43	4.31
Johnson	3.92	3.77	2.43	1.89	0.54	3.23
<i>1% level</i>						
Normal	4.48	4.31	2.43	2.16	0.81	3.50
Student t	3.36	3.50	1.35	1.89	0.54	2.70
Historical simulation	2.52	2.96	2.16	1.35	0.54	2.96
Peak over threshold	2.24	2.96	2.16	1.62	0.54	2.96
L-moment	2.24	2.43	2.16	1.35	0.54	2.96
Box-Cox	1.40	0.81	2.16	1.35	0.54	2.96
Johnson	1.40	1.08	1.62	0.81	0.54	1.89

For the period from October 2007 to March 2009 (the global financial crisis), this table reports the violation rates of our set of unconditional and conditional VaR estimators. Similar to Tables 2 and 4 we focus on three nominal error probabilities and six time series representing several asset classes.

Table 14: Violation rates – Global financial crisis

of all approaches. This method and the Johnson method tend to overestimate risk and thus can be considered the most conservative VaR estimation techniques.

Second, when applying conditional versions of the EVT-based estimators, we obtain far more acceptable forecasts with many instances where the more informative tests are passed. When concentrating on the 1% nominal error probability, which is particularly important for the Basel capital requirements, the Box-Cox method proves to be the best of the EVT-based methods. However, when extending our selection of methods to traditional non-EVT methods, we find

	ND	ST	HS	POT	LM	BC	JO
<i>Panel A: Dotcom crash</i>							
<i>Unconditional estimators</i>							
MAPE	120.58	90.93	93.28	86.30	95.03	72.73	64.16
Rank	7	4	5	3	6	2	1
MPE	66.16	39.82	43.06	29.94	44.81	0.36	-14.58
Rank	7	4	5	3	6	2	1
<i>Conditional estimators</i>							
MAPE	59.31	50.06	45.98	43.24	43.02	49.27	53.48
Rank	7	5	3	2	1	4	6
MPE	24.38	6.79	-5.47	-7.71	-7.40	-18.64	-50.14
Rank	7	6	5	3	4	2	1
<i>Panel B: Global financial crisis</i>							
<i>Unconditional estimators</i>							
MAPE	507.22	454.56	374.84	388.06	391.04	359.80	300.39
Rank	7	6	3	4	5	2	1
MPE	507.22	454.56	374.84	388.06	391.04	359.80	300.39
Rank	7	6	3	4	5	2	1
<i>Conditional estimators</i>							
MAPE	127.14	112.04	80.10	84.33	78.02	61.37	35.91
Rank	7	6	4	5	3	2	1
MPE	125.03	106.93	74.68	79.22	72.60	53.52	12.36
Rank	7	6	4	5	3	2	1

For each of our unconditional and conditional VaR estimators, this table reports the mean absolute percentage errors (MAPE) and mean percentage errors (MPE) across the three error probabilities and six time series in Tables 13 and 14, i.e., the errors in times of financial crises. Similar to Tables 3 and 5 it also presents rankings of the VaR estimators, where rank 1 (rank 7) is the best (worst) methodology in terms of MAPE or MPE.

Table 15: Violation summary – Dotcom crash and global financial crisis

that historical simulation outperforms this EVT-based approach. This result is of high practical importance because many large banks currently use or plan to use historical simulation methods (see Pritsker, 2006; Pérignon and Smith, 2010) but academic research urges them to switch to EVT-based methods (see Abad et al., 2014).

Finally, in a variety of sensitivity checks, we can show that the standard VaR estimator specifications and backtests used in practice (on which we have focused in our empirical analysis) are quite robust. That is, more sophisticated filter procedures, alternative sample size settings in the out-of-sample performance evaluation or simulation-based coverage tests, do not significantly alter the relative performance we have detected. This also holds when focusing on the performance of the VaR estimators in times of financial crisis. Thus, the key message here is that simple specifications can be quite powerful and more complex ones do not necessarily add significant value.

Even though we have used a variety of representative time-series reflecting important characteristics of asset returns and covering many phases of high market turbulences, one might still argue that conclusions based on these series cannot be sufficiently generalised. Future research might address this issue in two ways. First, the set of test assets could be significantly extended. For example, it might be interesting to have a look at the performance of EVT-based VaR estimators for highly liquid individual stocks (such as the components of the Dow Jones Industrial Average; see Taylor, 2014; Auer and Schuhmacher, 2015) or futures (such as the subindices of the GSCI; see Auer, 2015; Bianchi et al., 2015) which are typical components of investors' portfolios. Second one might conduct a suitable simulation study in which our VaR estimators are compared in different settings. However, such studies have problems of their own. Previous VaR-related simulation studies have typically generated random numbers based on very simple distributions (see McNeil and Frey, 2000) or standard stochastic processes (see Martins-Filho and Yao, 2006).

While such approaches allow determining the impact of, for example, different degrees of volatility clustering on VaR estimator performance, their general drawbacks are that the used theoretical models (i) naturally cannot capture all relevant features observed in empirical data and, (ii) if they can, the characteristics are highly idealised. Consequently, simulated extremes may be quite different from what we actually observe in empirical time series. Therefore, in order to fully understand differences in VaR estimator performance, future simulation-based comparisons must come up with more complex processes or rely on model-free bootstrap procedures delivering more realistic simulated time-series (for an interesting example, see Kinateder, 2016).

Another interesting topic for future research is related to the risk measure Expected Shortfall (ES) which is defined as the conditional expectation $ES_\alpha = E(r_t | r_t \leq VaR_\alpha)$ (see Eling and Schuhmacher, 2007) and has received increased interest in recent research because of its more favourable theoretical properties in comparison to the VaR (see Yamai and Yoshida, 2005).⁴⁰ While the VaR asks the question 'How bad can things get?', ES asks 'If things do get bad, what is our expected loss?'. By the definition of ES, errors in VaR measurement are linked to errors in ES estimates. However, for a closer look at the performance of ES estimators originating from our EVT approaches, we require two important inputs. First, we have to derive the corresponding explicit ES formulas which can usually be found in the articles proposing a new VaR estimation method (in our case, for example, McNeil and Frey, 2000; Martins-Filho and Yao, 2006) or in review articles covering a variety of different ES estimation techniques (see Nadarajah et al., 2014). Second, we have to consider that the evaluation and backtesting of ES is more complicated and requires alternative procedures and tests (see Wong, 2008). With these tools at hand, future research could contribute to a better understanding of the performance of EVT-based ES estimators which is important because most recent Basel Accords put significant emphasis on the ES as a preferable measure of market risk (see Chen, 2014).

⁴⁰Nevertheless, ES has its own shortcomings. For example, it is not consistent with right tail risk as measured by the convex order of degree three (see Hürlimann, 2004).

References

- Abad, P., Benito, S., López, C., 2014. A Comprehensive Review of Value at Risk Methodologies. *Spanish Review of Financial Economics* 12 (1), 15–32.
- Alexander, C., 2008. *Market Risk Analysis. Vol. IV - Value-at-Risk Models*. John Wiley & Sons, Chichester.
- Anderson, S., 2013. A History of the Past 40 Years in Financial Crises. *International Financing Review* 2000, 48–52.
- Angelidis, T., Benos, A., Degiannakis, S., 2004. The Use of GARCH Models in VaR Estimation. *Statistical Methodology* 1 (1-2), 105–128.
- Artzner, P., Delbaen, F., Eber, J., Heath, D., 1999. Coherent Measures of Risk. *Mathematical Finance* 9 (3), 203–228.
- Auer, B., 2015. Does the Choice of Performance Measure Influence the Evaluation of Commodity Investments? *International Review of Financial Analysis* 38, 142–150.
- Auer, B., Schuhmacher, F., 2015. Liquid Betting Against Beta in Dow Jones Industrial Average Stocks. *Financial Analysts Journal* 71 (6), 30–43.
- Baker, G. A., 1934. Transformation of Non-Normal Frequency Distributions into Normal Distributions. *Annals of Mathematical Statistics* 5 (2), 113–123.
- Bali, T., 2003. The Generalized Extreme Value Distribution. *Economics Letters* 79 (3), 423–427.
- Bali, T., 2007. A Generalized Extreme Value Approach to Financial Risk Measurement. *Journal of Money, Credit and Banking* 39 (7), 1613–1649.
- Bali, T., Neftci, S., 2003. Disturbing Extremal Behavior of Spot Rate Dynamics. *Journal of Empirical Finance* 10 (4), 455–477.
- Bali, T., Theodossiou, P., 2007. A Conditional-SGT-VaR-Approach with Alternative GARCH Models. *Annals of Operations Research* 151 (1), 241–267.
- Bali, T., Weinbaum, D., 2007. A Conditional Extreme Value Volatility Estimator Based on High Frequency Data. *Journal of Economic Dynamics and Control* 31 (2), 361–397.
- Bali, T. G., Mo, H., Tang, Y., 2008. The Role of Autoregressive Conditional Skewness and Kurtosis in the Estimation of Conditional VaR. *Journal of Banking and Finance* 32 (2), 269–282.
- Balkema, A., de Haan, L., 1974. Residual Life Time at Great Age. *Annals of Probability* 2 (5), 792–804.
- Barone-Adesi, G., Giannopoulos, K., 2001. Non-Parametric VaR Techniques: Myths and Realities. *Economic Notes* 30 (2), 167–181.
- Bartram, S., Bodnar, G., 2009. No Place to Hide: The Global Crisis in Equity Markets in 2008/2009. *Journal of International Money and Finance* 28 (8), 1246–1292.
- Basel Committee on Banking Supervision, 1995. An Internal Model-Based Approach to Market Risk Capital Requirements, www.bis.org.
- Basel Committee on Banking Supervision, 1996a. Overview of the Amendment to the Capital Accord to Incorporate Market Risks, www.bis.org.
- Basel Committee on Banking Supervision, 1996b. Supervisory Framework for the Use of "Backtesting" in Conjunction with the Internal Models Approach to Market Risk Capital Requirements, www.bis.org.
- Basel Committee on Banking Supervision, 2009. Revisions to the Basel II Market Risk Framework, www.bis.org.
- Basel Committee on Banking Supervision, 2011a. Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems, www.bis.org.
- Basel Committee on Banking Supervision, 2011b. Messages from the Academic Literature on Risk Measurement for the Trading Book, www.bis.org.
- Baur, D., Lucey, B., 2010. Is Gold a Hedge or a Safe Haven? An Analysis of Stocks, Bonds and Gold. *Financial Review* 45 (2), 217–229.
- Baur, D., McDermott, T., 2010. Is Gold a Safe Haven? International Evidence. *Journal of Banking and Finance* 34 (8), 1886–1898.
- Ben-David, I., Franzoni, F., Moussawi, R., 2012. Hedge Fund Stock Trading in the Financial Crisis of 2007-2009. *Review of Financial Studies* 25 (1), 1–54.
- Berkowitz, J., 2001. Testing Density Forecasts with Applications to Risk Management. *Journal of Business and Economic Statistics* 19 (4), 465–474.
- Berkowitz, J., Christoffersen, P., Pelletier, D., 2011. Evaluating Value-at-Risk Models with Desk-Level Data. *Management Science* 57 (12), 2213–2227.
- Berkowitz, J., O'Brien, J., 2002. How Accurate Are Value-at-Risk Models at Commercial Banks? *Journal of Finance* 57 (3), 1093–1111.
- Bianchi, R., Drew, M., Fan, J., 2015. Combining Momentum with Reversal in Commodity Futures. *Journal of Banking and Finance* 59, 423444.
- Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31 (3), 307–327.
- Bollerslev, T., Chou, R., Kroner, K., 1992. ARCH-Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics* 52 (1-2), 5–59.
- Bollerslev, T., Engle, R., Nelson, D., 1994. ARCH Models. In: Engle, R., McFadden, D. (Eds.), *Handbook of Econometrics. Vol. 4*. Elsevier, Amsterdam, pp. 2959–3038.

- Bordo, M., Landon-Lane, J., 2010. The Global Financial Crisis of 2007-08: Is it Unprecedented?, NBER Working Paper No. 16589.
- Bowman, K., Shenton, L., 2004. Johnson's System of Distributions. In: Encyclopedia of Statistical Sciences. John Wiley & Sons, New York.
- Box, G., Cox, D., 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society, Series B* 26 (2), 211–252.
- Brooks, C., Clare, A. D., Dalle Molle, J. W., Persaud, G., 2005. A Comparison of Extreme Value Theory Approaches for Determining Value At Risk. *Journal of Empirical Finance* 12 (2), 339–352.
- Campbell, J., Grossman, S., Wang, J., 1993. Trading Volume and Serial Correlation in Stock Returns. *Quarterly Journal of Economics* 108 (4), 905–939.
- Campbell, J., Lo, A., MacKinlay, A., 1997. *The Econometrics of Financial Markets*. Princeton University Press, Princeton.
- Campbell, R., Huisman, R., Koedijk, K., 2001. Optimal Portfolio Selection in a Value-at-Risk Framework. *Journal of Banking and Finance* 25 (9), 1789–1804.
- Campbell, S., 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 Test. *Journal of Financial Econometrics* 9 (2), 314–343.
- Carroll, R., Härdle, W., Mammen, E., 2002. Estimation in an Additive Model when the Components are Linked Parametrically. *Econometric Theory* 18 (4), 886–912.
- Castillo, E., 1988. *Extreme Value Theory in Engineering*. Academic Press, New York.
- Chen, J., 2014. Measuring Market Risk Under the Basel Accords: VaR, Stressed VaR, and Expected Shortfall. *IEB International Journal of Finance* 8, 184–201.
- Chen, Y., Lu, J., 2012. Value at Risk Estimation. In: Duan, J., Härdle, W., Gentle, J. (Eds.), *Handbook of Computational Finance*. Springer, Berlin Heidelberg, pp. 307–334.
- Cheng, W., Hung, J.-C., 2011. Skewness and Leptokurtosis in GARCH-Type VaR Estimation of Petroleum and Metal Asset Returns. *Journal of Empirical Finance* 18 (1), 160–173.
- Christoffersen, P., 1998. Evaluating Interval Forecasts. *International Economic Review* 39 (4), 841–862.
- Christoffersen, P., 2003. *Elements of Financial Risk Management*. Academic Press, New York.
- Ciner, C., Gurdgiev, V., Lucey, B., 2013. Hedges and Safe Havens: An Examination of Stocks, Bonds, Gold, Oil and Exchange Rates. *International Review of Financial Analysis* 29, 202–211.
- Claessens, S., Kose, M., 2013. *Financial Crises: Explanations, Types, and Implications*, IMF Working Paper No. 13/28.
- Cont, R., 2001. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance* 1 (2), 223–236.
- Cox, D., Hinkley, D., 1974. *Theoretical Statistics*. Chapman and Hall, London.
- Danielsson, J., de Vries, C., 2000. Value-at-Risk and Extreme Returns. *Annales d'Économie et de Statistique* 60, 239–270.
- Danielsson, J., Morimoto, Y., 2000. Forecasting Extreme Financial Risk: A Critical Analysis of Practical Methods for the Japanese Market. *Monetary and Economic Studies* 18 (2), 25–48.
- De Haan, L., Resnick, S., 1980. A Simple Asymptotic Estimate for the Index of a Stable Distribution. *Journal of the Royal Statistical Society* 42 (1), 83–87.
- Diebold, F. X., Schuermann, T., Stroughair, J., 2000. Pitfalls and Opportunities in the Use of Extreme Value Theory in Risk Management. *Journal of Risk Finance* 1 (2), 30–35.
- Diebold, J., Guégan, D., 1993. Tail Behaviour of the Stationary Density of General Non-Linear Autoregressive Processes of Order 1. *Journal of Applied Probability* 30 (2), 315–329.
- Domitrescu, E., Hurlin, C., Pham, V., 2012. Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Finance* 33 (1), 79–122.
- Dwyer, G., September 2009. *Stock Prices in the Financial Crisis*, Notes from the Vault, Federal Reserve Bank of Atlanta.
- Efron, B., 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia.
- El-Aroui, M., Diebold, J., 2002. On the Use of the Peaks Over Thresholds Method for Estimating Out-of-Sample Quantiles. *Computational Statistics and Data Analysis* 39 (4), 453–475.
- Eling, M., Schuhmacher, F., 2007. Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds? *Journal of Banking and Finance* 31 (9), 2632–2647.
- Embrechts, P., Klppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events: For Insurance and Finance*. Springer, Berlin.
- Embrechts, P., Resnick, S., Samorodnitsky, G., 1999. Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal* 3 (2), 30–41.
- Engle, R., Manganelli, S., 2004. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22 (4), 367–381.
- Engle, R., Patton, A., 2001. What Good is a Volatility Model? *Quantitative Finance* 1 (2), 237–245.

- Escanciano, J., Olmo, J., 2010. Backtesting Parametric Value-at-Risk With Estimation Risk. *Journal of Business and Economic Statistics* 28 (1), 36–51.
- Fan, J., 1992. Design-Adaptive Nonparametric Regression. *Journal of the American Statistical Association* 87, 998–1004.
- Fan, J., Yao, Q., 1998. Efficient Estimation of Conditional Variance Functions in Stochastic Regression. *Biometrika* 85 (3), 645–660.
- Ferreira, A., De Haan, L., 2015. On the Block Maxima Method in Extreme Value Theory: PWM Estimators. *Annals of Statistics* 43 (1), 276–298.
- Fisher, R. A., Tippett, L. H. C., 1928. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24 (2), 180–190.
- Flannery, M., Kwan, S., Nimalendran, M., 2013. The 2007-2009 Financial Crisis and Bank Opacity. *Journal of Financial Intermediation* 22 (1), 55–84.
- Florence, G., Ramachandran, K., 2011. Estimation of Parameters of Johnson’s System of Distributions. *Journal of Modern Applied Statistical Methods* 10 (2), 494–504.
- Fratzscher, M., 2012. Capital Flows, Push Versus Pull Factors and the Global Financial Crisis. *Journal of International Economics* 88 (2), 341–356.
- Gaglianone, W., Lima, L., Linton, O., Smith, D., 2011. Evaluating Value-at-Risk Models via Quantile Regressions. *Journal of Business and Economic Statistics* 29 (1), 150–160.
- 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.
- Gilli, M., Kellezi, E., 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics* 27 (2), 207–228.
- Gnedenko, B., 1943. Sur La Distribution Limite Du Terme D’une Série Aléatoire. *Annals of Mathematics* 4 (3), 423–453.
- Gumbel, E., 1958. *Statistics of Extremes*. Columbia University Press, New York.
- Hafner, C., 1998. *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility*. Physica, Heidelberg.
- Hansen, B., 1994. Autoregressive Conditional Density Estimation. *International Economic Review* 35 (3), 705–730.
- Härdle, W., Tsybakov, A., 1997. Local Polynomial Estimators of the Volatility Function in NonParametric Autoregression. *Journal of Econometrics* 81 (1), 223–242.
- Hill, B., 1975. A Simple General Approach to Inference about the Tail of a Distribution. *Annals of Statistics* 3 (5), 1163–1174.
- Ho, L., Burridge, P., Cadle, J., Theobald, M., 2000. Value-at-Risk: Applying the Extreme Value Approach to Asian Markets in the Recent Financial Turmoil. *Pacific-Basin Finance Journal* 8 (2), 249–275.
- Hoel, P., 1954. A Test for Markoff Chains. *Biometrika* 41 (3/4), 430–433.
- Hosking, J., 1989. Some Theoretical Results Concerning L-Moments, IBM Research Division, Technical Report.
- Hosking, J., 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society B* 52 (1), 105–124.
- Hosking, J., Wallis, J., 1987. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29 (3), 339–349.
- Hosking, J., Wallis, J., Wood, E., 1985. Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments. *Technometrics* 27 (3), 251–261.
- Hosking, J. R. M., Wallis, J. R., 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, Cambridge.
- Hsieh, D., 1993. Implications of Nonlinear Dynamics for Financial Risk Management. *Journal of Financial and Quantitative Analysis* 28 (1), 41–64.
- Huang, Y., Lin, B., 2004. Value-at-Risk Analysis for Taiwan Stock Index Futures: Fat Tails and Conditional Asymmetries in Return Innovations. *Review of Quantitative Finance and Accounting* 22 (2), 79–95.
- Huisman, R., Koedijk, K., Kool, C., Palm, F., 2001. Tail Index Estimates in Small Samples. *Journal of Business and Economic Statistics* 19 (2), 208–216.
- Hürlimann, W., 2004. Distortion Risk Measures and Economic Capital. *North American Actuarial Journal* 8 (1), 86–96.
- Hwang, S., Vallis Pereira, P., 2006. Small Sample Properties of GARCH Estimates and Persistence. *European Journal of Finance* 12 (6-7), 473–494.
- Jenkinson, A., 1955. The Frequency Distribution of the Annual Maximum (Minimum) Values of Meteorological Elements. *Quarterly Journal of the Royal Meteorological Society* 81 (348), 158–171.
- Johnson, N., 1949. Systems of Frequency Curves Generated by Methods of Translation. *Biometrika* 36 (1/2), 149–176.
- Jorion, P., 2007. *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd Edition. Mc-Graw-Hill, New York.
- Kinateder, H., 2016. Basel II versus III - A Comparative Assessment of Minimum Capital Requirements for Internal Model Approaches. *Journal of Risk* 18 (3), 25–45.

- Kuester, K., Mittnik, S., Paoletta, M., 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* 3 (2), 73–84.
- Landwehr, J., Matalas, N., Wallis, J., 1979. Probability Weighted Moments Compared with some Traditional Techniques in Estimating Gumbel Parameters and Quantiles. *Water Resources Research* 15 (5), 1055–1064.
- Lange, K., Little, R., Taylor, J., 1989. Robust Statistical Modeling Using the t Distribution. *Journal of the American Statistical Association* 84 (408), 881–896.
- Leadbetter, M., Lindgren, G., Rootzén, H., 1983. *Extremes and Related Properties of Random Sequences and Processes*. Springer, New York.
- Lee, C., Su, J., 2012. Alternative Statistical Distributions for Estimating Value-at-Risk: Theory and Evidence. *Review of Quantitative Finance and Accounting* 39 (3), 309–331.
- Linsmeier, T., Pearson, N., 2000. Value at Risk. *Financial Analysts Journal* 56 (2), 47–67.
- Linton, O., Nielsen, J., 1995. A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. *Biometrika* 82 (1), 93–100.
- Ljung, G., Box, G., 1978. On a Measure of Lack of Fit in Time Series Models. *Biometrika* 65 (2), 297–303.
- Longin, F., 2000. From Value at Risk to Stress Testing: The Extreme Value Approach. *Journal of Banking and Finance* 24 (7), 1097–1130.
- Lopez, J., 1997. Regulatory Evaluation of Value-at-Risk Models, Federal Reserve Bank of New York, Staff Report No. 33.
- Lopez, J., 1999. Methods for Evaluating Value-at-Risk Estimates. *Federal Reserve Bank of San Francisco Economic Review* 2, 3–17.
- Marimoutou, V., Raggad, B., Trabelsi, A., 2009. Extreme Value Theory and Value at Risk: Application to Oil Market. *Energy Economics* 31 (4), 519–530.
- Martins-Filho, C., Yao, F., 2006. Estimation of Value-at-Risk and Expected Shortfall based on Nonlinear Models of Return Dynamics and Extreme Value Theory. *Studies in Nonlinear Dynamics & Econometrics* 10 (2), Article 4, 1–41.
- McNeil, A., 1999. Extreme Value Theory for Risk Managers. In: *Internal Modeling and CAD II*. Risk Books, London, pp. 93–113.
- 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 (3-4), 271–300.
- Nadarajah, S., Zhang, B., Chan, S., 2014. Estimation Methods for Expected Shortfall. *Quantitative Finance* 14 (2), 271–291.
- Neftci, S., 2000. Value at Risk Calculations, Extreme Events, and Tail Estimation. *Journal of Derivatives* 7 (3), 23–37.
- Ofek, E., Richardson, M., 2003. DotCom Mania: The Rise and Fall of Internet Stock Prices. *Journal of Finance* 58 (3), 1113–1137.
- Oja, H., 1981. On Location, Scale, Skewness and Kurtosis of Univariate Distributions. *Scandinavian Journal of Statistics* 8 (3), 154–168.
- Paoletta, M., 2001. Testing the Stable Paretian Assumption. *Mathematical and Computer Modelling* 34, 1095–1112.
- Patton, A., 2004. On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation. *Journal of Financial Econometrics* 2 (1), 130–168.
- Pérignon, C., Smith, D., 2010. The Level and Quality of Value-at-Risk Disclosure by Commercial Banks. *Journal of Banking and Finance* 34 (9-11), 362–377.
- Pflug, G., Römisch, W., 2007. *Modeling, Measuring and Managing Risk*. World Scientific, Singapore.
- Pickands, J., 1975. Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics* 3 (1), 119–131.
- Pritsker, M., 2006. The Hidden Dangers of Historical Simulation. *Journal of Banking and Finance* 30 (2), 561–582.
- Reinhart, C., Rogoff, K., 2008. Is the 2007 US-Sub-Prime Financial Crisis so Different? An International Comparison. *American Economic Review* 98 (2), 339–344.
- Rosnick, S., 1987. *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.
- Rocco, M., 2014. Extreme Value Theory in Finance: A Survey. *Journal of Economic Surveys* 28 (1), 82–108.
- Ruppert, S., Wand, M., 1995. An Effective Bandwidth Selection for Local Least Squares Regression. *Journal of the American Statistical Association* 90 (432), 1257–1270.
- Scarrott, C., MacDonald, A., 2012. A Review of Extreme Value Threshold Estimation and Uncertainty Quantification. *REVSTAT - Statistical Journal* 10 (1), 33–60.
- Silvennoinen, A., Teräsvirta, T., 2009. Multivariate GARCH Models. In: Anderson, T., Davis, R., Kreiß, J., Mikosch, T. a. (Eds.), *Handbook of Financial Time Series*. Springer, Berlin Heidelberg, pp. 201–229.
- Smith, R., 1984. Threshold Methods for Sample Extremes. In: de Oliveira, J. (Ed.), *Statistical Extremes and Applications*. Springer, Dordrecht, pp. 621–638.
- Smith, R., 1987. Estimating Tails of Probability Distributions. *Annals of Statistics* 15 (3), 1174–1207.
- Tauchen, G., 2001. Notes on Financial Econometrics. *Journal of Econometrics* 100 (1), 57–64.
- Taylor, N., 2014. The Rise and Fall of Technical Trading Rule Success. *Journal of Banking and Finance* 40, 286–302.

- Tsay, R., 2005. *Analysis of Financial Time Series*, 2nd Edition. Wiley & Sons, Hoboken.
- Vogel, R., Fennessey, N., 1993. L Moment Diagrams Should Replace Product Moment Diagrams. *Water Resources Research* 29 (6), 1745–1752.
- von Mises, R., 1954. La Distribution De La Plus Grande de n Valeurs. In: *AMS Selected Papers*. Vol. 2. American Statistical Society, Providence, pp. 271–294.
- Wang, Q., 1990. Estimation of the GEV Distribution from Censored Samples by Method of Partial Probability Weighted Moments. *Journal of Hydrology* 120 (1-4), 103–114.
- Wong, W., 2008. Backtesting Trading Risk of Commercial Banks Using Expected Shortfall. *Journal of Banking and Finance* 32 (7), 1404–1415.
- Yamai, Y., Yoshida, T., 2002. Comparative Analyses of Expected Shortfall and Value-at-Risk: Their Estimation Error, Decomposition, and Optimization. *Monetary and Economic Studies* 20 (1), 87–122.
- Yamai, Y., Yoshida, T., 2005. Value-at-Risk Versus Expected Shortfall: A Practical Perspective. *Journal of Banking and Finance* 29 (4), 997–1015.
- Ziggel, D., Berens, T., Weiß, G., Wied, D., 2014. A New Set of Improved Value-at-Risk Backtests. *Journal of Banking and Finance* 48, 29–41.

Appendix

	CRSP	SP500	GSCI	GOLD	USGBI	\$/£
<i>Panel A: Unconditional estimators</i>						
<i>5% level</i>						
Normal	1.80	1.88	1.80	1.40	0.65	0.77
Student t	1.71	1.78	1.71	1.33	0.62	0.73
Historical simulation	1.83	1.86	1.79	1.27	0.67	0.73
Peak over threshold	1.78	1.82	1.78	1.23	0.66	0.74
L-moment	1.71	1.76	1.75	1.20	0.66	0.74
Box-Cox	1.94	2.00	1.78	1.34	0.66	0.74
Johnson	2.07	2.12	2.03	1.49	0.74	0.82
<i>2.5% level</i>						
Normal	2.16	2.26	2.15	1.66	0.78	0.92
Student t	2.20	2.29	2.18	1.68	0.79	0.93
Historical simulation	2.25	2.25	2.18	1.61	0.84	0.94
Peak over threshold	2.33	2.38	2.21	1.64	0.85	0.92
L-moment	2.21	2.28	2.16	1.57	0.86	0.92
Box-Cox	2.80	2.90	2.22	1.97	0.87	0.92
Johnson	2.69	2.69	2.45	2.07	0.92	1.06
<i>1% level</i>						
Normal	2.58	2.69	2.55	1.96	0.92	1.09
Student t	2.91	3.02	2.86	2.19	1.03	1.22
Historical simulation	2.76	2.99	2.74	2.16	1.04	1.17
Peak over threshold	3.13	3.20	2.77	2.21	1.09	1.16
L-moment	2.97	3.08	2.72	2.14	1.10	1.17
Box-Cox	4.19	4.33	2.81	2.99	1.14	1.17
Johnson	3.52	3.38	3.13	2.83	1.23	1.31
<i>Panel B: Conditional estimators</i>						
<i>5% level</i>						
Normal	2.26	2.06	2.11	1.21	0.60	0.82
Student t	2.14	1.95	2.00	1.15	0.57	0.78
Historical simulation	2.32	2.16	2.11	1.19	0.60	0.78
Peak over threshold	2.36	2.12	2.10	1.18	0.60	0.78
L-moment	2.33	2.10	2.08	1.19	0.60	0.79
Box-Cox	2.42	2.22	2.10	1.21	0.60	0.77
Johnson	2.81	2.50	2.37	1.30	0.69	0.90
<i>2.5% level</i>						
Normal	2.71	2.47	2.51	1.43	0.72	0.97
Student t	2.75	2.51	2.55	1.46	0.73	0.99
Historical simulation	3.08	2.71	2.63	1.57	0.78	1.00
Peak over threshold	3.03	2.74	2.60	1.55	0.77	0.96
L-moment	2.98	2.70	2.57	1.56	0.77	0.99
Box-Cox	3.28	3.06	2.60	1.66	0.80	0.97
Johnson	3.64	3.18	2.95	1.77	0.87	1.12
<i>1% level</i>						
Normal	3.24	2.95	2.98	1.69	0.85	1.15
Student t	3.64	3.32	3.34	1.88	0.96	1.29
Historical simulation	3.81	3.47	3.06	2.03	1.00	1.21
Peak over threshold	3.99	3.64	3.23	2.03	0.99	1.22
L-moment	3.93	3.59	3.18	2.04	1.00	1.24
Box-Cox	4.60	4.38	3.23	2.34	1.07	1.27
Johnson	4.27	3.88	3.64	2.50	1.08	1.39

For the period from March 2000 to December 2000 (the dotcom crash), this table reports the (absolute) average of daily VaR estimates produced by our set of unconditional and conditional VaR estimators. Similar to Tables 6 to 11 we focus on three nominal error probabilities and six time series representing several asset classes.

Table A1: Mean VaR estimates – Dotcom crash

	CRSP	SP500	GSCI	GOLD	USGBI	\$/£
<i>Panel A: Unconditional estimators</i>						
<i>5% Level</i>						
Normal	1.59	1.59	2.44	1.90	0.66	0.89
Student t	1.50	1.51	2.32	1.80	0.63	0.84
Historical simulation	1.56	1.55	2.44	2.02	0.67	0.91
Peak over threshold	1.55	1.55	2.42	2.01	0.66	0.90
L-moment	1.57	1.57	2.39	1.97	0.65	0.91
Box-Cox	1.58	1.59	2.44	2.03	0.66	0.89
Johnson	1.64	1.58	2.76	2.12	0.72	1.03
<i>2.5% Level</i>						
Normal	1.89	1.90	2.92	2.28	0.79	1.06
Student t	1.92	1.93	2.96	2.32	0.80	1.08
Historical simulation	2.06	2.11	2.99	2.70	0.84	1.14
Peak over threshold	2.07	2.11	2.98	2.73	0.85	1.13
L-moment	2.09	2.14	2.94	2.67	0.83	1.14
Box-Cox	2.18	2.25	3.04	2.82	0.85	1.14
Johnson	2.29	2.25	3.35	2.88	0.91	1.30
<i>1% Level</i>						
Normal	2.25	2.26	3.47	2.72	0.94	1.26
Student t	2.52	2.53	3.89	3.05	1.05	1.41
Historical simulation	3.02	3.15	3.74	3.70	1.07	1.46
Peak over threshold	2.80	2.92	3.73	3.65	1.10	1.44
L-moment	2.84	2.96	3.68	3.60	1.06	1.45
Box-Cox	3.10	3.27	3.84	3.93	1.10	1.47
Johnson	3.40	3.44	4.09	4.11	1.21	1.63
<i>Panel B: Conditional estimators</i>						
<i>5% Level</i>						
Normal	2.92	2.94	3.36	2.75	1.05	1.28
Student t	2.76	2.79	3.19	2.61	1.00	1.21
Historical simulation	3.15	3.16	3.43	2.77	1.07	1.35
Peak over threshold	3.08	3.11	3.42	2.79	1.05	1.33
L-moment	3.15	3.17	3.38	2.83	1.05	1.33
Box-Cox	3.18	3.21	3.42	2.82	1.05	1.33
Johnson	3.64	3.63	3.91	3.21	1.20	1.51
<i>2.5% Level</i>						
Normal	3.49	3.51	4.02	3.29	1.25	1.52
Student t	3.54	3.57	4.08	3.34	1.27	1.55
Historical simulation	4.03	4.14	4.06	3.66	1.29	1.66
Peak over threshold	3.81	3.92	4.11	3.67	1.29	1.65
L-moment	3.91	4.01	4.06	3.75	1.27	1.65
Box-Cox	4.17	4.32	4.11	3.81	1.29	1.65
Johnson	4.62	4.67	4.67	4.19	1.46	1.87
<i>1% Level</i>						
Normal	4.15	4.18	4.78	3.91	1.48	1.81
Student t	4.66	4.69	5.36	4.39	1.66	2.03
Historical simulation	4.82	4.98	4.97	5.16	1.49	1.99
Peak over threshold	4.83	5.07	4.93	4.89	1.58	2.03
L-moment	4.90	5.12	4.91	5.00	1.57	2.02
Box-Cox	5.67	6.02	4.93	5.22	1.60	2.03
Johnson	5.79	5.91	5.35	5.46	1.78	2.30

For the period from October 2007 to March 2009 (the global financial crisis), this table reports the (absolute) average of daily VaR estimates produced by our set of unconditional and conditional VaR estimators. Similar to Tables 6 to 11 we focus on three nominal error probabilities and six time series representing several asset classes.

Table A2: Mean VaR estimates – Global financial crisis