The Role of Model Risk in Extreme Value Theory for Capital Adequacy [*]

RALF KELLNER, DANIEL RÖSCH, HARALD SCHEULE[†]

November 19, 2015

Abstract

In the recent literature, methods from extreme value theory (EVT) have frequently been applied for the estimation of tail risk measures. While previous analyses show that EVT methods often lead to accurate estimates for risk measures, a potential drawback lies in high standard errors of point estimates of these methods as only a fraction of the data set is used. Thus, the aim of this paper is to comprehensively study the impact of model risk on EVT methods when determining the Value-at-Risk and Expected Shortfall. We distinguish between first order effects of model risk, which consist of misspecification and estimation risk, and second order effects of model risk which refer to the dispersion of risk measure estimates. We show that EVT methods are less prone to first order effects of model risk, however, they exhibit a higher sensitivity towards second order effects of model risk. We find that this can lead to severe Value-at-Risk and Expected Shortfall underestimations and should be reflected in regulatory capital models.

Keywords: Extreme Value Theory; Model Risk; Capital Requirements; Value-at-Risk; Expected Shortfall

I. INTRODUCTION

The accurate assessment of tail risk measures such as the Value-at-Risk and Expected Shortfall is of high relevance in practice and thus, subject to many scientific studies (see, e.g., Berkowitz and O'Brien, 2002; Nekhili et al., 2004; Taylor, 2008; Mancini and Trojani, 2011). A possibility for the estimation of tail risk measures is given through the application of methods from extreme value theory (EVT). Results show that these methods are capable of estimating tail risk measures and that their application can lead to more accurate estimates than traditional methods (see, e.g., Ghorbel and Trabelsi, 2009; Rufino and de Guia, 2011).

A potential drawback associated with EVT methods is that they only use a fraction of the data during the estimation process and hence, usually exhibit higher standard errors for their parameter estimates (see McNeil et al., 2005). This feature is of special relevance in the context of model risk which has gained increasing attention (see Basel Committee on Banking Supervision, 2010).

The 'true' data generating process (DGP) which generates realizations we can observe is unknown and has to be approximated through model assumptions and estimated model parameters. This exposes the quantification of all measures related to this process to model risk. In this paper we will differentiate between first order and second order effects of model risk.

First order effects of model risk consist of misspecification and estimation risk. Misspecfication risk describes the risk of choosing the wrong DGP as the true process. Estimation risk refers to estimating parameters which deviate to those of the true risk model. Parameter estimates are uncertain and the dispersion of the predictions for risk measures (which are based on parameter estimates and hence random numbers) increases with this uncertainty. This enhances the likelihood of more severe under- and overestimations for the true risk measure. We refer to this risk as second order effects of model risk which seem to be of special relevance for EVT methods as in comparison to models using all available data, standard errors and statistical quality of parameter estimates tend to be lower by nature. Thus, we define a first order effect as an effect of model specification and estimation, and a second order effect as an effect on the dispersion of a risk measure due to applying the estimated model.

The aim of this study is to comprehensively analyze first and second order effects of model risk related to EVT methods and compare these results with traditional methods. Moreover, we show which conclusions can be drawn with respect to the quantification of capital requirements and how these requirements can be determined in the presence of first and second order effects of model risk. To achieve this goal, this paper analyzes the accuracy and model risk of risk measures for market risk in relation to popular equity indices using in a broad range of applications.

Various and diverse applications of EVT methods exist in the recent literature. Actuarial studies use EVT for the estimation of losses. For example, in an early publication Rootzén and Tajvidi (1997) show how EVT can be used to assess large insurance claims. Brodin and Rootzén (2009) extend this analysis to a multivariate framework in which they show that a bivariate EVT model is well suited for the quantification of wind storm losses. A further contribution of EVT methods in the insurance literature is provided by Cébrian et al.

(2003) who use multivariate EVT to model the dependence between medical claims and corresponding costs for claims processing. Hereby, they show how their framework may lead to a lower degree of mispricing as it better captures the dependence structure.

Financial studies usually try to analyze the dependency of extreme values and their asymptotic behavior or focus on the estimation of the Value-at-Risk for financial data. Regarding the dependence structure, Longin and Solnik (2001) conduct an empirical analysis which uses EVT to examine asymptotic dependence behavior among financial markets. In their work, they find asymptotic dependence during bear markets not to converge towards zero, which strongly speaks against a Gaussian dependence structure. A further contribution in this field is given by Poon et al. (2004) who show how a multivariate EVT method can be used for the assessment of portfolio risk and how the true risk situation may be underestimated if other methods are applied, not taking into account asymptotic dependence. A similar approach is also used by Zhou (2010) who focuses on the development of an EVT framework in order to identify diversification effects in the presence of tail dependence.

Within these financial studies, the majority of prior contributions focuses on Value-at-Risk estimation using methods from EVT and compares corresponding results to so called classical or traditional models (see, e.g., Longin, 2000) with respect to backtesting performance. This approach can be seen in Longin (2000); McNeil and Frey (2000); Nekhili et al. (2004); Marinelli et al. (2007); Hotta et al. (2008); Ghorbel and Trabelsi (2009). Traditional models usually use time conditional methods for the mean and volatility of time series process (mostly ARMA-GARCH models or similar structures) in combination with the normal, Student t or empirical distribution (see, e.g., Longin, 2000; Hotta et al., 2008; Ghorbel and Trabelsi, 2009). All these studies show that univariate EVT methods perform better with respect to a correct risk assessment of the Value-at-Risk in almost all instances.

Furthermore, there has been steady research in the field of model risk. While Jorion (1996) analyzes the sample error related to Value-at-Risk estimates, Gibson et al. (1999) provide an overview for different sources of model risk of interest rate risk. Berkowitz and O'Brien (2002) empirically backtest the performance of internal Value-at-Risk models at commercial banks in order to draw conclusions about related estimation risk, whereas Talay and Zhang (2002) use a game theoretic framework for the analysis of model risk. Moreover, several analyses quantify estimation risk through determining confidence intervals for Value-at-Risk estimates under various model assumptions, i.e., the unconditional normal distribution, historical estimates, conditional GARCH models with different innovation assumptions such as normal or Student t errors, the Hill estimator, filtered historical simulation, Gram-Charlier and Cornish-Fisher expansion (see Pritisker, 1997; Christoffersen and Goncalves, 2005;

Chan et al., 2007). Besides this type of model risk quantification, Gourieroux and Zakoyan (2013) and Boucher et al. (2014) examine model risk through adjustment factors which need to be added to the original risk measure estimate in order to derive the desired behavior. The former analyze a GARCH model with normal and Student t innovations and the latter use unconditional extreme value distributions. A further domain in the literature related to model risk deals with the determination of model risk through deriving the bias in Value-at-Risk estimation. Bao and Ullah (2004) examine the bias in Value-at-Risk estimation in case of an ARCH model with normal errors and show that the bias is based on the residuals distribution assumption as well as the estimation error for the model's parameters. Inui and Kijima (2005) show how the bias in Value-at-Risk estimation can be reduced through using convex combinations of the Expected Shortfall and Hartz et al. (2006) reduce the bias for the Value-at-Risk of a normal-GARCH model through a resampling method which leads to better estimation results of the risk measure. Another contribution to the literature related to model risk is given by Escanciano and Olmo (2010, 2011) who analyze the impact of model risk on backtesting methods. Hereby, they provide backtesting methods which explicitly take into account estimation risk.

The literature has also provided capital frameworks for model risk. Kerkhof et al. (2010) provide a general model risk framework specifying estimation, misspecification and identification risk, in order to determine related capital charges and Alexander and Sarabia (2012) propose how a capital buffer for model risk can be calculated in case of quantile risk measures.

In summary, despite the support in literature for EVT methods and the general acknowledgment of model risk by prudential regulators, no paper exists to date that analyzes the effects of model risk for EVT methods and the impact on regulatory capital.

We contribute to previous work in multiple ways. To the best of our knowledge we are the first to analyze second order effects of model risk of EVT methods on corresponding capital requirements. We combine two fields of the literature: EVT and model risk. We analyze the potential impact of model risk using a simulation and an empirical analysis. Our work builds on Kerkhof et al. (2010) and Boucher et al. (2014) as we quantify the amount of misspecification and estimation risk (first order effects of model risk) through the determination of adjustment factors by means of numerical optimization. To the best of our knowledge, we are the first to use conditional EVT methods in this context and additionally examine model risk for the expected shortfall under the application of the peaks over threshold method. Moreover, in contrast to their work, we conduct an isolated analysis of second order effects of model risk which is of special importance for EVT methods.

4

We find a trade-off between accuracy (first order effect) and certainty (second order effect). EVT methods exhibit more accurate point estimates and are less prone to misspecification and estimation risk with respect to our data set. However, at the same time, we detect EVT methods to be more sensitive to second order effects of model risk than traditional methods, which increases the range of possible risk measure estimates and might lead to situations in which risk measure forecasts are further away from the true value than traditional methods. The dispersion further increases with the level of confidence.

The remainder of this paper is structured as follows. Section 2 provides a simulation study to motivate and clarify all sources of model risk. Section 3 contains the theoretical background for the estimation of risk measures, backtesting methods as well as the quantification of model risk, while empirical results are analyzed in Section 4. Section 5 concludes our main findings.

II. DIFFERENTIATION OF MODEL RISK SOURCES IN A SIMULATION STUDY

This section illustrates the two effects of model risk. We assume random losses to follow a standardized Student t distribution with four degrees of freedom for which we know that the true Value-at-Risk with confidence level $\alpha = 0.990$, $VaR_{0.990}$, equals 2.65.

### 1. *First Order Effects of Model Risk*

Consider a risk manager who is not aware of the true DGP and assumes the loss to follow a standard normal distribution. Without any parameter estimation errors this would lead to a point estimate of the $VaR_{0.990}$ equal to 2.33 and hence to an underestimation of approximately 12.1% which could be ascribed to the model risk source of misspecification. Now, if the risk manager correctly assumes a Student t distribution, the corresponding parameters (the mean, standard deviation and shape parameter) need to be estimated. Given a random sample of 1,000 realizations we fit a Student t distribution to the data and get three parameter estimates, e.g., 0.011, 0.976, 4.197 for the mean, standard deviation and shape. This leads to $\hat{VaR}_{0.990} = 2.59$ and hence to an underestimation of approximately 2.26% which can be ascribed to estimation risk. Combining misspecification and estimation risk, we fit a normal distribution to the data. This results in estimates for the mean and standard deviation equal to, e.g., -0.022 and 0.982, which leads to $\hat{VaR}_{0.990} = 2.26$ and an underestimation of 14.72%.[1] In a real-world setting, an isolated consideration of misspecification and estimation risk is not possible as we do not know the DGP and we label the combination of both effects first order effects of model risk.

_____

[1]Effects of misspecification and estimation risk can be commutated or offset each other in different situations.

2. *Second Order Effects of Model Risk*

We compare the estimation process of the standardized Student t distribution with four degrees of freedom to the peaks over threshold method from EVT which exclusively uses data above a certain threshold and directly estimates the tail of a distribution which follows a generalized pareto distribution (GPD) (see McNeil et al., 2005). The GPD depends on two parameters, i.e., a shape and a scaling parameter. The corresponding shape and scaling parameter to the assumed Student t distribution equal 0.25 and 0.50, such that for both methods we know the true parameters and have identical true $VaR_{0.990}$ values. We sample 1,000 random numbers from the true process and estimate both models. Given the estimated models, predictions for the $VaR_{0.990}$ can be made and the deviation from the true value is quantified. We repeat this process 100,000 times which, for both methods, leads to the distribution of deviations to the true value (in %) that can be seen in Figure 1.

Figure 1: Distributions of deviations from $VaR_{0.990}$ predictions to the true value for the Student t distribution and the GPD
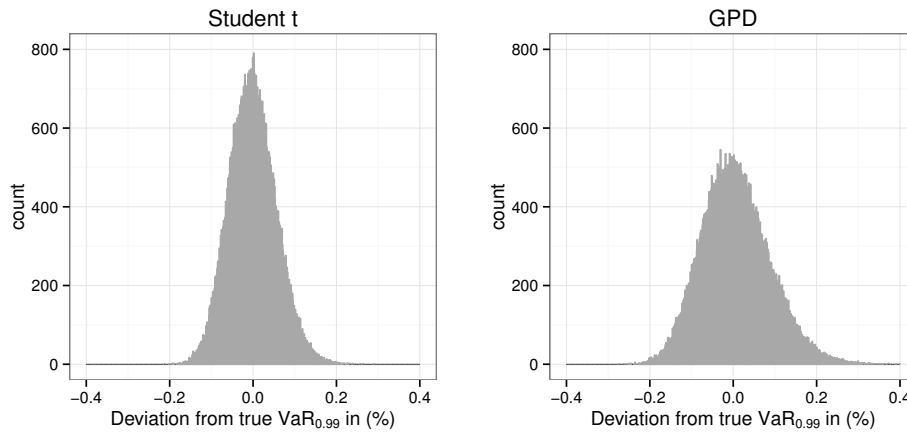
It can be observed that predicted values for the $VaR_{0.990}$ of the GPD exhibit a higher degree of dispersion than for the Student t distribution. This means that more predicted values of the GPD are further away from the true value in comparison to the Student t distribution.[2] Generally speaking, the results in Figure 1 highlight a further source of model risk that is not captured but related to first order effects of model risk, which is why we denote it as second order effects of model risk. To be more precise about the impact of this model risk source, we calculate empirical probabilities for potential underestimation.

---

[2]Note that this is true for over- and underestimation, however, in the following we focus on underestimation as this is more relevant for prudential regulation.

The results for our base case (Student t distribution with four degrees of freedom and random samples of size $1,000$) can be seen in the upper left of Figure 2 and they show that the probability for underestimating the true value of the $VaR_{0.990}$ is higher for the GPD in almost all cases. This seems to be a consequence from second order effects of model risk as through the higher dispersion of the EVT method, more severe underestimations occur with a higher frequency which increases the likelihood for such events. We find that this behavior worsens if the general amount of data values is reduced to 500 (see plot in the upper right of Figure 2) and is reduced for a higher amount of data, e.g., $2,000$ (see plot in the lower left of Figure 2).[3]

---

[3]Generally speaking, smaller time series imply a greater degree of uncertainty and longer time series may not be representative given structural changes.

Figure 2: Probabilities for the underestimation of $VaR_{0.990}$ in case of the Student t distribution and the GPD
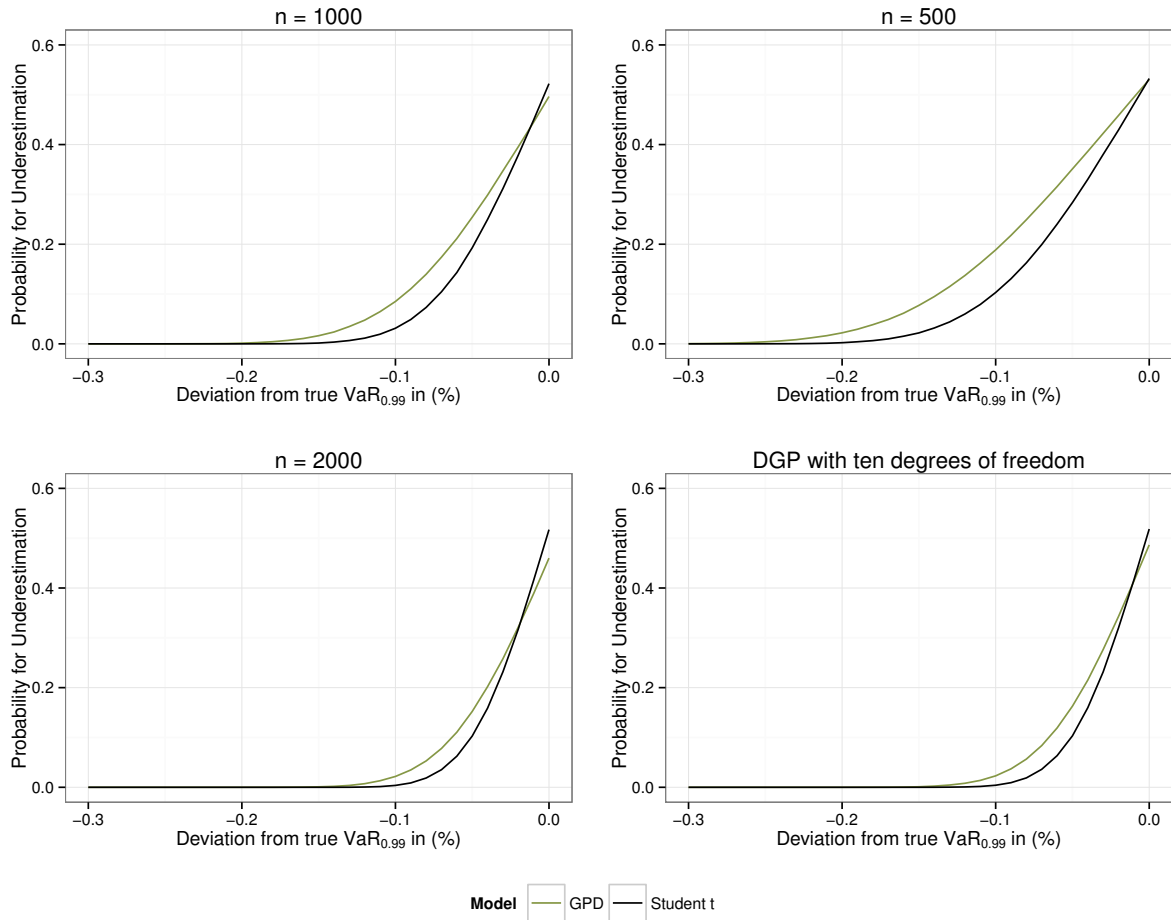
Note: For the results in the upper left we draw $1,000$ numbers from the DGP, while 500 and $2,000$ are drawn for the results in the upper right and the lower left. Results in the lower right are generated using a Student t distribution with ten degrees of freedom as the true DGP and random samples of size $1,000$.

In addition we find these results remain stable, if the DGP is a Student t distribution with ten degrees of freedom which corresponds to a shape parameter of 0.10 for the GPD. This leads to a less heavy tailed DGP. The plot in the lower right of Figure 2 exhibits that the probability for underestimating the true value ($VaR_{0.990} = 2.47$) is still higher for the EVT method. Overall, this section shows that a further dimension of model risk exists besides misspecification and estimation risk which is relevant in the context of risk management (especially for EVT methods).

III. Determination of risk measures and model risk

This section presents the theory to determine risk measures and their related model risk. First, the derivation of Value-at-Risk and Expected Shortfall is described. Second, we show how the accuracy of these models can be analyzed and third, we provide measures to quantify first and second effects of model risk.

1. *Determination of Risk Measures — Theoretical Background*

The aim of our analysis is to determine whether methods from EVT are more prone to prediction risk than traditional methods. We focus on financial data, i.e., negative log-returns of stock indices L_t, $t \in \mathbb{Z}$. L_t is a strictly stationary time series process of the form

$$L_t = \mu_t + \sigma_t X_t, \tag{1}$$

where μ_t and σ_t are measurable with respect to $\mathscr{F}_{t-1}$, the information available up to time $t-1$, and X_t are iid innovations with unknown distribution function F_X (see, e.g., McNeil and Frey, 2000). Moreover, we estimate μ_t by an $ARMA(p_1, q_1)$ process and σ_t by a $GARCH(p_2, q_2)$ process with orders $p_i, q_i, i = 1, 2$, which is estimated using a quasi-maximum likelihood estimation. This means that the parameters for the $ARMA(p_1, q_1) - GARCH(p_2, q_2)$ process are estimated assuming normality with respect to the innovations, even though we do not consider the innovations to be normally distributed. Resulting parameters of the $ARMA(p_1, q_1) - GARCH(p_2, q_2)$ process are consistent and asymptotic normally distributed, if the innovations' distribution has a finite fourth moment (see McNeil et al., 2005). Consequently, the parameters θ for the distribution of X_t are determined separately through maximum likelihood estimation after the $ARMA(p_1, q_1) - GARCH(p_2, q_2)$ process is estimated. For X_t, we assume a (semi-)parametric[4] distribution, i.e., either the generalized extreme value distribution (GEV) or the generalized Pareto distribution (GPD) with respect to methods from EVT or the normal, Student t or logistic distribution with respect to traditional methods.

Hence, given a certain distribution function for X_t and some confidence level $\alpha \in (0, 1)$, estimates for the daily[5] Value-at-Risk $\left(VaR_\alpha^{t,i}\right)$ and Expected Shortfall $\left(ES_\alpha^{t,i}\right)$ of the negative log-returns under the distribution assumption $i \in$ GEV, GPD, normal, Student t, logistic, are derived through (see, e.g., McNeil et al., 2005)

---

[4] The generalized Pareto distribution represents a semi-parametric distribution.

[5] Note that we use daily values as this approach is commonly chosen in practice. Usually, risk measures for periods of V days are approximated using $VaR_\alpha^V = VaR_\alpha^t \cdot \sqrt{V}$.

$$VaR_\alpha^{t,i} = \mu_t + \sigma_t VaR_\alpha^i, \qquad ES_\alpha^{t,i} = \mu_t + \sigma_t ES_\alpha^i, \tag{2}$$

where the unconditional Value-at-Risk and Expected Shortfall are defined through[6]

$$VaR_\alpha^i = inf\{x \in \mathbb{R} : F_X(x) \geq \alpha\}, \qquad ES_\alpha^i = \frac{1}{1-\alpha} \int_\alpha^1 VaR_v^i dv. \tag{3}$$

We focus our analysis to these two risk metrics as the Value-at-Risk is the current standard with respect to regulatory capital requirements and an ongoing discussion exists on the dominance of Value-at-Risk and Expected Shortfall. Note that the Basel Committee on Banking Supervision proposes to use the Expected Shortfall instead of the Value-at-Risk for internal models of market risk (see Basel Committee on Banking Supervision, 2013) and the Expected Shortfall is already applied for the Swiss Solvency Test for insurers. Table 1 provides a summary of the three steps for risk measure estimation.

Table 1: Procedure for estimating risk measures

Step	Procedure
1	Estimate the $ARMA(p_1, q_1) - GARCH(p_2, q_2)$ process
2	Estimate the distribution function for the residuals X_t from Equation 1 using the GEV, GPD, normal, Student t or logistic distribution
3	Determine risk measures according to Equation 2

2. *Assessing the Methods' Accuracy — Backtesting*

We conduct several statistical tests to analyze whether the applied models are suitable for the determination of the Value-at-Risk and Expected Shortfall. A time span of T days is taken into account for backtesting, for which we use a rolling window of D, $D < T$ days to estimate the $\hat{VaR}_\alpha^{t,i}$ and $\hat{ES}_\alpha^{t,i}$, $i \in$ GEV, GPD, normal, Student t, logistic (see Equation 2) at each point in time $t = \{t_D, ..., t_{T-1}\}$. By definition, the probability for the violation of the Value-at-Risk equals $1 - \alpha$. Thus, we introduce an indicator function I_t, which is given through

---

[6]See, e.g., McNeil et al. (2005) for a detailed explanation how to calculate the risk measures under given distributions.

$$I_t = \begin{cases} 1, \text{ if } L_t > V\hat{a}R_\alpha^{t,i}, \\ 0, \text{ if } L_t \leq V\hat{a}R_\alpha^{t,i}, \end{cases}$$

with a hypothesized success rate equal to $1 - \alpha$. This can be tested with a two-sided test under the null hypothesis that the expected success rate (violation rate, respectively) is realized using either a binomial or likelihood ratio test (see, e.g., McNeil and Frey, 2000; Angelidis et al., 2004). If the analyzed method systematically under- or overestimates the 'true' Value-at-Risk, the null hypothesis is rejected for a given significance level. Furthermore, we apply an additional test statistic for a likelihood ratio test, which simultaneously takes into account unconditional coverage and independence and thus, considers effects such as volatility clustering (see Christoffersen, 1998).

Backtesting the Expected Shortfall seems to be less straightforward than the Value-at-Risk. More concrete, the Expected Shortfall is not elicitable, which leads to difficulties with respect to backtesting and robust estimation (see Gneiting, 2011). However, valid methods for testing the Expected Shortfall exist, even though a direct comparison based on these backtesting results should be avoided (see Ziegel, 2014). We follow McNeil and Frey (2000) who present a backtesting method for the Expected Shortfall and introduce

$$R_t = \frac{L_t - ES_\alpha^{t,i}}{\sigma_t}, \tag{4}$$

which, conditional on $L_t > VaR_\alpha^t$, should exhibit an expected value of zero, if the Expected Shortfall is estimated correctly. This is tested using a bootstrap method without making any assumptions about the distribution of R_t. We conduct a two-sided as well as a one-sided test, while the latter, being more conservative, tests against the alternative hypothesis that the Expected Shortfall is systematically underestimated, i.e., the conditional mean is greater than zero. Summing up, we apply five tests, i.e., two conditional tests and one unconditional test for the Value-at-Risk and a two- and a one-sided test with respect to the Expected Shortfall. For each test we reject the respective model with the null hypothesis.

3. *Determination of Model Risk*

As the DGP is usually unknown, two important sources of model risk are (1) the choice of an inaccurate model, i.e., misspecification risk, and (2) the deviation between the estimated and the true parameter, i.e., estimation risk (see Boucher et al., 2014). In a real-world scenario

both sources of risk can not be identified on an isolated basis as they relate to each other and thus, are jointly analyzed.

In general, the realization of the parameter estimates relates to a data subset which represents random observations from the true DGP. This leads to a potential risk source that is not captured by misspecification and estimation risk. In order to quantify this type of risk, we analyze the distributions of possible risk measure estimates, e.g., higher ranges of risk estimates might lead to more severe underestimations of the true risk measure in extreme cases. In the following, we denote this risk as second order effects of model risk.

Our analysis examines first and second order effects of model risk, whereby we expect the latter to be the most relevant for EVT methods. Parameters of the GPD and the GEV exhibit generally higher standard errors due to their use of data during the estimation. For the estimation of the GEV, only maxima of a predefined time period (e.g., a week, a month, etc.) are used, while the GPD is estimated on the basis of values above a threshold such that in both cases a certain amount of the original data is not taken into account for the estimation. Higher standard errors reflect the sensitivity of parameter estimates to changes with respect to the data. The corresponding question is whether this sensitivity and uncertainty is directly reflected in corresponding risk measure estimates, and if so, what consequences should be taken into account from a risk management's perspective.

First Order Effects of Model Risk

To quantify first order effects of model risk, we proceed analogously to Boucher et al. (2014). We determine a linear adjustment factor that is necessary in order to reach the desired behavior for the risk measure's estimate. In case of the Value-at-Risk the failure rate of I_t should equal the predefined confidence level $1 - \alpha$. In case of the Expected Shortfall, the conditional mean of R_t should be equal to zero. Moreover, the adjusted risk measures should not be rejected for the statistical tests described in the previous subsection.[7] We conduct numerical optimization[8] and derive a constant adjustment factor A_j, $j = VaR_\alpha^i$, ES_α^i, $i \in$ GEV, GPD, normal, Student t, logistic, which leads to an adjusted risk measure $\left(\hat{VaR}_{A,\alpha}^{t,i}, \hat{ES}_{A,\alpha}^{t,i} \right)$ at time t

---

[7]The adjustment factor offsets the error in forecasting which stems from misspecification and estimation risk. However it does not offset the potential impact of variations in risk measure forecasts, i.e., second order effects of model risk.

[8]We use differential evolution to derive the adjustment factors (see Storn and Price, 1997). Moreover, to make sure that numerical optimization results are stable, we run the same optimization problem multiple times and find robust results which are available upon request from the authors.

$$VaR^{t,i}_{A,\alpha} = \hat{VaR}^{t,i}_\alpha + A_{VaR^i_\alpha} \cdot \hat{VaR}^{t,i}_\alpha, \qquad \hat{ES}^{t,i}_{A,\alpha} = \hat{ES}^{t,i}_\alpha + A_{ES^i_\alpha} \cdot \hat{ES}^{t,i}_\alpha. \qquad (5)$$

Both adjustment factors are solutions for the optimization problems given through

$$\min_{A_{VaR^i_\alpha}} |\hat{\alpha} - \alpha|, \qquad \min_{A_{ES^i_\alpha}} |E(R_t)|,$$

where $(1 - \hat{\alpha})$ is the empirical failure rate of I_t. Solutions for the minimization are only accepted during the optimization if the values for $|\hat{\alpha} - \alpha|$ and $|E(R_t)|$ stay below a small predefined border that is close to zero.[9] Note that the adjustment factor, as defined in Equation 5, should not be used for a dynamic risk adjustment as it is constant over time. This measure allows us to get an indication which models dominate, as a higher adjustment indicates a lower performance. We can therefore quantify the extent of misspecification and estimation risk by one number. Moreover, this measure could be extended into a dynamic setting through adapting the adjustment factor over time on the basis of a rolling window of a predefined number of past risk measure estimates. The adjustment factor may be interpreted as a capital buffer that is necessary to achieve the adjusted risk measure that should be held by a company in order to guarantee the confidence level in the presence of misspecification and estimation risk.

Second Order Effects of Model Risk

We analyze the range of possible estimates around the point estimate given the empirical data to capture the possible range of risk measure estimates if the true DGP is unknown. We capture second order effects of model risk through drawing random values for parameter estimates. In our analysis, each model is estimated using maximum likelihood, hence, we simulate N multivariate normally distributed random numbers for all parameter estimates by means of the Cholesky decomposition for the estimated covariance matrix $\hat{Cov}(\hat{\theta}) = \hat{\Sigma}$ (see, e.g., McNeil et al., 2005).[10] For each point in time t, we calculate N estimates of the Value-at-Risk and Expected Shortfall to get a distribution for each risk measure at

---

[9]See the empirical analysis for input values.

[10]As focus is laid on the distribution of the innovations, random numbers for their parameters are drawn while parameters for the econometric time series models are kept constant. To analyze potential interaction effects between second order effects of model risk of the econometric model and the innovations distributions additional analyses are conducted in which random numbers are drawn for all the parameters. We find that this increases the general level of model risk but does not impact our general findings.

time t. Based on this distribution we determine an additional quantity, which helps to quantify the degree of second order effects of model risk. This prediction buffer is denoted by $PB_\omega\left(\hat{VaR}_\alpha^{t,i}\right)$, $PB_\omega\left(\hat{ES}_\alpha^{t,i}\right)$, $i \in$ GEV, GPD, normal, Student t and logistic, and for both risk measures, it is determined through

$$PB_\omega\left(\hat{VaR}_\alpha^{t,i}\right) = \hat{VaR}_\alpha^{t,i} - \hat{VaR}_\alpha^{t,i}(\omega), \qquad PB_\omega\left(\hat{ES}_\alpha^{t,i}\right) = \hat{ES}_\alpha^{t,i} - \hat{ES}_\alpha^{t,i}(\omega),$$

with ω being the desired confidence level of second order effects of model risk and $\hat{VaR}_\alpha^{t,i}(\omega)$, $\hat{ES}_\alpha^{t,i}(\omega)$ representing the $\omega-$quantile of the risk measures estimates distribution at time t, such that when adding this quantity, one can be $(1-\omega)$-confident that the adjusted risk capital is sufficient to cover unexpected deviations with respect to changes of the parameter estimates (see Alexander and Sarabia, 2012). The idea behind this quantity is that given a higher degree of dispersion for risk measure estimates the $\omega-$quantile for lower/higher values of ω are further away from the true risk measure which increases prediction buffers. Hence, estimation methods which are more exposed to second order effects of model risk should exhibit higher values for $PB_\omega\left(\hat{VaR}_\alpha^{t,i}\right)$, $PB_\omega\left(\hat{ES}_\alpha^{t,i}\right)$ as their risk measure estimates exhibit a wider range of possible values. For the comparison of these measures in our analysis, we divide them through their corresponding point estimates to make the results comparable.[11]

In summary, adjustment factors $A_{VaR_\alpha^i}$, $A_{ES_\alpha^i}$ should give an implication for the degree of first order effects of model risk, while prediction buffers $\frac{PB_\omega\left(\hat{VaR}_\alpha^{t,i}\right)}{\hat{VaR}_\alpha^{t,i}}$, $\frac{PB_\omega\left(\hat{ES}_\alpha^{t,i}\right)}{\hat{ES}_\alpha^{t,i}}$ take into account the degree of second order effects of model risk.

IV. EMPIRICAL ANALYSIS

1. *Data Description and Input Values*

Our data consists of daily negative log-returns from four indices taken from the Yahoo Finance database, i.e., the S&P 500, Dax 30, Nikkei 225 and BSE Sensex from January 2000 until March 2014.[12] Within this time frame we use a rolling window of $D = 1,000$ days to estimate daily risk measure forecasts according to Equation 2, such that our first estimates refer to the day $D + 1 = 1,001$. The whole time span T is around 3,550 days but differs slightly for each index due to different trading days. Considering the orders of

---

[11] In addition, we took into account the standard deviation and the 95% confidence interval of risk measure estimates at time t. The results are robust and available from the authors upon request.

[12] The four indices are chosen as they represent economies with different characteristics.

the $ARMA(p_1, q_1) - GARCH(p_2, q_2)$ process, we decide to model all time series with an $ARMA(1,1) - GARCH(1,1)$ model as no further autocorrelation between lags of one and ten days for the residuals can be detected through a Box-Pierce test under this assumption.[13]

With respect to methods from EVT, we use a block size of $n = 50$ trading days for the block maxima method and a threshold u which equals the empirical 90th percentile of the data in case of the GPD.[14] We simulate $N = 100,000$ random numbers to derive the distribution for risk measure estimates at each time step and use a significance level of 5% for all statistical tests, i.e., an Augmented Dickey-Fuller test to test for stationarity of the time series before the $ARMA(1,1) - GARCH(1,1)$ model is estimated, a Box-Pierce test to check for autocorrelation in residuals X_t, the conditional tests for the Value-at-Risk violation ratio, the unconditional test for the Value-at-Risk violation ratio and the tests for the mean of R_t as defined in Equation 4.

We analyze different confidence levels $\alpha = (0.975, 0.990, 0.995)$, which we consider to be of special relevance due to regulatory requirements. Firstly, 97.5% is recommended through the Basel Committee on Banking Supervision (2013) if the Expected Shortfall is used instead of the Value-at-Risk. Secondly, 99% should be taken for market risk if the Value-at-Risk is applied (see Basel Committee on Banking Supervision, 2011).[15] Thirdly, the confidence level of 99.5% plays an important role with respect to insurers' regulation in the context of Solvency II. With respect to the numerical optimization of adjustment factors, we accept solutions which are up to 5% away of the expected failure rate $(1 - \alpha)$ in case of the Value-at-Risk and between $[-0.01; 0.01]$ in case of the Expected Shortfall.

2. *First Order Effects of Model Risk*

The expected failure rate equals $1 - \alpha$ by definition and indicates that the chosen risk model systematically underestimates the true risk measure if the estimated failure rate is above the expected failure rate and vice versa (see Section 3). Results for the estimated mean of R_t can be interpreted in a similar fashion. Table 2 displays estimation results and adjustment factors for the S&P 500 and Nikkei 225 for the Value-at-Risk and Expected Shortfall. Two rows are given for each method. The first row displays empirical failure rates for Value-at-Risk estimates and estimated mean values of R_t given confidence levels of $\alpha = (0.975, 0.990, 0.995)$. Corresponding outcomes for the tests related to backtesting are highlighted with symbols

---

[13]Details are available from the authors upon request.

[14]We conducted our analyses with different threshold levels and block sizes and detect the general findings to be robust for varying values.

[15]This confidence level refers to a time horizon of ten days. Our results are calculated on a daily basis. Risk measures may be adjusted for a longer time horizon as explained in Section 3.

as described below in Table 2. The second row presents adjustment factors as defined in Equation 5 in parentheses, while the smallest absolute values for adjustment factors are additionally highlighted in bold. Adjustment factors can be interpreted as the relative deviation from adequate to estimated capital requirements while models which exhibit the smallest adjustment factors are less prone to first order effects of model risk, as it needs by definition less adjustment to achieve the desired behavior of the risk measure. For example, in case of the S&P 500 for a confidence level of 0.975, the empirical failure rate for the GEV equals 0.02945 and thus, underestimates the true risk as the expected failure rate is given through 0.025. Hence, from an ex-post perspective 4.041 per cent of the Value-at-Risk estimates need to be added at each point in time to derive the expected failure rate. Results with respect to the Dax 30 and BSE Sensex index are illustrated in Table 7 analogously to Table 2.[16]

The true risk seems to be underestimated for the majority of the models. With respect to the Value-at-Risk in only five out of 60 cases the estimated failure rate is below the expected failure rate, while in only four out of 48 cases the estimated mean for R_t is below zero with respect to estimating the Expected Shortfall. Among nine instances in which risk is overestimated, results are generated using the GEV, GPD, logistic and Student t distribution. Consistent with previous studies, this underlines the inability of the normal distribution to model extreme events.

---

[16]We also analyzed an equally weighted portfolio of the four indices. We spare the illustration of these results as they behave similarly and do not provide any new insights.

Table 2: Failure rates $\sum_{t=D}^{(T-1)} I_t/(T-1-D)$ and estimates for the expected mean of R_t for given confidence levels α, backtesting results and adjustment factors $A_{VaR_\alpha^i}$, $A_{ES_\alpha^i}$

	Value-at-Risk					
	S&P 500			**Nikkei 225**		
$1-\alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
GEV	$0.02945^{*\dagger\circ}$	0.01627	$0.00775^{*\dagger\circ}$	$0.03001^{*\dagger\circ}$	$0.01120^{*\dagger\circ}$	$0.00480^{*\dagger\circ}$
	(0.04041)	(0.07166)	(0.04448)	(0.03992)	(0.02722)	(-0.00291)
GPD	$0.03100^{*\dagger\circ}$	$0.01201^{*\dagger\circ}$	$0.00620^{*\dagger\circ}$	$0.02641^{*\dagger\circ}$	$0.00920^{*\dagger\circ}$	$0.00520^{*\dagger\circ}$
	(0.04495)	**(0.03621)**	(**0.02382**)	**(0.01682)**	(-0.03056)	**(0.00283)**
Normal	0.04029	0.02363	0.01472	0.03441	0.01521	0.00920
	(0.16166)	(0.21291)	(0.20850)	(0.07464)	(0.09784)	(0.17626)
Student t	0.04029	0.01782	0.00814	0.03321	$0.01281^{*\dagger\circ}$	$0.00760^{*\dagger\circ}$
	(0.16774)	(0.15977)	(0.09140)	(0.07280)	(0.05726)	(0.12459)
Logistic	0.03758	0.01744	0.00852	$0.03081^{*\dagger\circ}$	$0.01000^{*\dagger\circ}$	$0.00560^{*\dagger\circ}$
	(0.15916)	(0.14718)	(0.10293)	(0.04186)	**(-0.00176)**	(0.02730)
	Expected Shortfall					
	S&P 500			**Nikkei 225**		
$1-\alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
GPD	$-0.00288^{\bullet+}$	$0.04922^{\bullet+}$	$0.07428^{\bullet+}$	$0.01127^{\bullet+}$	$0.23833^{\bullet+}$	$0.38136^{\bullet+}$
	(-0.00085)	(0.01579)	(0.02151)	(0.00424)	**(0.07492)**	**(0.10782)**
Normal	0.23636	$0.22503^{\bullet}$	$0.26296^{\bullet}$	0.19752	0.38715	0.54893
	(0.10311)	(0.08625)	(0.09252)	(0.08504)	(0.10617)	(0.15105)
Student t	$0.00210^{\bullet+}$	$-0.06461^{\bullet+}$	$-0.05165^{\bullet+}$	$0.09666^{\bullet+}$	$0.28372^{\bullet}$	$0.38585^{\bullet}$
	(0.00083)	(-0.02115)	(-0.01485)	(0.03956)	(0.08166)	(0.12046)
Logistic	$0.08190^{\bullet+}$	$0.01924^{\bullet+}$	$0.03380^{\bullet+}$	$0.00640^{\bullet+}$	$0.25591^{\bullet}$	$0.38236^{\bullet}$
	(0.03293)	**(0.00643)**	**(0.00910)**	**(0.00248)**	(0.07575)	(0.10992)

Note: Failure rates and expected means are given in the first row, the second row shows adjustment factors which are displayed in parentheses where the lowest adjustment factors are highlighted in bold. With respect to backtesting, symbols are defined as $*: H_0$ for binomial-test is not rejected; $\dagger: H_0$ for unconditional Likelihood ratio test is not rejected; $\circ: H_0$ for conditional Likelihood ratio test is not rejected.; $\bullet: H_0$ for mean is equal to zero is not rejected; $+: H_0$ for mean is smaller than zero is not rejected.

With respect to backtesting the Value-at-Risk, the GPD and GEV seem to be well suited for all confidence levels as except for the S&P 500 and $\alpha = 0.990$, none of the conducted backtests are rejected. At the same time almost all tests are rejected for the normal, Student t and logistic distribution in case of the S&P 500 and Dax 30, while the Student t and logistic distribution perform well in most of the cases for the Nikkei 225 and BSE Sensex. With respect to the backtesting of the Expected Shortfall, almost all statistical tests are rejected for the normal distribution, which leads to the conclusion that the mean for R_t differs from zero in nine out of twelve cases and is systematically underestimated (the mean of R_t is greater than zero) in twelve out of twelve cases, respectively.

Otherwise, none of the tests under the null hypotheses that the mean of R_t is equal to zero is rejected for the GPD, Student t and logistic distribution, while only in one (four, three) out of twelve cases the null that the mean of R_t is smaller than zero is rejected in case of the GPD (Student t, logistic) distribution. The adjustment factors are given in parentheses for each method. They should provide an indication for the amount of first order effects of model risk. Adjustment factors are higher the further away the failure rates are from the reference confidence level. The adjustment factors are therefore consistent with backtesting results. The GEV and GPD generate the lowest adjustment factors in all cases except one with regard to the Value-at-Risk, while for the Expected Shortfall, the logistic distribution seems to be more accurate than the other distributions as it leads to the lowest adjustment factors in seven out of twelve cases. At the same time, the GPD exhibits the lowest adjustment factor in four cases and the Student t distribution in one case.

The performance deterioration of the peaks over threshold method in case of the Expected Shortfall is related to the fact that the Expected Shortfall considers all values in the tail above the corresponding Value-at-Risk. Especially in case of the S&P 500 and Dax 30, the GPD seems to capture the Expected Shortfall less well than the Student t and logistic distribution. Our results show that for both indices the GPD's shape parameter ξ is negative at many points in time, which leads to a Pareto type II distribution.[17] Figure 3 exemplary illustrates the density distribution for a Pareto and a Pareto type II distribution. As it can be seen the Pareto type II distribution exhibits less probability mass in the tail of the distribution. As a consequence, even if the quantile, i.e., the Value-at-Risk, is accurately estimated, the risk situation for values higher than the Value-at-Risk tends to be underestimated such that the mean for R_t is underestimated. This seems to be true when estimating the Expected Shortfall for the S&P 500 and Dax. In the case of the Nikkei 225 and BSE Sensex ξ is positive at more points in time compared to the remaining indices, which leads

---

[17]Depending on the shape parameter $\xi > 0$ ($\xi = 0$, $\xi < 0$), the GPD is a Pareto (exponential, Pareto type II) distribution. See McNeil et al. (2005) for more detailed information.

to a Pareto distribution that exhibits more mass in the tail and thus captures events above the Value-at-Risk more accurately. This is reflected in the better performance of the GPD when estimating the Expected Shortfall for both indices.

As the results presented in Table 2 refer to the whole sample period, we additionally analyze the performance of each model if we split the sample period into three sub-periods. We conduct this additional analysis in order to examine how the models perform during times of turmoil. We create three sub-periods, i.e., the first estimation period lies between 2004 and June 2007, the second period between July 2007 and March 2010 and the last period between April 2010 until the end of our data sample.[18] Table 3 presents estimated failure rates for the Value-at-Risk and estimates of R_t for the Expected Shortfall given each subsample, confidence level and method in the case of the S&P 500.[19]

Figure 3: Comparison for the density of the GPD with a positive and a negative shape parameter

<hr>
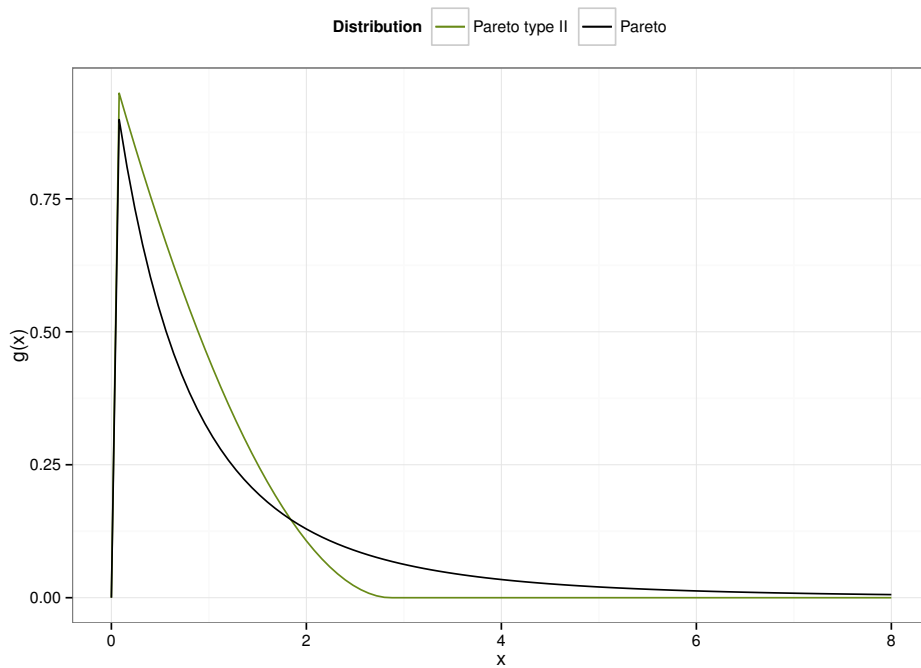
[18]Sub-periods refer to risk measure estimation, i.e., the 1,000 last observations are used to determine estimates for risk measures.

[19]Results for the remaining indices are robust and available from the authors upon request.

Table 3: Failure rates $\sum_{t=D}^{(T-1)} I_t/(T-1-D)$, estimates for the expected mean of R_t and adjustment factors $A_{VaR_\alpha^i}$, $A_{ES_\alpha^i}$ for the S&P 500 when dividing the time frame into three subsamples

|	Value-at-Risk								
	$1-\alpha=0.025$			**$1-\alpha=0.010$**			**$1-\alpha=0.005$**		
	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014
GEV	0.02043	0.06364	0.02385	0.01249	0.04091	0.01033	0.00681	0.02045	0.00397
	(-0.01675)	(0.39485)	**(-0.00708)**	(0.01818)	(0.48067)	**(0.00279)**	(0.01284)	(0.29449)	(-0.01984)
GPD	0.02724	0.05909	0.02385	0.01135	0.02955	0.00636	0.00568	0.01818	0.00238
	(0.01638)	**(0.34236)**	(-0.00861)	**(0.01638)**	**(0.26893)**	(-0.06150)	(0.00482)	**(0.22741)**	(-0.04651)
Normal	0.02724	0.07500	0.03736	0.01362	0.04773	0.02226	0.00681	0.03864	0.01192
	(0.01657)	(0.46475)	(0.16872)	(0.04303)	(0.58643)	(0.17005)	(0.01915)	(0.52194)	(0.18302)
Student t	0.02611	0.07727	0.03736	0.01249	0.04318	0.01272	0.00568	0.02273	0.00477
	(0.00304)	(0.48864)	(0.16475)	(0.01892)	(0.51513)	(0.07220)	**(0.00412)**	(0.34726)	**(-0.01602)**
Logistic	0.01930	0.07500	0.03736	0.00681	0.04318	0.01590	0.00341	0.02273	0.00715
	(-0.02858)	(0.46719)	(0.16719)	(-0.07716)	(0.52465)	(0.08542)	(-0.03139)	(0.35119)	(0.07293)
	Expected Shortfall								
	$1-\alpha=0.025$			**$1-\alpha=0.010$**			**$1-\alpha=0.005$**		
	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014
GPD	0.12993	0.04695	-0.15073	0.32082	-0.15811	0.04662	0.79916	-0.43611	0.22720
	(0.04974)	**(0.01702)**	(-0.05323)	(0.11674)	(-0.04719)	**(0.01458)**	(0.28011)	(-0.11289)	(0.06759)
Normal	0.20410	0.32866	0.18803	0.34661	0.31013	0.10910	0.70178	0.19835	0.16067
	(0.08877)	(0.14250)	(0.08251)	(0.13306)	(0.11819)	(0.04196)	(0.24983)	(0.06940)	(0.05663)
Student t	0.15212	0.09163	-0.13607	0.25663	-0.06346	-0.28683	0.66346	-0.17771	-0.43747
	(0.06396)	(0.03638)	(-0.05227)	**(0.09329)**	(-0.02070)	(-0.08822)	**(0.21724)**	(-0.05079)	(-0.11533)
Logistic	0.14011	0.13884	0.02086	0.45212	-0.00418	-0.08836	0.92608	-0.08199	-0.13498
	(0.05425)	(0.05562)	**(0.00853)**	(0.14753)	**(-0.00139)**	(-0.02992)	(0.36810)	**(-0.02618)**	**(-0.04054)**

The results given in Table 3 show that the estimation performance during the financial crisis is substantially worse than in the periods before and after the crisis for all models. This can be seen through high deviations from the expected behavior of all models and shows that no model seems to be capable of correctly forecasting the impact of the financial crisis. In the period before the financial crisis, we observe no extreme events such that even the normal distribution provides reasonable estimates for risk measures. The GPD deviates the least from the expected behavior during the financial crisis in the majority of all cases. Even though its risk measure estimates are too small during this period, it would have caused the lowest degree of financial distress if such a model would have been used. In the period after the financial crisis we find similar results as for the consideration over the whole period between 2004 and 2014. With respect to Value-at-Risk estimation, EVT methods exhibit the lowest adjustment factors and the logistic distribution seems to be most accurate with respect to Expected Shortfall estimation. We conclude this part of the analysis that in comparison to traditional methods, EVT methods seem to exhibit lower degrees of first order effects of model risk.

3. *Second Order Effects of Model Risk*

The degree of second effects of model risk is quantified through prediction buffers. The conjecture that EVT methods might be more prone to second order effects of model risk is based on the fact that these methods only use a fraction of the whole data set and thus, usually exhibit higher standard errors for their parameter estimates. First, we examine whether this behavior leads to more dispersed risk estimates that are quantified by EVT methods. In a second step, we analyze how second order effects of model risk impact the safety level for risk management. Figure 1 shows that a higher degree of second order effects of model risk might not necessarily lead to risk measure estimates that are not sufficient to cover critical losses. This issue will be addressed at the end of this analysis.

A risk measure distribution is generated for each method at each time step through Monte-Carlo simulation. Prediction buffers (see Section 3) are calculated using these distributions. The results are given in Figure 4 for the S&P 500 and the Nikkei 225, respectively. Results with respect to the Value-at-Risk and a confidence level of $\alpha = 0.990$ are illustrated in the two upper plots, while the two lower plots correspond to results for the Expected Shortfall and a confidence level of $\alpha = 0.975$. Overall, results are comparable for each index and we focus on the interpretation of the S&P 500 and the Nikkei 225. It can be seen that methods from EVT (GEV and GPD) exhibit higher prediction buffers (in relation to their point estimate).

Figure 4: Prediction buffers in case of the S&P 500 and Nikkei 225 for the Value-at-Risk with a confidence level $\alpha = 0.990$ and Expected Shortfall given a confidence level $\alpha = 0.975$ and $\omega = 0.05$
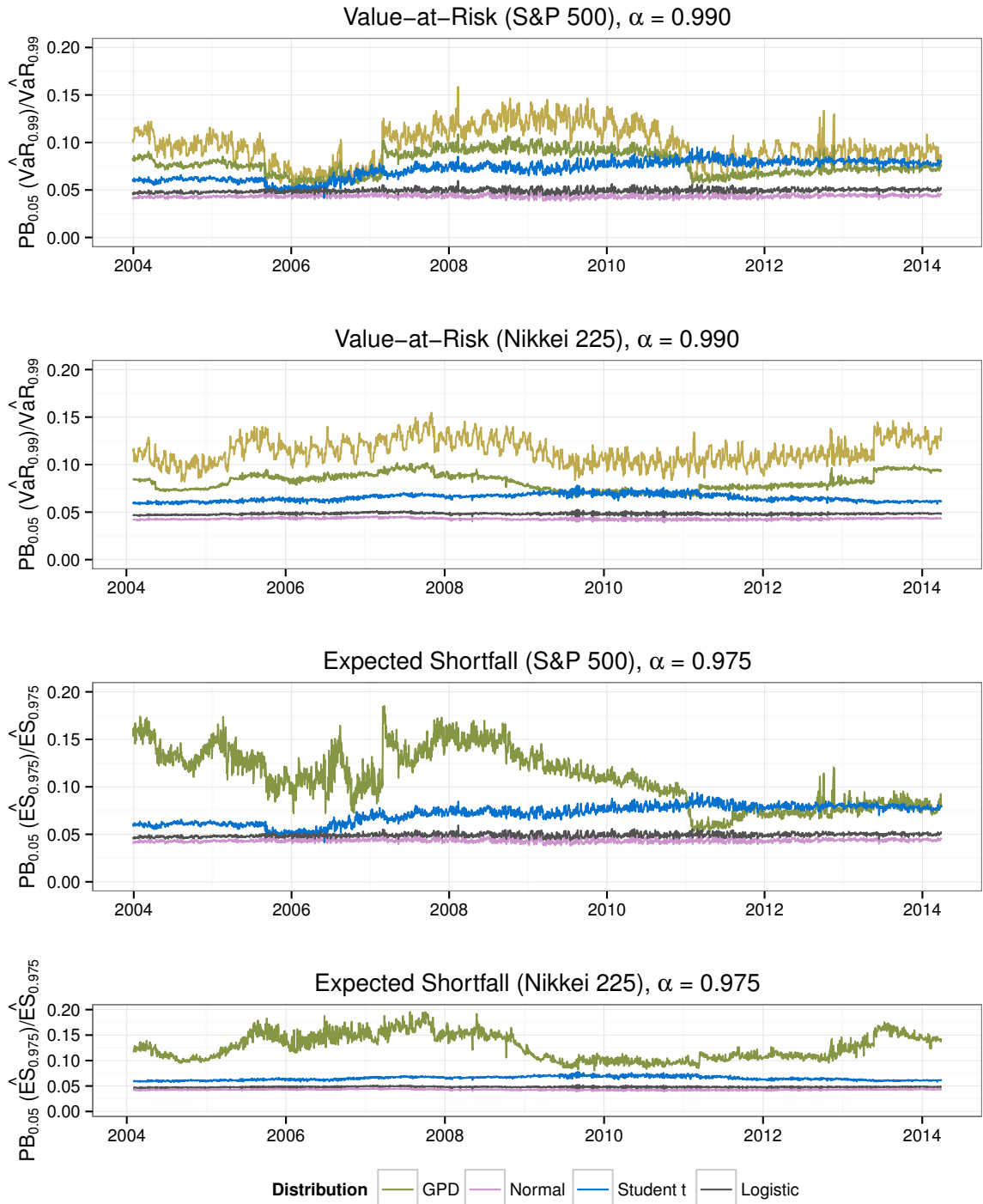
Note: For each method, the difference between the point estimate and the ω-quantile of the risk measure distribution at time t is illustrated in relation to the point estimate. Here ω equals 0.05. For all plots the order from top to bottom is: GEV, GPD, Student t, Logistic, Normal.

Starting with the results for the Value-at-Risk, the highest degree of prediction buffers over most of the sample period is given for the GEV, while the GPD follows most points in time. The traditional distributions show smaller values for prediction buffers, except for the Student t distribution which exceeds the prediction buffers of the GPD in some cases. We find that standard errors of the distributions' parameter estimates in relation to the distributions' parameter estimates are higher for the case of the GEV and GPD as compared to the normal, Student t and logistic distribution. This might explain the higher level of prediction buffers in the GEV and GPD cases. As the block maxima methods uses the least amount of data it tends to have the highest standard errors, which is translated to the variability of the Value-at-Risk estimates. The explanation for the higher level of prediction buffers in the case of the Student t distribution differs.

While the parameters' standard errors are low compared to the extreme value distributions, the point estimates for the Value-at-Risk are smaller than in case of the GEV and GPD (see results in Table 2 which indicate small Value-at-Risk estimates as the estimated failure rates are relatively high). Thus, the relation of prediction buffers to the point estimates is higher due to smaller values in the denominator.

Similar results can be found for the Expected Shortfall, such that the GPD leads to the highest prediction buffers over almost the whole time span. In addition, comparing the results of the GPD for the Value-at-Risk and the Expected Shortfall, we find higher prediction buffers for the Exptected Shortfall. The development of prediction buffers through time indicates that the degree of second order effects of model risk for the GPD seems to vary through time, such that its behavior might be influenced by changes in the economy, e.g., during the financial crisis, prediction buffers are relatively high. For all indices and both risk measures our results indicate that the sensitivity towards second order effects of model risk increases for higher confidence intervals. Figure 5 illustrates prediction buffers for increasing confidence levels at different times. Again it needs to be highlighted that prediction buffers relate to point estimates which allows us to better explain the results given in Figure 5.[20] We find that for the EVT distributions the level of prediction buffers increases above-average for higher confidence levels and remains constant for the normal, Student t and logistic distribution.

---

[20]Values for risk measures increase for each method. It is possible that the relation between prediction buffers and the level of the point estimate decreases for higher confidence levels as a relative consideration between prediction buffers of risk measure estimates and point estimates is taken into account.

Figure 5: Sensitivity of prediction buffers against increasing confidence levels in case of the S&P 500 for the Value-at-Risk and Expected Shortfall given different dates
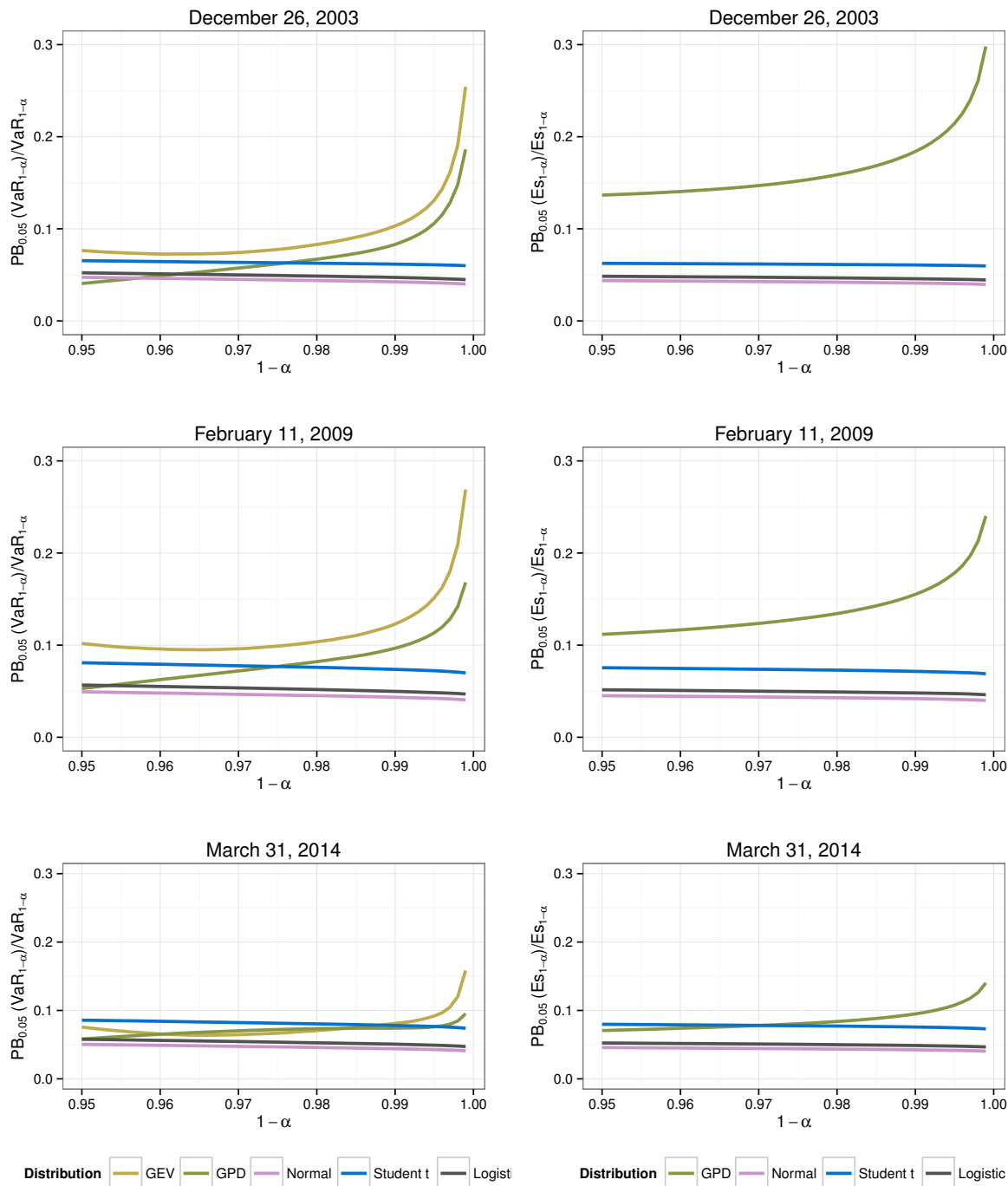
The results show that the absolute level of second order effects of model risk, which is measured through prediction buffers for each method, as well as point estimates for risk

measures increases for higher confidence levels. However the increase of prediction buffers and point estimates is proportionally linear in the case of the traditional distributions. In contrast, we find that the growth of prediction buffers increases over-proportionally to the growth in point estimates for the GEV and the GPD. This is an important result, as EVT distributions exhibit higher prediction buffers and in addition seem to be more sensitive with respect to an increase in the level of safety, i.e., the confidence level. These results seem to be very relevant for prudential regulation as the choice of model methodology may determine required capital buffers.

In summary, EVT methods are capable of estimating the Value-at-Risk and Expected Shortfall with respect to backtesting methods in our analysis. Moreover, the degree of first order effects of model risk is comparably small. However, due to their estimation methods (which use only a subset of the whole data set) they exhibit higher standard errors for parameter estimates, which leads to more dispersed risk measure estimates and thus, to higher prediction buffers. The remaining question is whether this is critical from a risk management perspective and how such a risk source could be handled within a risk management model.

4. *Implications of Second Order Effects of Model Risk for Risk Management*

Risk measures may be estimated well below the mean of the estimates' distribution as a consequence of the large parameter uncertainty. This might lead to a low amount of safety capital that is not sufficient to protect the company against future losses. Hence, we conduct an additional analysis for which we consider all sources of model risk analyzed in this paper at the same time. To achieve this goal we perform the same analysis as for adjustment factors illustrated in Table 2 and 7 and use the lower ω-quantile of risk measure estimates instead of point estimates. Thus, an additional safety buffer is included which takes a possible negative consequence of second order effects of model risk into account. Moreover, it is possible through this procedure to isolate the extent of second order effects of model risk and to examine whether risk measure estimates derived through EVT methods may exhibit a higher potential downside with respect to the determination of capital requirements.

Results (over the whole sample period) for adjustment factors using the $\omega = 0.025, 0.050, 0.100$-quantile are displayed in Table 4 with respect to the Value-at-Risk estimation and in Table 5 with respect to the Expected Shortfall estimation for the S&P 500 and Nikkei 225.[21] For each method adjustment factors decrease the higher ω and they are all higher in comparison to those derived with the point estimates.

---

[21]See Table 8 and 9 for the Dax 30 and BSE Sensex index.

Table 4: Adjustment factors when using the $\omega = 0.025,\ 0.050,\ 0.100$ quantile of the risk measure estimate distributions in case of the Value-at-Risk

	$1-\alpha$	S&P 500			Nikkei 225		
		0.025	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
$\hat{VaR}_\alpha^{t,GEV}(0.025)$		0.20748	0.21098	0.22477	0.18894	0.20134	0.21714
$\hat{VaR}_\alpha^{t,GPD}(0.025)$		**0.14679**	**0.13968**	0.15872	**0.10207**	0.09161	0.14552
$\hat{VaR}_\alpha^{t,Normal}(0.025)$		0.23189	0.27994	0.27243	0.13697	0.15830	0.23787
$\hat{VaR}_\alpha^{t,Student\ t}(0.025)$		0.27694	0.23445	0.18741	0.16477	0.14967	0.21335
$\hat{VaR}_\alpha^{t,Logistic}(0.025)$		0.24305	0.21790	**0.14589**	0.10564	**0.05976**	**0.10246**
$\hat{VaR}_\alpha^{t,GEV}(0.050)$		0.18300	0.18330	0.19634	0.16136	0.16043	0.17494
$\hat{VaR}_\alpha^{t,GPD}(0.050)$		**0.12564**	**0.12004**	0.13857	**0.08456**	0.07463	0.12715
$\hat{VaR}_\alpha^{t,Normal}(0.050)$		0.21995	0.26582	0.26219	0.12653	0.14799	0.23630
$\hat{VaR}_\alpha^{t,Student\ t}(0.050)$		0.25785	0.21330	0.16933	0.14978	0.12504	0.18780
$\hat{VaR}_\alpha^{t,Logistic}(0.050)$		0.23101	0.20579	**0.12859**	0.09647	**0.05485**	**0.07889**
$\hat{VaR}_\alpha^{t,GEV}(0.100)$		0.15359	0.15666	0.15928	0.12288	0.13999	0.14225
$\hat{VaR}_\alpha^{t,GPD}(0.100)$		**0.1076**	**0.10501**	**0.11375**	**0.06961**	0.05107	0.10546
$\hat{VaR}_\alpha^{t,Normal}(0.100)$		0.20575	0.25136	0.24973	0.11356	0.13641	0.21935
$\hat{VaR}_\alpha^{t,Student\ t}(0.100)$		0.23586	0.19089	0.15019	0.13288	0.10980	0.16838
$\hat{VaR}_\alpha^{t,Logistic}(0.100)$		0.21571	0.19213	0.12461	0.08380	**0.03738**	**0.06306**

Note: Adjustment factors for the estimation of the Value-at-Risk are displayed in this table similar to Table 2. Lower quantiles from the distributions of risk measure estimates are used instead of point estimates in order to analyze the impact of prediction risk. The smallest adjustment factors, given an index and a confidence interval, are highlighted in bold.

The degree of second order effects of model risk can be quantified on an isolated basis when examining absolute deviations between the adjustment factors using point estimates and adjustment factors using the ω-quantile of the risk measure distributions. For example, with respect to the Value-at-Risk estimation of the S&P 500 with a confidence level of 97.5% through the GPD results in an adjustment factor of 0.04495 when using the point estimate (see Table 2) and 0.14679 when using the 2.5% quantile of the Value-at-Risk distribution for the GPD. Hence, the impact of second order effects of model risk can be quantified through an additional adjustment equal to $0.14679 - 0.04495 = 0.10184$.

The consideration of second order effects of model risk might alter the decision for the best method. Given the results for Value-at-Risk estimation in Table 2 and 7, EVT methods exhibit the lowest adjustment factors in eleven out of twelve cases. For $\omega = 0.025\ (0.050,\ 0.100)$

this changes such that EVT methods have the lowest adjustment factors in six (seven, nine) cases out of twelve. This shows that there is a potential downside risk, which could result in using over-optimistic estimates for the true risk measure. This also holds true for the estimation of the Expected Shortfall. Comparing Table 2 and 7 with 5 and 9 shows that with respect to EVT methods, the number of lowest adjustment factors decreases from four to zero for all values of ω taken into account in the analysis.

Table 5: Adjustment factors when using the $\omega = 0.025,\ 0.050,\ 0.100$ quantile of the risk measure estimate distributions in case of the Expected Shortfall

	S&P 500			Nikkei 225		
$1 - \alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
$\hat{VaR}_\alpha^{t,GPD}(0.025)$	0.10370	0.11913	0.15783	0.09638	0.15499	0.22796
$\hat{VaR}_\alpha^{t,Normal}(0.025)$	0.12259	0.11955	0.11002	0.10863	0.13183	0.17549
$\hat{VaR}_\alpha^{t,Student\ t}(0.025)$	0.04376	**-0.00468**	**-0.00316**	0.05952	0.10888	0.16042
$\hat{VaR}_\alpha^{t,Logistic}(0.025)$	**0.04216**	0.01425	0.01078	**0.02648**	**0.08213**	**0.12132**
$\hat{VaR}_\alpha^{t,GPD}(0.050)$	0.09321	0.09253	0.14606	0.07895	0.14200	0.18763
$\hat{VaR}_\alpha^{t,Normal}(0.050)$	0.12014	0.10298	0.10079	0.10762	0.12156	0.16577
$\hat{VaR}_\alpha^{t,Student\ t}(0.050)$	**0.03019**	**-0.00533**	**-0.00424**	0.05742	0.08579	0.15124
$\hat{VaR}_\alpha^{t,Logistic}(0.050)$	0.03610	0.01401	0.01043	**0.01650**	**0.08010**	**0.11078**
$\hat{VaR}_\alpha^{t,GPD}(0.100)$	0.07722	0.06548	0.12184	0.06170	0.13890	0.17112
$\hat{VaR}_\alpha^{t,Normal}(0.100)$	0.11877	0.09558	0.09530	0.10234	0.11783	0.16408
$\hat{VaR}_\alpha^{t,Student\ t}(0.100)$	**0.02589**	-0.01179	**-0.00434**	0.05373	0.08322	0.15092
$\hat{VaR}_\alpha^{t,Logistic}(0.100)$	0.03577	**0.01064**	0.01011	**0.01049**	**0.07684**	**0.11003**

Note: Adjustment factors for the estimation of the Expected Shortfall are displayed in this table similar to Table 2. Lower quantiles from the distributions of risk measure estimates are used instead of point estimates in order to analyze the impact of prediction risk. The smallest adjustment factors, given an index and a confidence interval, are highlighted in bold.

For the Expected Shortfall it can be observed that for $\omega = 0.025$ (0.050, 0.100) adjustment factors are higher than those necessary for the normal distribution in seven (seven, five) out of twelve cases. However, in some cases EVT methods still perform better than traditional methods in the presence of second order effects of model risk, especially in the case of the Value-at-Risk estimation.

Table 6: Adjustment factors $A_{VaR_\alpha^i}$, $A_{ES_\alpha^i}$ according to Table 4 for the S&P 500 when dividing the time frame into three subsamples and using the $\omega = 0.025, 0.010$ quantile of the risk estimate distribution

	Value-at-Risk								
	$1-\alpha = 0.025$			$1-\alpha = 0.010$			$1-\alpha = 0.005$		
	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014
GEV ($\omega = 0.025$)	0.07726	0.57199	0.12981	0.14951	0.67802	0.13561	0.10224	0.84064	0.09238
GEV ($\omega = 0.100$)	0.04262	0.50916	0.06736	0.10494	0.61849	0.09347	0.07222	0.79091	0.05554
GPD ($\omega = 0.025$)	0.09702	**0.47254**	**0.09165**	0.13231	**0.55208**	**0.04982**	0.07081	**0.72407**	**0.06304**
GPD ($\omega = 0.100$)	0.06854	**0.42532**	**0.05585**	0.09017	**0.51039**	**0.01135**	0.04210	**0.66482**	**0.01543**
Normal ($\omega = 0.025$)	0.07331	0.54210	0.24025	0.10019	0.71078	0.21094	0.05331	0.88903	0.25820
Normal ($\omega = 0.100$)	0.05548	0.50779	0.20949	0.07778	0.68247	0.20544	0.03942	0.85088	0.19035
Student t ($\omega = 0.025$)	0.08267	0.59764	0.27603	0.10593	0.64731	0.18925	0.03029	0.81033	0.06558
Student t ($\omega = 0.100$)	0.05204	0.53117	0.23274	**0.06611**	0.60822	0.13361	0.02412	0.77402	0.05794
Logistic ($\omega = 0.025$)	**0.03196**	0.53444	0.25433	**-0.00585**	0.58271	0.16450	**-0.01039**	0.76022	0.14535
Logistic ($\omega = 0.100$)	**0.01084**	0.51118	0.22269	-0.06826	0.54709	0.13778	**-0.02183**	0.74289	0.11069
	Expected Shortfall								
	$1-\alpha = 0.025$			$1-\alpha = 0.010$			$1-\alpha = 0.005$		
	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014	2004 - 2007	2007 - 2010	2010 - 2014
GPD ($\omega = 0.025$)	0.19420	0.09334	0.05976	0.26611	0.06129	0.08638	0.33111	0.03512	0.13827
GPD ($\omega = 0.100$)	0.15038	**0.06220**	0.04422	0.19187	0.02453	0.05638	0.30415	0.02368	0.12279
Normal ($\omega = 0.025$)	0.10766	0.16809	0.09470	0.14871	0.15081	0.09712	0.27916	0.14236	0.09365
Normal ($\omega = 0.100$)	0.09636	0.15525	0.08449	0.13880	0.13246	0.08056	0.26642	0.12484	0.08476
Student t ($\omega = 0.025$)	0.08227	0.09101	**-0.00511**	**0.12096**	**0.03680**	-0.05083	**0.25198**	0.05657	-0.07898
Student t ($\omega = 0.100$)	0.07374	0.07278	-0.02674	**0.11858**	**0.01622**	-0.06112	**0.21975**	0.04354	-0.10575
Logistic ($\omega = 0.025$)	**0.06951**	**0.08778**	0.02352	0.17214	0.03881	**0.00333**	0.42763	**0.02439**	**-0.00368**
Logistic ($\omega = 0.100$)	**0.06121**	0.07285	**0.01952**	0.15221	0.02728	**-0.00942**	0.40279	**0.02039**	**-0.01402**

This is in line with the observation of Figure 4 which illustrates that estimates for the Expected Shortfall are more prone to prediction risk. These findings can also be observed for the consideration of three periods. We exemplary show adjustment factors for the S&P 500 when using the $\omega = 0.025, 0.100$ quantile of the risk measure estimates distribution in Table 6.[22] Overall, our results emphasize the potential impact of second order effects of model risk and how original risk measure estimates could be adjusted, in order to improve a risk model and protect it against the potential impact of first and second order effects of model risk.

V. CONCLUSION

This paper examines model risk for EVT methods. In the recent literature, EVT methods are often applied to estimate tail risk measures and compared to traditional methods and good backtesting results are detected in most of these analyses. An attribute associated with methods from EVT is that they produce higher standard errors for parameter estimates as less data is used compared to traditional methods. The question of interest is whether this increases model risk with respect to risk measure estimates of these methods and if so, how this behavior should be handled with respect to the quantification of the risk situation and capital adequacy.

We apply two EVT and three traditional methods to estimate the Value-at-Risk and Expected Shortfall and examine different sources of model risk in this context. Our empirical results show that EVT methods are less prone to misspecification and estimation risk (first order effects of model risk), but exhibit higher degrees of second order effects of model risk in most of the cases. Furthermore, the sensitivity towards second order effects of model risk increases for higher safety levels.

The economic impact of these results has to be evaluated in the context of the respective regulatory regime. For example, according to the Basel regulation, regulatory capital for market risk is generally calculated as the sum of a specific risk charge plus the maximum of the previous day's 99%-VaR and the average 99%-VaR multiplied by a punitive factor that is generally equal to three and increases from three to four for more than five exceedance events in a year. In other words, if the number of exceedance events is less than five no punitive capital charges are expected as the avarage VaR is underestimated by less than 2% (i.e., 5 events over 250 trading days).

When analyzing second effects of model risk, we find that certain events exist in which EVT

_____

[22]Results for the remaining indices are robust.

methods might underestimate the true risk to a higher degree than the traditional models applied in this analysis. This is even true for the normal distribution which seems not to be well suited for risk estimation in most of the cases. Regulators may require additional capital for this kind of risk and this may imply an economic capital advantage for traditional models relative to EVT techniques. Such capital buffers have been discussed and can be determined by means of the quantification measures provided in this paper.

REFERENCES

Alexander, C. and Sarabia, J. M. (2012), 'Quantile Uncertainty and Value-at-Risk', *Risk Analysis: An International Journal* **32**(8), 1293–1308.

Angelidis, T., Benos, A. and Degiannakis, S. A. (2004), 'The Use of GARCH Models in VaR Estimation', *Statistical Methodology* **1**(2), 105–128.

Bao, Y. and Ullah, A. (2004), 'Bias of Value-at-Risk', *Finance Research Letters* **1**(4), 241–249.

Basel Committee on Banking Supervision (2010), 'Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems', *Bank for International Settlements* .

Basel Committee on Banking Supervision (2011), 'Revisions to the Basel II Market Risk Framework', *Bank for International Settlements* .

Basel Committee on Banking Supervision (2013), 'Fundamental Review of the Trading Book: A Revised Market Risk Framework', *Bank for International Settlements* .

Berkowitz, J. and O'Brien, J. (2002), 'How Accurate are Value-at-Risk Models at Commercial Banks?', *Journal of Finance* **57**(3), 1093–1111.

Boucher, C., Danielsson, J., Kouontchou, P. S. and Mailet, B. B. (2014), 'Risk models-at-risk', *Journal of Banking and Finance* **44**, 72–92.

Brodin, E. and Rootzén, H. (2009), 'Univariate and Bivariate GPD Methods for Predicting Extreme Wind Storm Losses', *Insurance: Mathematics and Economics* **44**(3), 345–356.

Chan, N., Deng, S., Peng, L. and Xia, Z. (2007), 'Interval Estimation of Value-at-Risk based on GARCH Models with Heavy-Tailed Innovations', *Journal of Econometrics* **137**(2), 556–576.

Christoffersen, P. (1998), 'Evaluating Interval Forecasts', *International Economics Review* **39**(4), 841–862.

Christoffersen, P. and Goncalves, P. (2005), 'Estimation Risk in Financial Risk Managment', *Journal of Risk* **7**(3), 1–28.

Cébrian, A. C., Denuit, M. and Lambert, P. (2003), 'Analysis of Bivariate Tail Dependence using Extreme Value Copulas: An Application to SOA Medical Large Claims Database', *Belgian Actuarial Bulletin* **3**, 33–41.

Escanciano, J. and Olmo, J. (2010), 'Backtesting Parametric Value-at-Risk with Estimation Risk', *Journal of Business and Economic Statistics* **28**(1), 36–51.

Escanciano, J. and Olmo, J. (2011), 'Robust Backtesting Test for Value-at-Risk', *Journal of Financial Econometrics* **9**(1), 132–161.

Ghorbel, A. and Trabelsi, A. (2009), 'Measure of Financial Risk using Conditional Extreme Value Copulas with EVT Margins', *Journal of Risk* **11**(4), 51–85.

Gibson, R., L'Habitand, F. S., Pistre, N. and Talay, D. (1999), 'Interest Rate Model Risk: An Overview', *Journal of Risk* **1**(3), 37–62.

Gneiting, T. (2011), 'Making and Evaluating Point Forecasts', *Journal of the American Statistical Association* **106**, 746–762.

Gourieroux, C. and Zakoyan, J. M. (2013), 'Estimation Adjusted VaR', *Econometric Theory* **29**(4), 735–750.

Hartz, C., Mittnik, S. and Paolella, M. (2006), 'Accurate Value-at-Risk Forecasting based on the Normal-GARCH Model', *Computational Statistics and Data Analysis* **51**(4), 2295–2312.

Hotta, L., K., Lucas, E., C. and Palaro, H., P. (2008), 'Estimation of VaR using Copula and Extreme Value Theory', *Multinational Finance Journal* **12**(3), 205–218.

Inui, K. and Kijima, M. (2005), 'On the Signficance of Expected Shortfall as a Coherent Risk Measure', *Journal of Banking and Finance* **29**(4), 853–864.

Jorion, P. (1996), 'Risk: Measuring the Risk in Value-at-Risk', *Financial Analysts Journal* **52**(6), 47–56.

Kerkhof, J., Melenberg, B. and Schumacher, H. (2010), 'Model Risk and Capital Reserves', *Journal of Banking and Finance* **34**(1), 267–279.

Longin, F. (2000), 'From Value-at-Risk to Stress Testing: The Extreme Value Approach', *Journal of Banking and Finance* **24**(7), 1097–1035.

Longin, F. and Solnik, B. (2001), 'Extreme Correlation of Internatioal Equity Markets', *Journal of Finance* **56**(2), 649–676.

Mancini, L. and Trojani, F. (2011), 'Robust Value-at-Risk Prediction', *Journal of Financial Econometrics* **9**(2), 281–313.

Marinelli, C., D'addona, S. and Rachev, S. T. (2007), 'A Comparison of Some Univariate Models for Value-at-Risk and Expected Shortfall', *International Journal of Theoretical and Applied Finance* **10**(6), 1043–1075.

McNeil, A. and Frey, R. (2000), 'Estimation of Tail-Related Risk Measure for Heteroscedastic Financial Time Series: An Extreme Value Approach', *Journal of Empirical Finance* **7**, 271–300.

McNeil, A., Frey, R. and Embrechts, P. (2005), *Quantitative Risk Management - Concepts, Techniques and Tools*, Princeton University Press.

Nekhili, R., Altay-Salih, A. and Caner, S. (2004), Quantifying Foreign Exchange Market Risk at Different Time Horizons, *in* 'Proceedings of the Eleventh Annual Multinational Finance Society (MFS) Conference', pp. 184–197.

Poon, S. H., Rockinger, M. and Tawn, J. (2004), 'Extreme Value Dependence in Financial Markets: Diagnostics, Models and Financial Implications', *Review of Financial Studies* **17**(2), 581–610.

Pritisker, M. (1997), 'Evaluating Value-at-Risk Methodologies: Accuracy versus Computational Time', *Journal of Financial Services Research* **12**(2), 201–242.

Rootzén, H. and Tajvidi, N. (1997), 'Extreme Value Statistics and Wind Storm Losses: A Case Study', *Scandinavian Actuarial Journal* **1**, 70–94.

Rufino, C., C. and de Guia, E., G. (2011), 'Empirical Comparison of Extreme Value Theory Vis-a-Vis other Methods of VaR Estimation using ASEAN+3 Exchange Rates', *Business and Economics Review* **20**(2), 9–22.

Storn, R. and Price, K. (1997), 'Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continous Spaces', *Journal of Global Optimization* **11**(4), 341–359.

Talay, D. and Zhang, Z. (2002), 'Worst Case Model Risk Management', *Finance and Stochastics* **6**(4), 517–537.

Taylor, J., W. (2008), 'Using Exponentially Weighted Quantile Regression to Estimate Value-at-Risk and Expected Shortfall', *Journal of Financial Econometrics* **6**(3), 382–406.

Zhou, C. (2010), 'Dependence Structure of Risk Factors and Diversification Effects', *Insurance: Mathematics and Economics* **46**(3), 531–540.

Ziegel, J. F. (2014), 'Coherence and Elicitabiliy', *Mathematical Finance* (forthcoming).

Table 7: Failure rates $\sum_{t=D}^{(T-1)} I_t/(T-1-D)$ and estimates for the expected mean of R_t for given confidence levels α, backtesting results and adjustment factors $A_{VaR_\alpha^i}$, $A_{ES_\alpha^i}$

	Value-at-Risk					
	Dax 30			**BSE Sensex**		
$1-\alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
GEV	0.02621*†∘	0.01177*†∘	0.00684*†∘	0.02330*†∘	0.00948*†∘	0.00611*†∘
	(0.02413)	(**0.05130**)	(0.02901)	(-0.04477)	(**-0.00682**)	(**0.07632**)
GPD	0.02545*†∘	0.01215*†∘	0.00570*†∘	0.02607*†∘	0.01106*†∘	0.00711*†∘
	(**0.00513**)	(0.06058)	(**0.02583**)	(**0.00519**)	(0.03694)	(0.09511)
Normal	0.03836	0.01709	0.01291	0.03239	0.01698	0.01224
	(0.10284)	(0.19730)	(0.19138)	(0.06971)	(0.09784)	(0.34756)
Student t	0.03798	0.01519	0.00912	0.03233	0.01383*†	0.00829
	(0.09201)	(0.13148)	(0.08440)	(0.06328)	(0.08899)	(0.22215)
Logistic	0.03608	0.01443∘	0.00836	0.03002*†∘	0.01303*†∘	0.00869
	(0.07563)	(0.07346)	(0.05165)	(0.04623)	(0.08632)	(0.22865)
	Expected Shortfall					
	Dax 30			**BSE Sensex**		
$1-\alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
GPD	0.08909•+	0.15331•+	0.43071•	0.04259•+	0.10442•+	−0.08161•+
	(0.03385)	(0.05044)	(0.13365)	(**0.01604**)	(**0.03291**)	(-0.02241)
Normal	0.19391	0.36517	0.32103•	0.27066	0.38419	0.39064
	(0.08485)	(0.14013)	(0.11328)	(0.11798)	(0.10217)	(0.13749)
Student t	0.01598•+	0.08417•+	0.03631•+	0.08064•+	0.17072•	0.15750•
	(0.00647)	(0.02837)	(0.01096)	(0.03246)	(0.05747)	(0.04725)
Logistic	0.00679•+	0.06498•+	0.02976•+	0.09271•+	0.16162•	0.04907•+
	(**0.00271**)	(**0.02156**)	(**0.00877**)	(0.03685)	(0.05353)	(**0.01441**)

Note: Failure rates and expected means are given in the first row, the second row shows adjustment factors which are displayed in parentheses where the lowest adjustment factors are highlighted in bold. With respect to backtesting, symbols are defined as $*: H_0$ for binomial-test is not rejected; $\dagger: H_0$ for unconditional Likelihood ratio test is not rejected; $\circ: H_0$ for conditional Likelihood ratio test is not rejected.; $\bullet: H_0$ for mean is equal to zero is not rejected; $+: H_0$ for mean is smaller than zero is not rejected.

Table 8: Adjustment factors when using the $\omega = 0.025,\ 0.050,\ 0.100$ quantile of the risk measure estimate distributions in case of the Value-at-Risk

	Dax 30			BSE Sensex		
$1 - \alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
$\hat{VaR}_\alpha^{t,GEV}(0.025)$	0.14297	0.29536	0.22749	0.14061	0.17224	0.28360
$\hat{VaR}_\alpha^{t,GPD}(0.025)$	**0.10077**	0.17121	0.15274	**0.11234**	**0.14080**	0.26751
$\hat{VaR}_\alpha^{t,Normal}(0.025)$	0.16360	0.24978	0.26003	0.14037	0.22833	0.41944
$\hat{VaR}_\alpha^{t,Student\ t}(0.025)$	0.20172	0.34109	0.17969	0.15245	0.19736	0.32648
$\hat{VaR}_\alpha^{t,Logistic}(0.025)$	0.14932	**0.14642**	**0.11825**	0.11547	0.15204	**0.25611**
$\hat{VaR}_\alpha^{t,GEV}(0.050)$	0.12720	0.20059	0.19137	0.10466	0.13916	**0.23169**
$\hat{VaR}_\alpha^{t,GPD}(0.050)$	**0.08389**	0.14965	0.10268	**0.09105**	**0.11420**	0.24433
$\hat{VaR}_\alpha^{t,Normal}(0.050)$	0.15995	0.23582	0.25219	0.12653	0.21799	0.40630
$\hat{VaR}_\alpha^{t,Student\ t}(0.050)$	0.18165	0.21837	0.15863	0.13655	0.18160	0.31627
$\hat{VaR}_\alpha^{t,Logistic}(0.050)$	0.13686	**0.12724**	**0.10007**	0.10471	0.13286	0.24288
$\hat{VaR}_\alpha^{t,GEV}(0.100)$	0.10479	0.17149	0.14241	**0.06464**	**0.10236**	**0.17963**
$\hat{VaR}_\alpha^{t,GPD}(0.100)$	**0.06468**	0.12501	**0.08772**	0.08668	0.10299	0.21325
$\hat{VaR}_\alpha^{t,Normal}(0.100)$	0.14179	0.22936	0.23586	0.10840	0.20881	0.39169
$\hat{VaR}_\alpha^{t,Student\ t}(0.100)$	0.15951	0.19799	0.14905	0.11901	0.16170	0.29827
$\hat{VaR}_\alpha^{t,Logistic}(0.100)$	0.12289	**0.11479**	0.09464	0.08884	0.13140	0.22968

Note: Adjustment factors for the estimation of the Value-at-Risk are displayed in this table similar to Table 7. Lower quantiles from the distributions of risk measure estimates are used instead of point estimates in order to analyze the impact of prediction risk. The smallest adjustment factors, given an index and a confidence interval, are highlighted in bold.

Table 9: Adjustment factors when using the $\omega = 0.025,\ 0.050,\ 0.100$ quantile of the risk measure estimate distributions in case of the Expected Shortfall

	Dax 30			BSE Sensex		
$1-\alpha$	**0.025**	**0.010**	**0.005**	**0.025**	**0.010**	**0.005**
$\hat{ES}_{\alpha}^{t,GPD}(0.025)$	0.10471	0.17951	0.20815	0.11756	0.17608	0.19704
$\hat{ES}_{\alpha}^{t,Normal}(0.025)$	0.10555	0.16955	0.12763	0.12950	0.15839	0.15364
$\hat{ES}_{\alpha}^{t,Student\ t}(0.025)$	0.03955	0.05245	**0.03668**	0.06433	0.06913	**0.05231**
$\hat{ES}_{\alpha}^{t,Logistic}(0.025)$	**0.01886**	**0.05041**	0.04457	**0.04910**	**0.06810**	0.05688
$\hat{ES}_{\alpha}^{t,GPD}(0.050)$	0.09530	0.17177	0.19877	0.10927	0.15343	0.15336
$\hat{ES}_{\alpha}^{t,Normal}(0.050)$	0.10042	0.15546	0.12523	0.12569	0.15233	0.14821
$\hat{ES}_{\alpha}^{t,Student\ t}(0.050)$	0.03179	0.05020	**0.02856**	0.05269	0.06881	0.05111
$\hat{ES}_{\alpha}^{t,Logistic}(0.050)$	**0.01294**	**0.04551**	0.03461	**0.04773**	**0.05785**	**0.04692**
$\hat{ES}_{\alpha}^{t,GPD}(0.100)$	0.07716	0.15421	0.17644	0.10354	0.12358	0.12042
$\hat{ES}_{\alpha}^{t,Normal}(0.100)$	0.09881	0.14043	0.12496	0.12484	0.14897	0.14452
$\hat{ES}_{\alpha}^{t,Student\ t}(0.100)$	0.03155	0.04735	**0.02845**	0.05066	0.06758	0.04931
$\hat{ES}_{\alpha}^{t,Logistic}(0.100)$	**0.01144**	**0.04472**	0.03139	**0.04530**	**0.05547**	**0.04460**

Note: Adjustment factors for the estimation of the Expected Shortfall are displayed in this table similar to Table 7. Lower quantiles from the distributions of risk measure estimates are used instead of point estimates in order to analyze the impact of prediction risk. The smallest adjustment factors, given an index and a confidence interval, are highlighted in bold.