Risk Models–at–Risk*

Christophe M. Boucher
A.A.Advisors-QCG (ABN AMRO), Variances and Univ. Lorraine (CEREFIGE)

Jón Daníelsson
Systemic Risk Centre, London School of Economics

Patrick S. Kouontchou
Variances and Univ. Lorraine (CEREFIGE)

Bertrand B. Maillet
A.A.Advisors-QCG (ABN AMRO), Variances,

Univ. La Reunion and Orleans (CEMOI, LEO/CNRS and LBI)

December 2013

Abstract

The experience from the global financial crisis has raised serious concerns about the accuracy of standard risk measures as tools for the quantification of extreme downward risk. A key reason for this is that risk measures are subject to model risk due, e.g., to specification and estimation uncertainty. While the authorities would like financial institutions to assess model risk, there is no accepted approach for such computations. We propose a remedy for this by a general framework for the computation of risk measures robust to model risk by empirically adjusting imperfect risk forecasts by outcomes from backtesting, considering the desirable quality of VaR models such as the frequency, independence and magnitude of violations. We also provide a fair comparison between the main risk models using the same metric that corresponds to model risk required corrections.

Keywords: Model Risk, Value–at–Risk, Backtesting.
J.E.L. Classification: C50, G11, G32.

1   Introduction

Recent crises have laid bare the failures of standard risk models. High levels of model risk caused models to underforecast risk prior to crisis events, to be slow to react as a crisis unfolds, and then slow to reduce risk levels post–crisis. It is as if the risk models got it wrong in all states of the world. Addressing this problem provides the main motivation for our work. In particular, we explicitly adjust risk forecasts for model risk by their historical performance, so that a risk model learns from its past mistakes. While our focus is on Value–at–Risk (VaR), the analysis applies equally to other risk measures such as expected shortfall (ES).

While there is no single definition of model risk,[1] it generally relates to the uncertainty created by not knowing perfectly the true data generating process (DGP). This inevitably means that any practical definition is linked to such an uncertainty and thus is context dependent. In our case, the end product is a risk forecast, so model risk is the uncertainty in risk forecasting arising from estimation error and the use of an incorrect model. This double uncertainty is responsible both for the range of plausible risk estimates (see, e.g. Beder, 1995), and more generally the inability to forecast risk width acceptable accuracy.
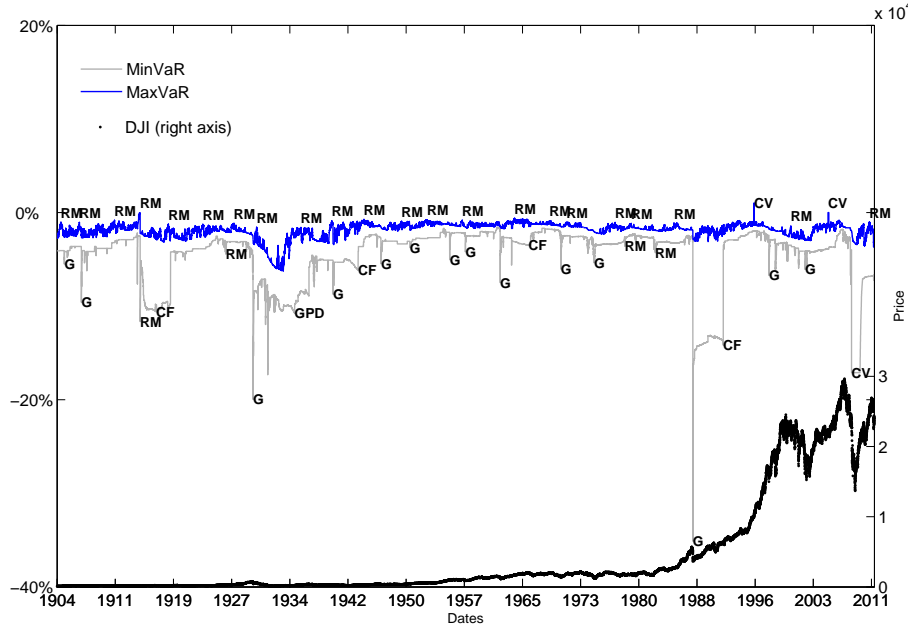
To formalize this, in our view a risk forecast model should meet three desirable criteria: the expected frequency of violations, the absence of violation clustering and a magnitude of violations consistent with the underlying distributional assumptions. These three criteria provide the lens through which to view our empirical results.

We can motivate our contribution by means of an example represented in Figure 1, where, for each day in a sample of the Dow Jones (DJIA) index over a century, we show the outcomes from applying state–of–the–art VaR forecast methods. We also show periodically which method generated the highest and the lowest forecasts. By highlighting the wide disparity between the most common risk forecast methods, the figure illustrates one of the biggest challenges faced by risk managers. Typically, the VaR does not vary much, but when it does, it reacts sharply but belatedly to extreme returns. The range of plausible VaR forecasts is large, where the models producing the highest and lowest forecasts frequently change position across time. Even right after WWII, during a relatively quiet period for financial markets, the

---

[1]In the finance literature, the term "model risk" frequently applies to uncertainty about the risk factor distribution (e.g. Gibson, 2000; Jorion, 2009–a and 2009–b). Although, the term is sometimes used in a wider sense (e.g. Derman, 1996; Crouhy et al., 1998).

most conservative VaR can be four times the most aggressive one.

Figure 1: DJIA and the range of daily 99% VaR forecasts

Daily DJIA index returns from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to dynamically re–estimate parameters for the various methods. The letters "H", "N", "t", "CF", "RM", "G", "CV", "GEV", "GPD" stand for, respectively, historical, normal, Student, Cornish-Fisher, exponential weighted moving average (EWMA or RiskMetrics), GARCH, CAViaR, GEV and GPD methods for VaR calculation.
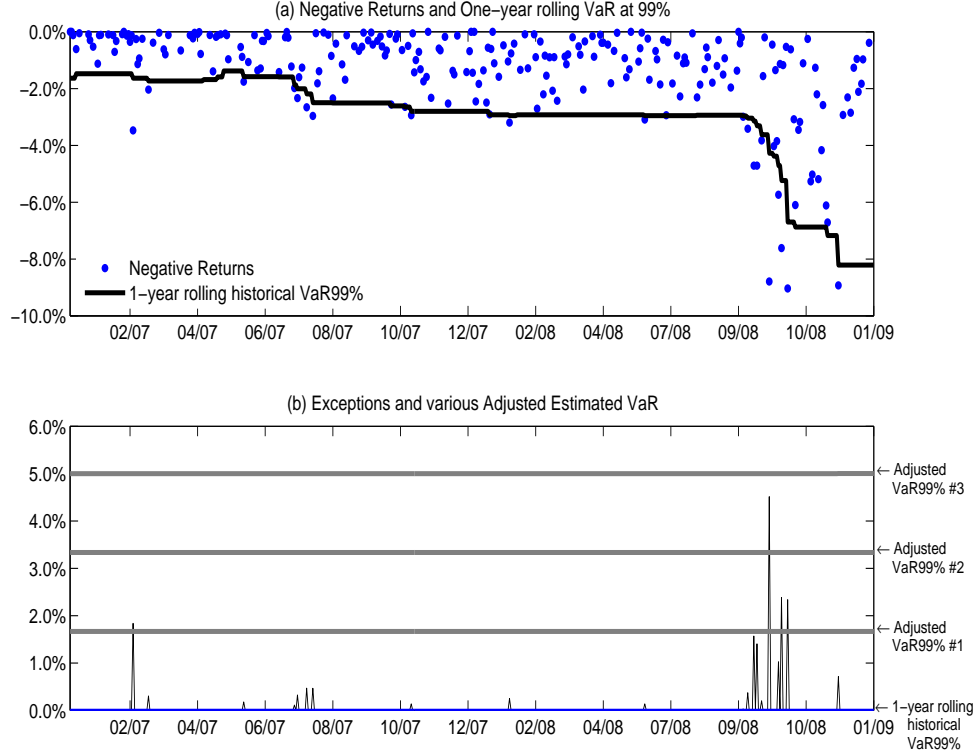
As in Daníelsson et al. (2011), the main conclusion from this brief analysis is that risk managers face a large range of plausible forecast methods and their associated model risk, having to choose between desirable criteria such as performance, degree of conservativeness or forecast volatility. This challenge motivates our main objective where we propose a general method for the correction of imperfect risk estimates, whatever the risk model.

We illustrate our approach by considering events around the Lehman Brothers' collapse, as presented in Figure 2 for the period of January 1st, 2007 to January 1st, 2009. The Figure displays peaks–over–VaR for one–year rolling daily historical 99% VaR on the S&P 500 index.

The figure shows that the hits are excessively frequent, highly autocorrelated and, around October 2008, far from the estimated VaR, even if it progressively adjusted after the hits. This suggests that an optimal buffer would make the VaR forecast more robust.

However, it is not trivial to calculate the buffer, after all, the properties of hits are significantly different in terms of frequency, dependence and size, depending both on the underlying VaR model and probability level as well

Figure 2: S&P500 negative returns and daily 99% VaR forecasts around the 2008 Lehman Brothers's event

(a) Negative Returns and One–year rolling VaR at 99%

(b) Exceptions and various Adjusted Estimated VaR

Daily S&P500 index from the 1st January, 2003 to the 1st January, 2009. The figure presents peak–over–VaR based on the four–year rolling daily historical 99% VaR on the S&P 500 index, as well as corrected VaR estimates with various *ad hoc* incremental buffers (numbered from #1 to #3).

as the magnitude of the buffer. A large (respectively small) buffer correction will lead to a too conservative (too little) protection. The question for the risk manager is then how to ex ante fix the size of this buffer, as illustrated by the three arbitrary correction factors labelled #1, #2 or #3, on the right–hand side y–axis in Figure 2.

In the financial literature, a number of papers have considered estimation risk for risk models, see for instance Gibson et al., 1999; and Talay and Zheng, 2002. The issue of estimation risk for VaR has been considered for the identically and independently distributed return case by, for example, Pritsker (1997) and Jorion (2007). Estimation risk in dynamic models has also been studied by several authors. Berkowitz and O'Brien (2002) observe that the usual VaR estimates are too conservative. Figlewski (2004) examines the effect of estimation errors on the VaR by simulation. The bias of the VaR estimator, resulting from parameter estimation and misspecified distribution,

is studied for ARCH(1) models by Bao and Ullah (2004). In the identical and independent setting, Inui and Kijima (2005) show that the nonparametric VaR estimator may have a strong positive bias when the distribution features fat–tails. Christoffersen and Gonçalves (2005) study the loss of accuracy in VaR and ES due to estimation errors and constructed bootstrap predictive confidence intervals for risk measures. Hartz et al. (2006) propose a re–sampling method based on bootstrap to correct the bias in VaR forecasts for the Gaussian GARCH model. For GARCH models with heavy–tailed distributions, Chan et al. (2007) derive the asymptotic distributions of extremal quantiles. Escanciano and Olmo (2009, 2010 and 2011) study the effects of estimation risk on backtesting procedures. They show how to correct the critical values in standard tests used, when assessing the quality of VaR models. Gouriéroux and Zakoïan (2013) quantify in a GARCH context the effect of estimation risk on measures for estimation of portfolio credit risk and show how to adjust risk measures to account for estimation error. Gagliardini et al. (2012) propose estimation and granularity adjustments for VaR, whilst Lönnbark (2010) derives adjustments of interval forecasts to account for parameter estimation.

In the context of extreme risk measures, our work also relates to Kerkhof et al. (2010), who first propose an incremental market risk capital charge calibrated on the backtesting framework of the regulators. Our present work documents the proposed methodology and complements their approach, generalizing the tests used for defining the buffer. Alexander and Sarabia (2012) also explicitly deal with VaR model risk by quantifying VaR model risk and propose an adjustment to regulatory capital based on a maximum relative entropy *criterion* to some benchmark density. In a similar manner, Breuer and Csiszár (2012 and 2013) and Breuer et al. (2012) define model risk as an amplified largest loss based on a distribution which is at a reasonable, Mahalanobis or Kullback–Leibler, distance to a reference density.

We start with a controlled experiment, whereby we simulate an artificial long time–series which exhibits the salient features of financial return data. We then estimate a range of VaR forecast models with this data, both identifying model risk and more importantly dynamically adjusting the risk forecasts with respect to such risk.

The conclusions from this exercise lead us to a number of interesting conclusions. First, by dynamically adjusting for estimation bias we significantly improve the performance of every method, suggesting that such an approach might be valid in routine applications of risk forecasting. Second, the model bias is large in general, and sometimes to the same order as the VaR measure itself, and very different across methods. Finally, the bias strongly depends

upon the probability confidence level. This suggests that a commonly advocated approach of probabilities shifting — whereby we estimate a model with one probability to better estimate a VaR with a less extreme probability — is not valid.

The Monte Carlo results motivate our main contribution, the development of a practical method for dealing with model uncertainty. Since we do not know the "true" model, we instead learn from history by evaluating the historical errors in order to use them to dynamically adjust future forecasts. We reach a range of empirical conclusions from this exercise.

1. The magnitude of corrections can sometimes be large, especially around the 1929 and 2008 crises, ranging from 0 to 15% for some methods to more than 100% in some circumstances;

2. The EWMA and GARCH VaR are among the preferred models, since the minimum correction to pass main backtests are among the smallest;

3. Regardless of the model, a ten year sample period is needed to have a fairly good idea of the magnitude of the required correction;

4. The model risk of the correction buffer can be measured and the buffer fine–tuned according to the link between the confidence level on the required correction. This enables risk managers to explicitly tailor the buffer to major financial stress episodes such as the Great Depression of 1929 or the 2008 crisis, if they choose to do so;

5. By considering multivariate indexes and portfolios, we find that the model risk adjustment buffer is in line with the multiple k imposed by regulators (from 3 to 5);

6. The general methodology can be used to gauge the plausibility of traditional handpicked stress–test scenarios.

The outline of the paper is as follows: Section 2 evaluates the extent to which elementary model risks affect VaR estimates based on realistic simulations. Section 3 proposes a practical method to provide VaR estimates robust to model risk. Section 4 finally concludes, whilst the Appendix follows, outlining some description and examples of model risks and the main backtesting methods used in the paper.

2 Analysis of estimation and specification errors

Consider the best case scenario where we know the DGP but where the sample size is small. In this case, the estimated VaR will inevitably be an imperfect estimate of the theoretic, or true, VaR. In particular, there exists a ε that makes the equality between the theoretic and empirical exact:

$$\text{ThVaR}(\theta_0, \alpha) = \text{EVaR}(\hat{\theta}, \alpha) + \varepsilon, \tag{1}$$

where $\hat{\theta}$ denotes the estimated parameters, θ_0 the true parameters and α the probability level of the VaR. The theoretic true VaR is denoted by $\text{ThVaR}(\theta_0, \alpha)$ and the estimated VaR by $\text{EVaR}(\hat{\theta}, \alpha)$.

We hereafter denote the bias ε by the function[2] $\text{bias}(\hat{\theta}, \theta_0, \alpha)$. In this best case scenario (when the true VaR is known), we know the bias function, and can therefore obtain the perfect estimation adjusted VaR (PEAVaR), with the estimated VaR ($\text{EVaR}(\hat{\theta}, \alpha)$), by:

$$\text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{EVaR}(\hat{\theta}, \alpha) + \text{bias}(\hat{\theta}, \theta_0, \alpha). \tag{2}$$

As a general rule, the smaller α is, the better we forecast VaR and identify the bias function. The reason is that, for a given sample size, the number of quantiles increases along with decreasing α, so the effective sample size used in the forecasting exercise increases. As the probabilities become more extreme, so does the accuracy of the VaR forecasts decrease, for example because fewer observations are used in the estimation. Consequently, it is harder to model the shape of the tail than the shape of the interior distribution. For this reason, it might be tempting to forecast VaR slightly closer to the center of the distribution, perhaps at $\alpha = 95\%$, and then use those estimation results to get at the VaR for more extreme probability levels, like $\alpha = 99\%$ or $\alpha = 99.9\%$. This is often referred to as *probability shifting*.

2.1 Probability shifting

We can analyze the impact of probability shifting within our framework by defining two random probabilities, $\tilde{\alpha}^*$ and $\tilde{\alpha}^{**}$, so that:

$$\begin{cases} \text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{EVaR}(\hat{\theta}, \tilde{\alpha}^*) = \text{ThVaR}(\theta_0, \tilde{\alpha}^*) \\ \text{EVaR}(\hat{\theta}, \tilde{\alpha}^{**}) = \text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{ThVaR}(\theta_0, \alpha), \end{cases} \tag{3}$$

---

[2]See the Appendices for examples of such bias functions in various contexts of model risks.

or equivalently, with F and $\hat{F}$ representing, respectively, the theoretic and estimated cumulative density functions:

$$\begin{cases} \tilde{\alpha}^* = F\left[\hat{F}^{-1}(\alpha)\right] \\ \tilde{\alpha}^{**} = \hat{F}\left[F^{-1}(\alpha)\right], \end{cases} \tag{4}$$

with $\hat{F}^{-1}(\alpha) = \text{EVaR}(\hat{\theta}, \alpha)$ and $F^{-1}(\alpha) = \text{ThVaR}(\theta_0, \alpha)$.

If one were to use $\tilde{\alpha}^*$ instead of α, the bias adjusted VaR results, whilst $\tilde{\alpha}^{**}$ achieves the opposite, mapping the probability corresponding to the biased VaR, to the theoretic VaR.

It follows that if $\tilde{\alpha}^* > \alpha > \tilde{\alpha}^{**}$, the estimated VaR is biased towards zero, whilst if $\tilde{\alpha}^* < \alpha < \tilde{\alpha}^{**}$, it is biased towards minus infinity.

2.2 Monte Carlo examination

Many potential sources of error can significantly impact on the accuracy of risk forecasts. The sources one is most likely to encounter in day–to–day risk forecasting, and certainly in most academic studies, are estimation and specification errors. For this reason, we investigate these two in detail by means of Monte Carlo experiments.

We consider below the distribution of the errors between the poorly estimated VaR and the true VaR when considering, alternatively, estimation risk, specification uncertainty or both. We first specify a GDP from which we generate data. We then treat the DGP as unknown and forecast VaR for the simulated data.

As before, the true parameters are θ_0, but we now also have the true parameters of the misspecified model, indicated by θ_1, as well as its estimate $\hat{\theta}_1$. In this case, we indicate the estimated VaR by $\text{EVaR}(\hat{\theta}_1, \alpha)$ and define the perfect model risk adjusted VaR (denoted herein PMAVaR) by:

$$\text{PMAVaR}(\hat{\theta}_1, \alpha) = \text{EVaR}(\hat{\theta}_1, \alpha) + \text{bias}(\theta_0, \hat{\theta}_1, \alpha). \tag{5}$$

We first present the theoretical framework related to the correction procedure in a static setting for the sake of simplicity. However, in the subsequent empirical application, we also consider the dynamic properties of our correction procedure that is proposed at date t based on the conditional information available at date $t - 1$.

8

2.2.1 The true model

The DGP needs to be sufficiently general to capture the salient features of financial return data. Because we are not limited by the need to estimate a model, we can specify a DGP that might be difficult, to the point of impossible, to estimate in small samples. The DGP we employ is a second order Markov–switching generalized autoregressive conditionally heteroskedastic with Student–t disturbances (hereafter denoted MS(2)-GARCH(1,1)-t)[3] as in Frésard et al. (2011) in a VaR context.[4]

More precisely, the DGP is:

$$r_t = \mu_{s_t} + \sigma_{s_t} z_t, \tag{6}$$

where the z_t innovations series are independently and identically distributed as a standard Student distribution with υ degrees of freedom ($z_t \sim \text{iid} St(0, 1, \upsilon)$), and $\sigma_{s_t}^2 = \omega_{s_t} + \alpha_{s_t} \varepsilon_{t-1}^2 + \beta_{s_t}^2 \sigma_{s_{t-1}}^2$, with $s_t \in \{1, 2\}$ characterizes the state of the market, μ_{s_t} is the mean return and with υ degrees of freedom, and where $\omega_{s_t} > 0$, $\alpha_{s_t} \geq 0$, $\beta_{s_t} \geq 0$ are the parameters of the GARCH(1,1) in the two states, and $\varepsilon_t = r_t - \mu_{s_t}$ the return innovations with the fat tails of a Student density with a υ degree of freedom.

The state is modelled with a Markov chain whose matrix of transition probabilities is defined by $p_{ij} = \Pr(s_t = j | s_{t-1} = i)$. Appropriately chosen restrictions on the GARCH coefficients ensure that σ_t^2 is almost strictly positive.

Using this DGP, we first simulate a long artificial series of 360,000 daily returns with estimated parameters on the daily DJIA from the 1st January, 1990 to the 20th September, 2011[5]. We then forecast various VaRs using 1,000 observations, and finally compute main statistics of the forecast error,

---

[3]See Hamilton and Susmel, 1994; Gray, 1996; Klaassen, 2002; Haas et al., 2004, for more details on the process.

[4]As a complement (not reported here for space reasons, but available on demand in a web Appendix), we also made use of other alternative frameworks: a Student *versus* a normal density, as well as Brownian, Lévy and Hawkes processes, with the same qualitative response with a relative model error for VaR ranging from $5 - 15\%$ in the simplest cases (Gaussian estimation risk with 250 observations) to as large as 200% when the process is complex and the sample small (the case of Hawkes processes).

[5]The estimated parameters of the MS(2)-GARCH(1,1) model on the DJI Index are $\omega_1 = 3.1699e^{-006}$, $\beta_1 = 0.90801$, $\alpha_1 = 0.0733081$, $\omega_2 = 2.509e^{-005}$, $\beta_2 = 0.10453$, $\alpha_2 = 0.0064734$, $\mu_1 = 0.00$, $\mu_2 = 0.00$, $\upsilon = 5.56$, $p_{11} = 0.99654$ and $p_{22} = 0.99328$. Bauwens et al. (2010) obtain approximately the same results on the S&P. This estimation is crucial since the transition probabilities between states and auto-regressive parameters both affect the persistence of the simulated processes. Our estimates are here very similar to those exhibited in the literature (e.g., Bauwens et al., 2010; Billio et al., 2012; Frésard et al., 2011). Moreover, when artificially considering different probabilities related to the second

measured by differences between the asymptotic VaR (computed with the true simulated DGP on 360,000 observations) and empirical ones recovered from limited samples.

2.2.2 Misspecification and a parameter estimation uncertainty

Our focus is on the annualized daily 95%, 99% and 99.5% VaR. Table 1 illustrates the model risk of VaR estimates, defined as the implication of model misspecification and a parameter estimation uncertainty. We examine this model risk by comparing simulations and estimates corresponding to a normal GARCH(1,1) and a MS(2)-GARCH(1,1)-t. The columns represent respectively the average adjusted VaR according to specification and/or estimation errors, the theoretic VaR, the average, the minimum and maximum values of the adjustment terms. Note that a negative adjustment term indicates that the estimated VaR (which is a negative return) should be more conservative (more negative).

We present the estimation bias ($\mathrm{bias}(\theta_0, \hat{\theta}, \alpha)$), in Panel A of Table 1, when we simulate a simple model (Normal GARCH(1,1)) and use the appropriate methodology for computing the VaR (Normal GARCH VaR). This bias arises only due to the small estimation sample size (1,000) and is zero for the full 360,000 sized sample. However, the dispersion of this estimation bias is quite large since the minimum and the maximum values of the bias (or adjustment term) represent about 50% of the true VaR. For example, with $\alpha = 99\%$, the minimum and maximum biases are respectively equal to -33% and +32% for a true VaR of -60%.

The specification bias ($\mathrm{bias}(\theta_0, \theta_1, \alpha)$) is presented in Panel B of Table 1, where the quantiles were modelled by a GARCH(1,1) VaR. Within this specific illustration, the risk model is fully explained by the discrepancy between the DGP and the assumed simple risk model used (since the parameters are here known and the estimation bias is zero by definition); the specification bias is thus constant and depends upon the choice of the risk model specification. The average specification bias is large here; it is negative and increases in absolute terms with α, which indicates that extreme risks of the MS(2)-GARCH(1,1)-t DGP are generally underestimated by the GARCH(1,1) parametric VaR model.

---

state, we find the same qualitative results in what we are interested in: model risk of risk models. Last but not least, when we have adopted other representations of financial returns (either using processes or densities), we again reach the same order of magnitude of the worst errors of forecasting (additional results available upon request).

Table 1: Conditional simulated errors associated with the 95%, 99% and 99.5% VaR: GARCH(1,1) *versus* MS(2)-GARCH(1,1)-t

pair	Panel A. GARCH(1,1) DGP and GARCH(1,1) VaR with Estimation Error					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-36.16%	-36.16%	.00%	.02%	-19.53%	19.60%
$\alpha = 99.00\%$	-59.70%	-59.70%	.00%	.04%	-32.66%	32.02%
$\alpha = 99.50\%$	-70.99%	-70.99%	.00%	.06%	-38.35%	38.03%

	Panel B. MS(2)-GARCH(1,1)-t DGP and GARCH(1,1) VaR with Specification Error					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-30.78%	-36.16%	-5.38%	-5.38%	-5.38%	-5.38%
$\alpha = 99.00\%$	-43.83%	-59.70%	-15.87%	-15.87%	-15.87%	-15.87%
$\alpha = 99.50\%$	-48.61%	-70.99%	-22.38%	-22.38%	-22.38%	-22.38%

	Panel C. MS(2)-GARCH(1,1)-t DGP and GARCH(1,1) VaR with Specification and Estimation Errors					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-28.97%	-36.16%	-7.19%	-8.83%	-21.70%	18.99%
$\alpha = 99.00\%$	-41.28%	-59.70%	-18.42%	-20.76%	-38.88%	18.02%
$\alpha = 99.50\%$	-45.78%	-70.99%	-25.20%	-27.79%	-47.84%	15.03%

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. These statistics were computed with the results of 360,000 simulated series of 1,000 daily returns according to a specific DGP (rescaled GARCH(1,1) for Panel A and MS(2)-GARCH(1,1)-t for Panels B and C) using an annualized Normal GARCH VaR (in all Panels). The columns represent, respectively, the average adjusted VaR according to specification and/or estimation errors, the theoretical VaR, the average, the minimum and the maximum value of the adjustment terms. A negative adjustment term indicates that the estimated VaR (negative return) should be more conservative (more negative). Panel A presents GARCH(1,1) DGP and/or estimated GARCH VaR; Panel B relates to a MS(2)-GARCH(1,1) DGP with estimated GARCH VaR; Panel C refers to an estimated MS(2)-GARCH(1,1) DGP with results from an estimated GARCH VaR.

The estimation and specification biases are captured simultaneously in Panel C. These components of model risk are jointly considered and, in the worst cases, they merely add up in an independent manner. We compute the global error — denoted $bias(\theta_0, \theta_1, \hat{\theta}_1, \alpha)$ in its most general formulation — as the difference between the true VaR and the estimated VaR according to a misspecified VaR model estimated on a limited sample. As in Panel B, where

a normal GARCH(1,1) VaR is used with a simulated MS(2)-GARCH(1,1)-t, the average bias is negative and increases in absolute terms with α. The mean errors are thus equivalent to the specification bias component, but the dispersion of the model risk realizations is inflated by the estimation bias.

2.2.3   Probability shifting

We illustrate the impact of probability shifting and model risk in Table 2, which shows the two modified probability levels $\tilde{\alpha}^*$ and $\tilde{\alpha}^{**}$. The former is associated to the true density and corresponds to the (mis-)estimated $(1-\alpha)$-VaR, whilst the latter, associated to the estimated VaR, corresponds to the $(1-\alpha)$-VaR without model error.

The gap between $\tilde{\alpha}^*$ and α can be interpreted as a measure of the model risk of the risk model. The gap between $\tilde{\alpha}^{**}$ and α can also be analyzed as the probability shift that we should apply using a specific model of VaR to reach the true VaR.

This alternative representation of the model risk of risk models shows that $\tilde{\alpha}^{**}$ is often unreachable and cannot be used for correcting the estimated VaR. For instance, the maximum associated with the 99.5% VaR in Panel C has to be superior to 100%, which cannot in practice be discriminated from the maximum, i.e. when associated with the 100% probability. More generally, $\tilde{\alpha}^{**}$ is frequently superior to α, (and $\tilde{\alpha}^*$ generally inferior to α) which can be interpreted as an under–estimation of the risk using the proposed model of VaR (the estimated VaR is too aggressive).

This suggests that the recent call of some authorities for more extreme quantiles (see, e.g. FSA, 2006), i.e. VaR 99.5% or 99.9%, is not warranted since in some cases the real VaR appears below the worst estimated return.

Finally, our results show, surprisingly, that the mean bias is not a simple increasing function of the VaR and, accordingly, of the level of probability associated to the VaR. The expected adjustment associated to the 99.5% (99%) probability level is, for instance, four (two) times larger than the expected adjustment associated to the 95% probability level and represents an increase of nearly 15% (10%). The relation between the model risk and the probability associated to the VaR is not linear and depends on several components.

The implemented estimated VaR should be corrected by an adjustment corresponding to the global bias linked to the potential model risk error. However, the true perfect VaR is generally unknown by definition. The proposed adjustments are thus impossible to quantify accurately outside a pure academic

simulation exercise.

Table 2: Probability shifts associated with 95%, 99% and 99.5% annualized VaR: GARCH(1,1) *versus* MS(2)–GARCH(1,1) quantiles

	Probability $\tilde{\alpha}^*$ associated to the true density corresponding to the (mis-)estimated VaR				Probability $\tilde{\alpha}^{**}$ associated to the biased empirical density corresponding to the perfect VaR			
	Panel A. GARCH(1,1) DGP and GARCH(1,1) VaR with Estimation Error							
Estimated VaR	Mean Shift	Median Shift	Min Shift	Max Shift	Mean Shift	Median Shift	Min Shift	Max Shift
$\alpha = 95.00\%$	94.19%	94.24%	90.37%	99.31%	94.51%	94.26%	94.36%	99.88%
$\alpha = 99.00\%$	98.92%	98.95%	96.83%	99.92%	99.05%	99.08%	98.49%	99.99%
$\alpha = 99.50\%$	99.25%	99.38%	98.71%	99.97%	99.47%	99.09%	99.98%	N.R.
	Panel B. MS(2)-GARCH(1,1)-t DGP and GARCH(1,1) VaR with Specification Error							
Estimated VaR	Mean Shift	Median Shift	Min Shift	Max Shift	Mean Shift	Median Shift	Min Shift	Max Shift
$\alpha = 95.00\%$	95.81%	95.81%	95.81%	95.81%	97.29%	97.29%	97.29%	97.29%
$\alpha = 99.00\%$	98.64%	98.64%	98.64%	98.64%	99.92%	99.92%	99.92%	99.92%
$\alpha = 99.50\%$	99.07%	99.07%	99.07%	99.07%	99.99%	99.99%	99.99%	99.99%
	Panel C. MS(2)-GARCH(1,1)-t DGP and GARCH(1,1) VaR with Specification and Estimation Errors							
Estimated VaR	Mean Shift	Median Shift	Min Shift	Max Shift	Mean Shift	Median Shift	Min Shift	Max Shift
$\alpha = 95.00\%$	94.15%	94.29%	82.43%	99.44%	97.44%	98.47%	85.69%	N.R.
$\alpha = 99.00\%$	97.71%	97.94%	89.81%	99.88%	99.78%	99.98%	96.27%	N.R.
$\alpha = 99.50\%$	98.35%	98.56%	91.71%	99.92%	99.93%	N.R.	98.32%	N.R.

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. These statistics were computed with the results of 360,000 simulated series of 1,000 daily returns according to a specific DGP (rescaled GARCH(1,1) for Panel A and MS(2)-GARCH(1,1)-t for Panels B and C) using an annualized Normal GARCH VaR (in all Panels). The columns represent, respectively, the average Estimated VaR according to specification and/or estimation errors, the mean, the minimum and the maximum of the modified probability level $\tilde{\alpha}^*$, the mean, the minimum and the maximum of the modified probability level $\tilde{\alpha}^{**}$. The letters N.R. stand for "Not Reached", *i.e.* condition on bounds is not met even for 100.00%. Panel A presents GARCH(1,1) DGP and/or estimated GARCH VaR; Panel B relates to a MS(2)-GARCH(1,1) DGP with estimated GARCH VaR; Panel C refers to an estimated MS(2)-GARCH(1,1) DGP with results from an estimated GARCH VaR.

3 An economic valuation of model risk

While the illustration above is focused on the controlled experiment where the modeller knows the true model, in reality the true is not known. To address this, we propose a practical method for dealing with model uncertainty, that makes use of past historical errors related to specific estimated models. While it is not possible to optimally adjust for biases, we can approximate them by adjusting the VaR forecasts by the model's historical performance. More concretely, historical errors are used to adjust future forecasts by identifying the minimum correction factor needed to pass backtest criteria.

We first define the imperfect model adjusted VaR (IMAVaR) as:

$$\text{IMAVaR}(\hat{\theta}_1, \alpha) = \text{EVaR}(\hat{\theta}_1, \alpha) + \text{adj}(\theta_0, \theta_1, \hat{\theta}_1, \alpha), \tag{7}$$

where $\text{EVaR}(\cdot)$ is an estimated VaR with a specific risk model, $\hat{\theta}_1$ are model parameters estimated with T observations, and $\text{adj}(\theta_0, \theta_1, \hat{\theta}_1, \alpha)$ the minimum VaR adjustment for the risk model, so that:

$$\text{IMAVaR}(\hat{\theta}_1, \alpha, n) = \underbrace{\sup}_{\text{VaR} \in \mathbb{R}} \left\{ \text{VaR}(\alpha)^* \right\}, \tag{8}$$

where the symbol $\mathbb{R}$ refers to real numbers set, $\text{VaR}(\cdot)^*$ is a set of VaRs, from a model, and $\text{IMAVaR}(\cdot)$ is the lowest acceptable VaR, as perhaps identified by the authorities.

The better the VaR model and the lower the minimum required adjustment and vice–versa. The next step is to make explicit the process that defines the limit of VaR that bounds the IMAVaR.

3.1 General backtest procedures

A variety of tests have been proposed in the literature to gauge the accuracy of VaR estimates. In our view, there are three desirable properties that should be met by a risk model: the expected frequency of violations, the absence of violation clustering and the consistency of exception magnitudes to the underlying statistical model in the parametetric case.

3.1.1 Frequency

The unconditional coverage test (Kupiec, 1995) is based on comparing the observed number of violations to the expected[6]. The hit variable, obtained

---

[6]Note that the Basel "traffic light" backtesting framework is directly inspired by this unconditional coverage test. Escanciano and Pei (2012) show, however, that this uncon-

from the ex post observation of EVaR($\cdot$) violations for threshold α and time t, denoted $I_t^{\text{EVaR}}(\alpha)$, is defined as:

$$I_t^{\text{EVaR}(\cdot)}(\alpha) = \begin{cases} 1 \text{ if } r_t < -\text{EVaR}(\hat{\theta}, \alpha)_{t-1} \\ 0 \text{ otherwise,} \end{cases}$$

where r_t is the return at time t, with $t = [1, 2, \ldots, T]$.

If we assume that $I_t^{\text{EVaR}}(\cdot)$ is iid, then, under the unconditional coverage hypothesis (Kupiec, 1995), the total number of VaR exceptions, denoted $\text{Hit}_t^{\text{EVaR}}(\alpha)$, follows a binomial distribution (Christoffersen, 1998), denoted $B(T, \alpha)$:

$$\text{Hit}_t^{\text{EVaR}(\cdot)}(\alpha) = \sum_{t=1}^{T} I_t^{\text{EVaR}(\cdot)}(\alpha) \sim B(T, \alpha). \tag{9}$$

Under the null hypothesis, the likelihood ratio, $LRuc$, has the asymptotic distribution:

$$LRuc^{I_t^{\text{VaR}(\cdot)}(\alpha)} = 2\left\{\log\left[\widehat{\alpha}^{T_I}(1-\widehat{\alpha}^{T-T_I})\right] - \log\left[\alpha^{T_I}(1-\alpha^{T-T_I})\right]\right\} \xrightarrow{d} \chi^2(1), \tag{10}$$

where the symbol $\xrightarrow{d}$ denotes the convergence in distribution of the test statistic, $T_I = T \times E\left[I_t^{\text{EVaR}(\cdot)}\right]$ is the number of exceptions and $\widehat{\alpha} = T_I/T$ is the unconditional coverage.

3.1.2 Independence

Christoffersen (1998) proposed a test for the independence of violations:

$$\text{LRind}^{I_t^{\text{EVaR}}(\alpha)} = 2\left[\log L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11}) - \log L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi)\right] \xrightarrow{d} \chi^2(1), \tag{11}$$

where $\pi_{ij} = \text{Pr}\left[I_t^{\text{EVaR}}(\alpha) = j | I_{t-1}^{\text{EVaR}} = i\right]$ is a Markov chain that reflects the existence of an order 1 memory in the process $I_t^{\text{EVaR}}(\alpha)$, $L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11}) = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{10}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}$ is thus the likelihood under the hypothesis of the first–order Markov dependence, $L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi)$ is the likelihood under the hypothesis of independence, such as $\pi_{01} = \pi_{11} = \pi$, with T_{ij} the number of observations in the state j for the current period and at state i for the previous period, $\pi_{01} = T_{01}/(T_{00} + T_{01})$, $\pi_{11} = T_{11}/(T_{10} + T_{11})$ and $\pi = (T_{01} + T_{11})/T$.

---

ditional test is always inconsistent in detecting non–optimal VaR forecasts based on the historical method. In the following, nevertheless, we consider for our adjustment procedure three of the main tests (including the unconditional coverage test), as well as their bootstrapped corrected versions.

3.1.3 Magnitude

A third class of tests focuses on the magnitude of the losses experienced when VaR limits are violated. While this is not relevant for methods such as historical simulation, it provides a useful evaluation of the parametric approaches. Berkowitz (2001), for instance, proposes a hypothesis test for determining whether the magnitudes of observed VaR exceptions are consistent with the underlying VaR model, such as:

$$\text{LRmag}^{\gamma_{t+1}} = 2\left[L_{\text{mag}}^{\gamma_{t+1}}(\mu, \sigma) - L_{\text{mag}}^{\gamma_{t+1}}(0, 1)\right] \xrightarrow{d} \chi^2(2), \qquad (12)$$

where γ_{t+1} is the magnitude variable of the observed VaR exceptions, μ and σ are unconditional mean and standard deviation of γ_{t+1} series, and where:

$$
\begin{aligned}
L_{\text{mag}}^{\gamma_{t+1}}(\mu, \sigma) = \quad &\sum\nolimits_{\{\gamma_{t+1}=0\}} \log\left\{1 - \Phi\left\{\frac{\Phi^{-1}(\alpha) - \mu}{\sigma}\right\}\right\} \\
+ \quad &\sum\nolimits_{\{\gamma_{t+1}\neq 0\}} \left\{-\tfrac{1}{2}\log(2\pi\sigma^2) - \frac{(\gamma_{t+1}-\mu)^2}{2\sigma^2} - \log\left\{\Phi\left\{\frac{\Phi^{-1}(\alpha) - \mu}{\sigma}\right\}\right\}\right\}.
\end{aligned}
$$

For both unconditional and conditional coverage tests, Escanciano and Olmo (2009, 2010 and 2011) alternatively approximate the critical values of these tests by using a sub–sampling bootstrap methodology, since they show that the coverage VaR backtest is affected by model misspecification.

3.2 A desirable VaR and the backtests

Under the H_0 hypothesis, a desirable VaR passes each of these three test criteria:

$$
\begin{cases}
LRuc^{I_t^{\text{VaR}(\cdot)}(\alpha)} \xrightarrow{d} \chi^2(1) \text{ for the hit test;} \\[2mm]
\text{LRind}_t^{I^{\text{VaR}(\cdot)}(\alpha)} \xrightarrow{d} \chi^2(1) \text{ for the independence test;} \qquad (13) \\[2mm]
\text{LRmag}_t^{\gamma_{t+1}(\alpha)} \xrightarrow{d} \chi^2(2) \text{ for the exception magnitude test.}
\end{cases}
$$

We now have to search for the minimal adjustment value q^* that allows us to pass all the tests (one–by–one or jointly). For a given VaR forecast and the bounding range for the tests above, we can obtain the $IMAVaR$ that respects conditions (10), (11) and/or (12) (or their sub–sampled versions). More precisely, given a sequence of predictions $\left\{VaR_t(\hat{\theta}, \alpha) : t = [1, \cdots, T]\right\}$, we construct the set of values $q \in \mathbb{R}$ such that the sequence $\left\{VaR_t(\hat{\theta}, \alpha) + q : t = \right.$

$[1, \cdots, T]\}$ passes several backtests. If we denote the set of accepted adjustments by $A_T(\alpha)$, the optimal adjustment is given by:[7]

$$q_T^* = \arg \underbrace{\min_{q \in A_T(\alpha)}} \{q\}. \tag{14}$$

We use a numerical optimisation technique to solve the program (14): During the adjustment process, we search for the optimal adjustment, starting with a large negative value of q^*, increasing it slowly, until the adjusted VaR allows us to pass all the tests.[8]

The program (14) gives the optimal value of adjustment of the imperfect VaR estimation to become a desirable VaR. This means that the H_0 hypothesis is true for the selected backtest method, so that the test statistic is lower than critical values for all tests at the threshold α. In what follows, in order to distinguish the effect of each test, we provide each correction separately, corresponding to each of the tests taken alone.[9]

As a first illustration, Figure 3 provides the minimum adjustments (errors), denoted $\underline{q}^*$ as solutions of the program (14). We first only consider the hit test, for the historical, the Gaussian and the GARCH VaRs computed on the DJIA over one century of daily data. The figure represents the minimal adjustment (in a percentage of the underlying VaR) necessary to respect the hit ratio criteria according to the VaR level of confidence (95% to 99.5%). This minimal adjustment is here considered as a proxy for the economic value of the model risk; it is expressed as a proportion of the observed average VaR.
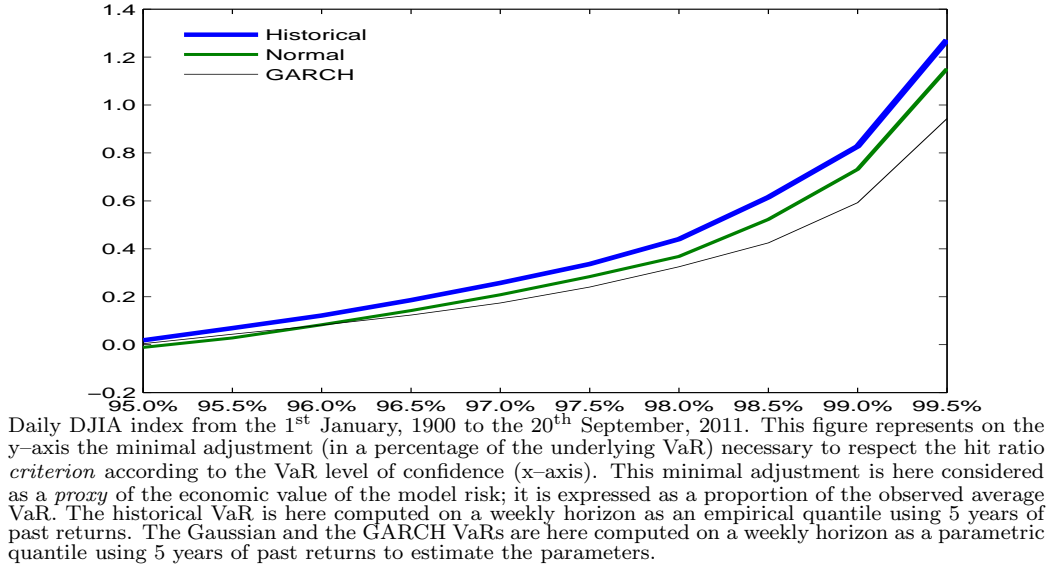
In other words, we show the minimal constant that should be added to the quantile estimation for reaching a VaR sequence that passes the hit test at all times (here with full information at time T). We can see that the corrections range from (almost) 0 to 140% and increase with the quantile. The comparison between the three methods favors the GARCH method, since the error is lower for all quantiles and the difference between methods (with

---

[7]On a theoretical basis, $A_T(\alpha)$ might, of course, be empty and q_T can be positive. However, as the sample gets large, these two situations are very unlikely since some negative errors might soon appear (please see Figure 4 below).

[8]We used a looped grid-search algorithm, adding successively a small increment on the top of the VaR (+.1% of the EVaR at each step of the loop), starting from the maximum positive value and increasing until the test is finally passed at a given probability threshold.

[9]A generalization of the basic procedure allows the risk manager for simple time–varying corrections, where the original sequence is modified as $\left\{ VaR_{t_1}(\hat{\theta}_1, \alpha) + q_1 : t_1 = [1, \cdots, T_1], \cdots, VaR_{t_k}(\hat{\theta}_k, \alpha) + q_k : t_k = [k, \cdots, T_1 + k - 1], \cdots \right\}$, and the optimization is done in all the arguments $(q_1, \cdots, q_k, \cdots)$ with the optimal adjustment at the end being the maximum of the sequence $(q_1, \cdots, q_k, \cdots)$.

Figure 3: Minimum model risk adjustment factor for the hit test associated with historical, Gaussian and GARCH VaRs on the DJIA, for a range of probabilities

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. This figure represents on the y–axis the minimal adjustment (in a percentage of the underlying VaR) necessary to respect the hit ratio *criterion* according to the VaR level of confidence (x–axis). This minimal adjustment is here considered as a *proxy* of the economic value of the model risk; it is expressed as a proportion of the observed average VaR. The historical VaR is here computed on a weekly horizon as an empirical quantile using 5 years of past returns. The Gaussian and the GARCH VaRs are here computed on a weekly horizon as a parametric quantile using 5 years of past returns to estimate the parameters.

full information on the total sample) is quite similar and rather independent of the confidence level.
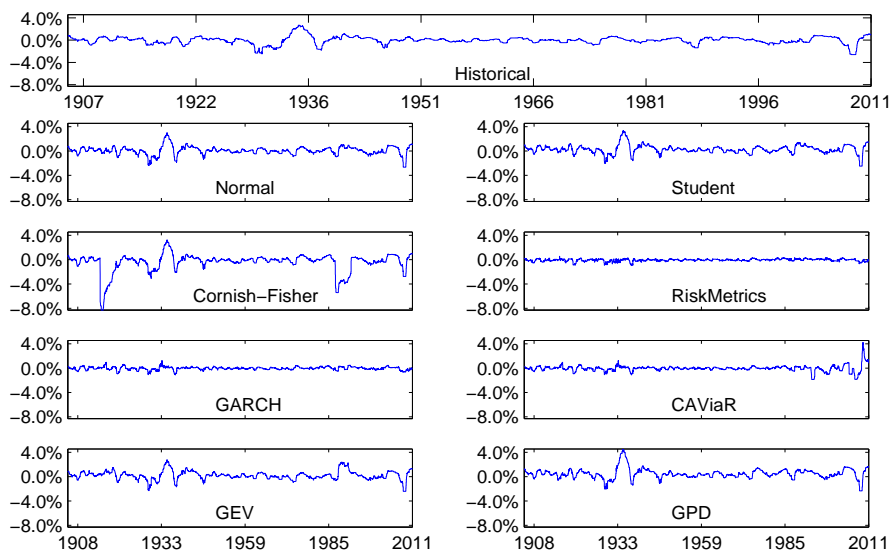
3.3 VaR model comparisons

We apply the general adjustment method presented above, obtained for the daily DJIA index from January 1st, 1900 until March 2nd, 2011 (29,002 daily returns). We use a moving window of four years (1,040 daily returns) to re–estimate parameters dynamically for the various methods. Forecasted VaR are computed dynamically for each method for the final 29,957 days (about 108 years). The out–of–sample exercise consists of a rolling forecast scheme with a window of four years (1,040 daily returns) to re–estimate parameters dynamically. Then, we use one year of out–of–sample daily forecasts to calibrate the correction based on the backtesting procedures. The backtesting experiment to correct the risk model of VaR estimates is then based on a ratio of the out–of–sample to in–sample size equal to .24, i.e. 250/1,040), which is sufficiently close to zero, as required for a valid out–of–sample exercise as shown by West (1996), McCracken (2000), Escanciano and Olmo (2010), and Escanciano and Pei (2012). This comparison considers daily estimation of the 95%, 99% and 99.5% conditional VaR.

18

This leaves the choice of the VaR forecast method. While there is a large number of techniques that could be used, we restrict ourselves to the most common in practice, in particular historical simulation and several parametric approaches based on Gaussian or Student–t return distributions, as well as the Cornish–Fisher VaR; see Cornish and Fisher, 1937; Favre and Galeano, 2002). We also employ three dynamic methods, EWMA, GARCH(1,1) and CAViaR (Engle and Manganelli, 2004). Finally, we complement these methods by using two extreme densities for the returns such as the GEV distribution and the GPD (see e.g., Engle and Manganelli, 2001).

Figure 4 shows the optimal adjustment factor for the various risk models for a 95% VaR estimated with the DJIA, in particular the daily correction factors that pass the hit test over the past year of daily returns (over the period from $t - 250$ to t). The magnitude can sometimes be large (specifically around the 1929 and 2008 crises), ranging from 0 to 15% EWMA or to more than 100% in some circumstances (for the Cornish–Fisher VaR). We also see that the most extreme VaR violations happened during the Great Depression for all measures. Dynamic measures, such as EWMA, GARCH and CAViaR, also demonstrate some superiority over unconditional parametric methodologies.
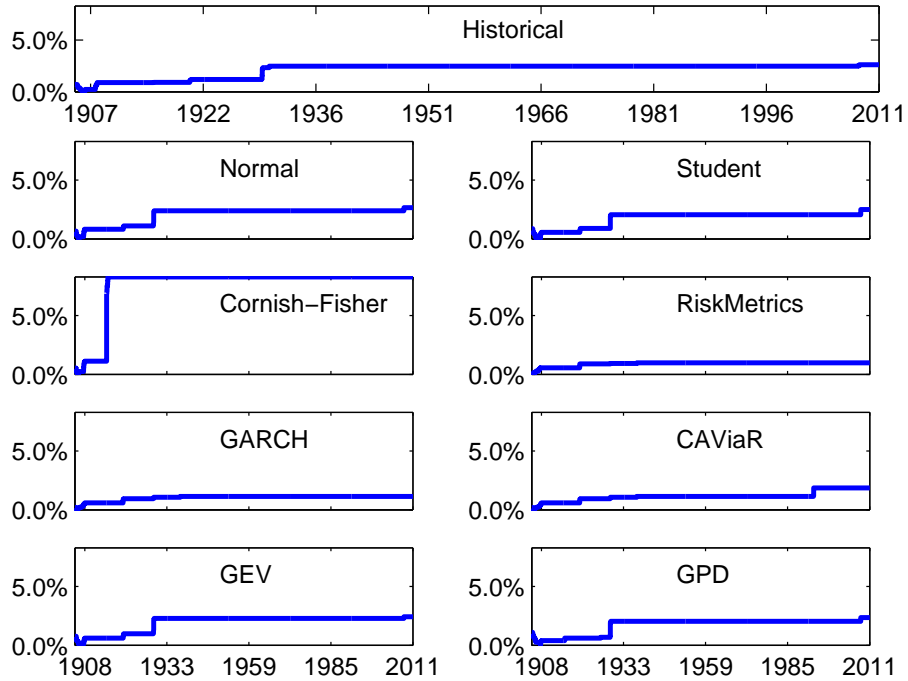
Figure 4: Dynamic optimal adjustment on the daily 95% VaR

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to re–estimate parameters dynamically for the various methods.

Figure 5 illustrates the evolution of the maximum required corrections for all VaR methods under consideration (maxima of the historical correction record needed from January 1st, 1900 to the current date t, which were already represented in Figure 4)[10]. These corrections are for the hit test, from the general program aiming to correct today's VaR with the historical maximum of the minimum correction that has been necessary since the beginning of the series (expressed here in relative terms compared to the level of VaR).

Figure 5: Optimal dynamic absolute value of minimum negative adjustments for the hit test for different methods and the 95% VaR

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to re–estimate parameters dynamically for the various methods.

Figure 6 illustrates the minimum dynamic adjustment required for passing the hit test for a randomly chosen first date of implementation. More precisely, the exercise consists of choosing a first date and then computing the dynamic adjustment until the end of the sample; repeating this exercise 30,000 times, whilst ultimately keeping, for each horizon, the minimum correction obtained. The optimal adjustments are here expressed in terms

[10]We did the same estimation and backtesting with a 10–year sample for VaR. We obtained the same qualitative results and saw that the choice of the size for VaR estimation is not crucial in our case. The results are available on demand.

of a percentage of their maximum value over the whole sample. For each horizon (x–axis in Figure 6), the correction (on the y–axis) thus corresponds to the worst case scenario, i.e. the smallest correction required in the various samples of the same horizon).

The figure shows that, depending on the VaR method, the time period length for having almost all of the maximum correction factors varies from 18 years (GEV) to 46 years (CAViaR). Moreover, regardless of the model, the major part (80% or so) of the correction factors is reached after 10 years. This means that, whatever the VaR model, most of the greatest surprises have been faced after a decade of history (even in the worst scenario when the sample is amongst the least turbulent ones). In other words, at least ten years are needed to have a fairly good idea of the magnitude of the required correction factors.

Figure 6: Optimal dynamic relative adjustment for the hit test for different starting dates and 95% VaR by horizon (in years)

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to dynamic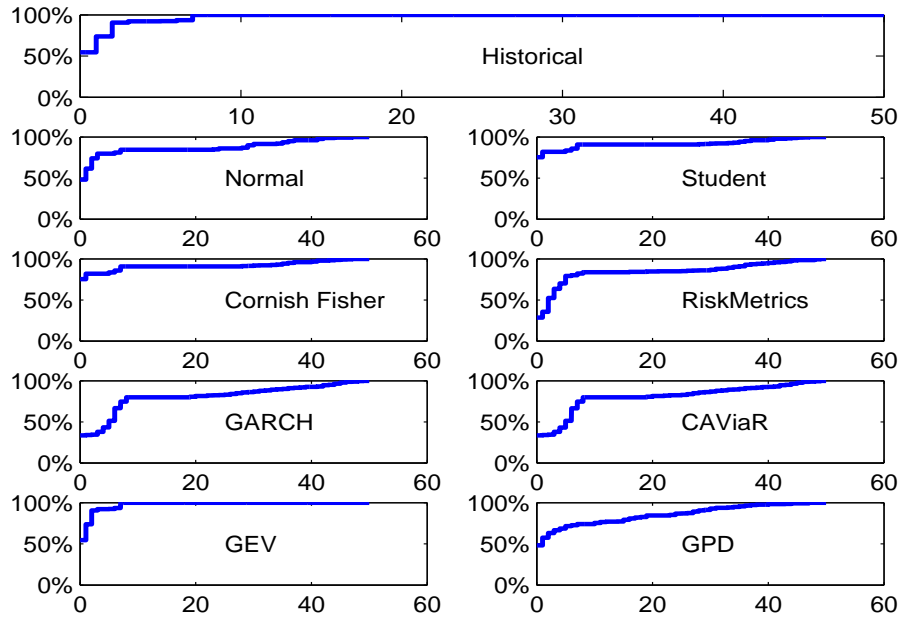ally re–estimate parameters for the various methods. This figure illustrates the dynamic negative adjustment required for passing the hit test (see Figure 4), having randomly chosen the first date of implementation. Optimal relative negative adjustments are here expressed in terms of percentage of their maximum value over the whole sample.

We next consider the three main qualities of VaR models as a generalization of the approach by Kerkhof et al. (2010). Table 3 reports the various minimum required corrections related to the three main categories of tests,

together with their Escanciano and Olmo (2009, 2010 and 2011) bootstrapped corrected versions. We first note that the hit test is less permissive when the bootstrapped critical values are used, whilst the tests of independence and magnitude impose very severe corrections (to the order of 100% in relative terms for some tests).

Table 3: Minimum model risk for 95% daily VaR models for various validity tests with a 5% confidence level

Method	Mean VaR	$\underline{q}_1$	$\underline{q}_1^*$	$\underline{q}_2$	$\underline{q}_2^*$	$\underline{q}_3$	$\underline{q}_3^*$
Historical	-1.60%	-2.61%	-2.03%	-4.85%	-3.24%	-3.10%	-5.90%
Normal	-1.68%	-2.66%	-1.86%	-4.62%	-2.76%	-2.76%	-5.49%
Student	-1.89%	-2.49%	-1.86%	-4.25%	-2.85%	-3.11%	-6.30%
CF	-1.26%	-8.29%	-7.48%	-8.40%	-8.86%	-8.40%	-8.86%
EWMA	-1.59%	-.98%	-.65%	-2.03%	-1.02%	-1.02%	-2.89%
GARCH	-1.61%	-1.13%	-.96%	-2.57%	-1.15%	-1.20%	-2.46%
CAViaR	-1.66%	-1.87%	-1.55%	-2.59%	-2.22%	-2.08%	-2.56%
GEV	-1.84%	-2.42%	-1.99%	-4.47%	-2.99%	-2.80%	-6.97%
GPD	-2.11%	-2.35%	-1.67%	-4.43%	-2.63%	-2.71%	-6.51%

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to dynamically re–estimate parameters for the various methods. The variable $\underline{q}_1$ refers to the hit test; $\underline{q}_2$ to the independence test; $\underline{q}_3$ to the magnitude test; and $\underline{q}_1^*$, $\underline{q}_2^*$, $\underline{q}_3^*$ correspond to their resampling versions, following Escanciano and Olmo (2009, 2010 and 2011).

According to the the unconditional coverage test at a 5% level, EWMA is the best model for estimating the DJIA index 95% VaR, followed by GARCH and then GEV. The independence test favours the conditional methods, with the best result for the GARCH model. Finally, when considering the magnitude of the violations — the most severe test — once again the dynamic measures show some superiority, whilst the extreme density VaR exhibits weakness.

3.4   Generalized model risk of model risk

Finally, we compare our method with classical stress–test exercises. We first present the extent to which the required calibrated correction factors can provide an insurance against major historical financial crises. Then, we compare the correction factors implied by the various backtests to correct the model risk of risk models, to a typical stress–test scenario.

Three implicit levels of confidence are required: the probability level of the VaR under consideration, the thresholds in the various tests applied for computing the required correction and, finally, the degree of confidence we want to put on the solidity of the buffer. Typically, a high probability VaR focus will increase the model risk, whilst a more severe test level leads to a lower risk. Consequently, a high incremental buffer leads to a high protection against the model risk that is realized during extreme events on the market. By contrast, a reduced buffer decreases the insurance against these major turbulent episodes and, then, ultimately increases failures of (corrected) risk models.
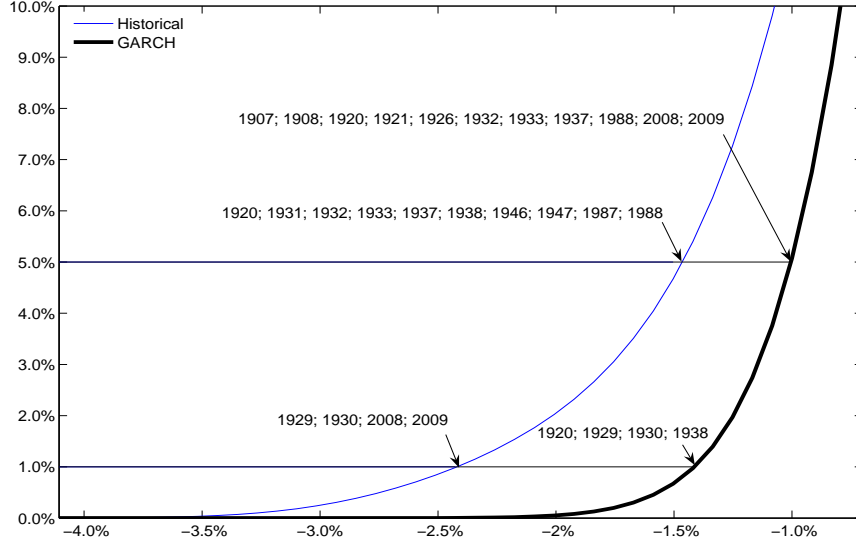
Figure 7 below illustrates this link between the level of the buffer, here translated into protection against the more severe historical crises, and the degree of confidence associated to the buffer. The Figure represents the cumulative density functions of required adjustments (in the last century of the DJIA) for, respectively, the historical and GARCH(1,1) VaR at a 95% confidence level, with a threshold for the hit test fixed at 5%. The series of dates stand for years corresponding to the largest exceptions for the two VaR methods for certain levels of confidence (on the y–axis) and related corrections (on the x–axis). We note here that the GARCH VaR leads to smaller corrections in general. We also see that if we accept a 5% model risk, we are, unsurprisingly, not protected anymore against the 5% biggest shocks in the data (such as, for instance, those of 1929, 1930, 2008 and 2009 for the historical method).

We then compare the correction applied to assess the robustness of risk estimates with the correction implied by a typical stress test exercise for usual portfolio profiles by imposing handpicked shocks for each investment class. We provide these comparisons in terms of factor k used by regulators for determining capital (k being between 3 and 5).

Thus, we first present in Table 4 (Panel A and B) the various (model risk free) minimum corrections corresponding to the three tests (frequency, independence and magnitude) at a 5% confidence level for a 95% GARCH VaR applied to financial series of daily return on indexes and profiled portfolios on the period from December 31st, 1986 to November 28th, 2011. We consider four asset classes as well as three investment profiles combining these asset classes (defensive, balanced and aggressive portfolios).[11]

---

[11] For the bonds, we use the "Merrill Lynch U.S. Treasuries/Agencies-Master AAA" index before 01/01/1998 and the "J.P. Morgan EMU Global Aggregate Bond AAA All Maturities" after; for the equity class we use a composite index "95% MSCI Europe Index + 5% MSCI World Index"; for the real estate class we get the "European Real Estate Investment and Services Index" and for commodity, the "CRB Spot Index".

Figure 7: The empirical cumulative density function of optimal adjustment values for the hit test of a 95% daily historical and GARCH VaR

Daily DJIA index from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) for computing the VaR. The threshold for the hit test is here fixed at 5% and we use a Gaussian kernel smoothing density (see Bowman and Azzalini, 1997).

We express the outcomes as a percentage of VaR in Table 4 (Panel A), whilst presenting them as k ratios of corrected VaR out of estimated VaR in Panel B of Table 4. The correction factors in Panel A of Table 4 for single indexes range from -3.65% (for $\underline{q}_3$ — magnitude correction for the commodity index) to -63.83% (for $\underline{q}_3^*$ — bootstrapped magnitude correction for the real estate index). For the various profiles, we see that the correction factor is lower than 1% for the defensive profile and goes to 10% or so for the aggressive one (and to -35.05% when considering the most severe test of magnitude). When these correction factors are expressed in terms of k ratios in Panel B of Table 4, they range from 1.01 to 3.66 which is in line with the official k ratio between 3 and 5.

We can now compare the correction factors, calibrated based on our framework, with a standard stress–test approach supposing some typical shocks on various asset classes. As underlined by Breuer and Csiszár (2012), stress tests with hand–picked scenarios are subject to two significant criticisms. First, arbitrary severe scenarios may be too implausible. Second, some other stress scenarios leave open the question of whether there are more severe scenarios of similar plausibility. If the considered scenarios are harmless, either because stress testers lack proficiency or wish to hide risks, stress tests convey a feel-

24

ing of safety which might be false. If they are merely unrealistic, they lead falsely to excessively high capital. Our proposed strategy can help to gauge the severity (and plausibility) of an ad hoc handpicked specific scenario.

Focusing indeed on the k ratios, Panel C of Table 4 reports the implied corrections on annual 95% GARCH VaR in the case of a hypothetical stress. With the given intensity of shocks considered here[12] (-30% for the equity index, -40% for the real estate, -30% for commodity and -20% for bonds over a one–year horizon), k ratios vary from 1.90 (for $\underline{q}_2$ - independence correction for the equity index) to 4.99 (for $\underline{q}_3$ — magnitude test for the real estate index) for the single indexes, and from 1.54 (for $\underline{q}_2$ - independence correction for the balanced profile) to 6.10 (for $\underline{q}_3$ — magnitude test for the aggressive portfolio).

If we now compare the results in Panel C of Table 4 (ad hoc stress tests) to those in Panel B of Table 4 (calibrated empirical corrections), the arbitrary implied corrections of the stress test scenarios appear to be far more severe for almost all indexes and portfolios (except for the balanced one and the independence test). We thus conclude that this illustrative stress–test is very conservative. In other words, because k ratios are almost higher in Panel C of Table 4 than in Panel B of Table 4 (on average by 80%), this stress–test seems to be relatively robust to the impact of model risk for the risky assets.

---

[12]The amplitude of the shocks is directly inspired from recommendations of the Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS).

Table 4: Minimum model risk for a 95% GARCH–VaR, k ratio model risk confidence levels for a 95% GARCH–VaR and 95% stress–VaR for 5% validity tests on various portfolios

Portfolio	$\underline{q}_1$	$\underline{q}_1^*$	$\underline{q}_2$	$\underline{q}_2^*$	$\underline{q}_3$	$\underline{q}_3^*$
Panel A. Minimum annualized model risk for a 95% GARCH-VaR						
Equity	-10.15%	-7.14%	-9.86%	-15.12%	-44.80%	-16.44%
Real estate	-12.65%	-10.32%	-16.53%	-18.93%	-63.83%	-25.03%
Commodity	-6.39%	-6.25%	-5.29%	-6.99%	-13.76%	-3.65%
Bond	-9.89%	-9.62%	-10.27%	-10.54%	-18.44%	-13.62%
Defensive profile	-.08%	-.08%	.00%	-.21%	-1.04%	-.26%
Balanced profile	-4.63%	-4.36%	-5.88%	-6.52%	-15.79%	-8.74%
Aggressive profile	-9.28%	-8.38%	-8.52%	-11.62%	-35.05%	-12.72%
Panel B. Minimum k ratio model risk confidence levels for a 95% GARCH–VaR						
Equity	1.35	1.25	1.34	1.53	2.56	1.57
Real estate	1.40	1.33	1.53	1.60	3.03	1.80
Commodity	1.65	1.64	1.54	1.72	2.41	1.37
Bond	2.43	2.39	2.48	2.52	3.66	2.97
Defensive profile	1.15	1.15	1.01	1.40	3.00	1.50
Balanced profile	1.45	1.42	1.57	1.63	2.52	1.84
Aggressive profile	1.42	1.38	1.38	1.52	2.58	1.57
Panel C. Minimum k ratio model risk confidence levels of 95% stress–VaR						
Equity	2.54	2.25	1.90	2.71	4.81	3.29
Real estate	3.11	2.84	3.18	3.80	4.99	4.84
Commodity	2.89	2.89	3.27	2.92	3.83	3.19
Bond	3.51	3.50	3.52	3.56	4.04	3.76
Defensive profile	2.11	2.11	2.11	2.12	2.20	2.16
Balanced profile	2.63	2.50	1.54	2.63	5.10	3.81
Aggressive profile	3.08	2.94	1.83	3.78	6.10	4.94

Datasource: DataStream and Bloomberg. Daily data from the 31st December, 1986 to the 28th November, 2011; computations by the authors. The asset classes as detailed in Footnote 11. A moving window of four years (1,040 daily returns) is used to re-estimate parameters dynamically for the various methods. "Defensive Profile" corresponds to a mixed portfolio compound with 10% bond +90% Liquidity; "Balanced Profile" 30% equity+10% Real Estate +10% commodity + 40% bond + 10% liquidity; and "Aggressive Profile" 70% equity + 15% real estate + 15% commodity. The variable $\underline{q}_1$ refers to the hit test; $\underline{q}_2$ to the independence test; $\underline{q}_3$ to the magnitude test; and $\underline{q}_1^*$, $\underline{q}_2^*$, $\underline{q}_3^*$ correspond to their resampling versions, following Escanciano and Olmo (2009, 2010 and 2011). Panel A gives the minimum annualized corrections for backtest at 5% confidence level on a 95% GARCH–VaR, Panel B the minimum k–ratio (adjustment/VaR)for a 95% GARCH–VaR and Panel C the minimum k ratio model in the stress–VaR context for 5% validity tests. The following shocks are considered for Panel C: -30% for the Equity index, -40% for the real estate, -30% for commodity and -20% for bonds over a one–year horizon.

Taken altogether, our results suggest that some VaR models are preferred (e.g. the dynamic approaches such as the EWMA, CAViaR and GARCH models), whilst others should be avoided (e.g. the Cornish–Fisher VaR or extreme distribution based VaR) when comparing the minimum correction to pass the frequency/hit test. Moreover, the independence and the magnitude tests lead to more severe corrections on the estimated VaR than the frequency test does. But whatever the model, the magnitude of the correction factors can be sometimes exceptionally large, especially during major financial crisis episodes such as the Great Depression of 1929 or the crisis of 2008. This is why there is a direct link between the confidence level on the required ex post correction (on the full historical sample), and the insurance against these major historical financial turmoils. However, we also show that a 10 year sample of observations for calibrating the minimum correction to be added, is sufficient to have a fairly good idea of the magnitude of the model risk of risk models.

4   Conclusion

Standard risk measures failed to forecast extreme risks and regulators require that financial institutions quantify this model risk of risk models. We propose to adjust risk forecasts for model risk by the historical performance of the model. In other words, the risk model learns from its past mistakes.

We first examine standard risk models by assessing how well they forecast risk from a simulated process, designed to realistically capture the salient features of financial returns. The experiment shows that model risk is significant and ever present, in some cases, so large that it exceeds the actual risk forecast.

In our main contribution, we then propose a methodology for explicitly incorporating model risk corrections into risk forecasting by taking into account the models' performance on a range of standard back testing methodologies.

The general setup also enables us to evaluate the performance of standard risk forecast models, by applying the basic principle that the lower the model risk correction factor, the lower the model risk and, therefore, the better the model. The results show that dynamic methods, such as EWMA, CAViaR and GARCH VaR, have an advantage over static approaches such as Gaussian and extreme density approaches. Somewhat surprisingly, the very simple historical simulation approach is, if not the best method, close to the best.

We conclude by proposing an approach that provides a tailored methodology for risk managers where they can explicitly relate the degrees of confidence

in the correction factor to the distribution of past violations. In this, the manager addresses three concerns: the VaR probability, the severity of tests and the trust we want to put into the correction buffer. This can, for example, enable a risk manager to explicitly consider extreme events, such as 1929 and 2008, or alternatively disregard their impact on risk forecasts.

The Basel Committee has recently proposed (BCBS, 2013) the use of a stressed risk forecast as the main input into the current risk forecast. Such an approach is an improvement over the existing methodology, and is partially consistent with our methodology. The Committee indeed proposes to rescale the risk forecasts by the ratio of the stressed and unstressed risk factors, such as the adjusted current risk forecast becomes more conservative and thus less prone to exceptions. However, our proposal deals with this in a more precise way. First, we adjust risk forecasts by their past errors, which mainly come from these distressed periods. Second, we consider a confidence level about the required correction factor linked the insurance against major financial stress episodes. Finally, we define proper criteria for adjusting the risk forecasts based on some properties of forecast errors such as their frequency, their independence and their magnitude. In our view, the Basel Committee proposal still ignores the model risk of risk forecasts and consists of an adjustment of the current risk without an explicit criterion.

Our work can be extended in several ways. Our general correction framework can be used when comparing the various tests of a desirable VaR proposed in the literature (Berkowitz et al., 2011). The second extension could be to apply some specific VaR models when judging the riskiness of some non–linear products using, this time, several pricing models. In the same vein, evaluating the impact on asset allocation of integrating the model risk of risk measures could be of interest, especially for asset allocation paradigms depending on risk budgets, e.g. safety first criteria. The third extension could be found in generalizing the comparison considering several time–horizons (e.g. Cheridito and Stadje, 2009; Hoogerheide et al., 2011) or several quantile levels (Colletaz et al., 2013). The fourth extension is about alternative backtests when calibrating our model risk correction (see appendix for a list of tests), in particular the D-test of Escanciano and Pei (2012) that combines the nonparametric weighted backtest with the test of independence of Christoffersen (1998) and offers good finite-sample size and power properties. Another approach would to adopt the same methodology leading to an estimated multi–VaR, built as a portfolio of various VaR models (see Abdous and Remillard, 1995), directly aiming to minimize the model risk (McAleer et al., 2013). Finally, using the same metric of corrections, the quality of other VaR based measures in a context of systemic risk measures (such as

Marginal Expected Shortfall or CoVaR) would be worth considering (e.g. Danielsson et al., 2011; Benoit et al., 2013; Löffler and Raupach, 2013).

5 Bibliography

Abdous B. and B. Remillard, (1995), "Relating Quantiles and Expectiles under Weighted-Symmetry", *Annals of the Institute of Statistical Mathematics 47(2)*, 371-384.

Aït-Sahalia Y., J. Cacho-Diaz and R. Laeven, (2013), "Modeling Financial Contagion using Mutually Exciting Jump Processes", *Princeton Working Paper*, 44 pages.

Angelidis T. and S. Degiannakis, (2007), "Backtesting VaR Models: a Two-Stage Procedure", *Journal of Risk Model Validation 1(2)*, 27-48.

Alexander C. and J.M. Sarabia, (2012), "Quantile Uncertainty and Value-at-Risk Model Risk", *Risk Analysis 32(8)*, 1293-1308.

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

Basel Committee on Banking Supervision, (2009), *Revisions to the Basel II Market Risk Framework*, Bank for International Settlements, 35 pages.

Basel Committee on Banking Supervision, (2010), *Basel III: Towards a Safer Financial System*, Bank for International Settlements, 8 pages.

Basle Committee on Banking Supervision, (2013), *Fundamental Review of the Trading Book: A Revised Market Risk Framework*, Bank for International Settlements, 127 pages.

Bauwens L., A. Preminger and J. Rombouts, (2010), "Theory and Inference for a Markov Switching GARCH Model", *Journal of Econometrics 13(2)*, 218-244.

Beder T., (1995), "VaR: Seductive but Dangerous", *Financial Analysts Journal 51(5)*, 12-24.

Benoit S., G. Colletaz, Ch. Hurlin and Ch. Perignon, (2013), "A Theoretical and Empirical Comparison of Systemic Risk Measures", *SSRN Working Paper*, 42 pages.

Berkowitz J., (2001), "Testing Density Forecasts with Applications to Risk Management", *Journal of Business and Economics Statistics 19(4)*, 465-474.

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

Berkowitz J., P. Christoffersen and D. Pelletier, (2011), "Evaluating Value-at-Risk Models with Desk-Level Data", *Management Science 57(12)*, 2213-2227.

Billio M., R. Casarin and A. Osuntuyi, (2012), "Efficient Gibbs Sampling for Markov Switching GARCH Models", *Working Paper 35*, University of Ca'Foscari, 40 pages.

Bowman A. and A. Azzalini, (1997), *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, 204 pages.

Breuer T. and I. Csiszár, (2012), "Measuring Model Risk", mimeo, 19 pages.

Breuer T. and I. Csiszár, (2013), "Model Risk Plausibility Constraints", *Journal of Banking and Finance 37(5)*, 1552-1559.

Breuer T., M. Jandacka, J. Mencía and M. Summer, (2012), "A Systematic Approach to Multi-period Stress Testing of Portfolio Credit Risk", *Journal of Banking and Finance 36(2)*, 332-340.

Campbell S., (2007), "A Review of Backtesting and Backtesting Procedures", *Journal of Risk 9(2)*, 1-19.

Candelon B., G. Colletaz , C. Hurlin and S. Tokpavi, (2011), "Backtesting Value-at-Risk: a GMM duration-based test", *Journal of Financial Econometrics 9(2)*, 314-343.

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

Chan N., M. Getmansky, S. Haas and A. Lo, (2006), *Systemic Risk and Hedge Funds* in *The Risks of Financial Institutions*, Carey-Stulz (Eds), 235-338.

Cheridito P. and M. Stadje, (2009), "Time-inconsistency of VaR and Time-consistent Alternatives", *Finance Research Letters 6(1)*, 40-46.

Christoffersen P., (1998), "Evaluating Interval Forecasts", *International Economic Review 39(4)*, 841-862.

Christoffersen P. and S. Gonçalves, (2005), "Estimation Risk in Financial Risk Management", *Journal of Risk 7(3)*, 1-28.

Christoffersen P. and D. Pelletier, (2004), "Backtesting Value-at-Risk: A Duration-Based Approach", *Journal of Financial Econometrics 2*, 84-108.

Colletaz G., Ch. Hurlin and Ch. Pérignon, (2013), "The Risk Map: A New Tool for Validating Risk Models", *Journal of Banking and Finance 37(10)*, 3843-3854.

Chordia T., R. Roll and A. Subrahmanyam, (2001), "Market Liquidity and

Trading Activity", *Journal of Finance 56(2)*, 501-530.

Cornish E. and R. Fisher, (1937), "Moments and Cumulants in the Specification of Distributions", *Review of the International Statistical Institute 5(4)*, 307-320.

Crouhy M., D. Galai and R. Mark, (1998), "Model Risk", *Journal of Financial Engineering 7*, 267-288.

Daníelsson J., (2002), "The Emperor has No Clothes: Limits to Risk Modelling", *Journal of Banking and Finance 26(4)*, 1273-1296.

Daníelsson J., K. James, M. Valenzuela and I. Zer, (2011), "Model Risk of Systemic Risk Models", *mimeo*, 26 pages.

Derman E., (1996), "Model Risk", *Risk 9(5)*, 34-37.

Dumitrescu E., Ch. Hurlin and V. Pham, (2012), "Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests", *Finance 33(1)*, 79-112.

Engle R. and S. Manganelli, (2001), "Value-at-Risk Models in Finance", *ECB Working Paper 75*, 40 pages.

Engle R. and S. Manganelli, (2004), "CAViaR: Conditional AutoRegressive Value-at-Risk by Regression Quantile", *Journal of Business and Economic Statistics 22(4)*, 367-381.

Escanciano J. and P. Pei, (2012), "Pitfalls in Backtesting Historical Simulation Models", *Journal of Banking and Finance 36(8)*, 2233-2244.

Escanciano J. and J. Olmo, (2009), "Specification Tests in Parametric Value-at-Risk Models", in *Financial Risks*, Gouriéroux - Jeanblanc (Eds), Economica, 49-62.

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

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

Favre L. and J. Galeano, (2002), "Mean Modified Value-at-Risk Optimization with Hedge Funds", *The Journal of Alternative Investments 5(2)*, 21-25.

Figlewski S., (2004), "Estimation Error in the Assessment of Financial Risk Exposure", *Working Paper,* New-York University, 48 pages.

Financial Services Authority, (2006), "Solvency II: A New Framework for Prudential Regulation of Insurance in the EU", *FSA Discussion Paper*, 1-64.

Frésard L., Ch. Pérignon and A. Wilhelmsson, (2011), "The Pernicious Ef-

fects of Contaminated Data in Risk Management", *Journal of Banking and Finance 35(10),* 2569-2583.

Gagliardini P. and Ch. Gouriéroux, (2013), "Granularity Adjustment for VaR Risk Measures: Systematic vs Unsystematic Risks", *International Journal of Approximate Reasoning 54(6),* 717-747.

Gagliardini P., Ch. Gouriéroux and A. Monfort, (2012), "Micro-information, Non-linear Filtering and Granularity", *Journal of Financial Econometrics 10(1),* 1-53.

Getmansky M., A. Lo and I. Makarov, (2004), "An Econometric Analysis of Serial Correlation and Illiquidity in Hedge-Fund Returns", *Journal of Financial Economics 74(3),* 529-609.

Gibson R. (Editor), (2000), *Model Risk: Concepts, Calibration and Pricing,* Risk Books.

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

Giot P. and J. Grammig, (2005), "How Large is Liquidity Risk in an Automated Auction Market?", *Empirical Economics 30(4),* 867-888.

Gordy M., (2003), "A Risk-Factor Model Foundation for Rating-based Bank Capital Rules", *Journal of Financial Intermediation 12(1),* 199-232.

Gouriéroux C. and J.-M. Zakoïan, (2013), "Estimation-Adjusted VaR", *Econometric Theory 29(4),* 735-770.

Gray S., (1996), "Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process", *Journal of Financial Economics 42(1),* 27-62.

Haas M., (2005), "Improved Duration-Based Backtesting of Value-at-Risk", *Journal of Risk 8(2),* 17-36.

Haas M., S. Mittnik and M. Paolella, (2004), "A New Approach to Markov-Switching GARCH Models", *Journal of Financial Econometrics 2(4),* 493-530.

Hamilton J. and R. Susmel, (1994), "AutoRegressive Conditional Heteroskedasticity and Changes in Regime", *Journal of Econometrics 64(1-2),* 307-333.

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

Hoogerheide L., F. Ravazzolo and H. van Dijk, (2011), "Backtesting Value-at-Risk using Forecasts for Multiple Horizons, A Comment on the Forecast Rationality Tests of A.J. Patton and A. Timmermann", *Tinbergen Institute*

Discussion Paper TI 2011-131/4, 15 pages.

Hurlin C. and S. Tokpavi, (2006), "Backtesting Value-at-Risk Accuracy: A Simple New Test", *Journal of Risk 9(2)*, 19-37.

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

Jorion P., (2007), *Value-at-Risk: The New Benchmark for Managing Financial Risk*, McGraw-Hill, 600 pages.

Jorion P., (2009-a), *Financial Risk Manager Handbook*, Wiley, 717 pages.

Jorion P., (2009-b), "Risk Management Lessons from the Credit Crisis", *European Financial Management 15(5)*, 923-933.

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

Klaassen F., (2002), "Improving GARCH Volatility Forecasts with Regime-Switching GARCH", *Empirical Economics 27(2)*, 363-394.

Kupiec P., (1995), "Techniques for verifying the Accuracy of Risk Measurement Models", *Journal of Derivatives 3(2)*, 73-84.

Kyle A., (1985), "Continuous Auctions and Insider Trading", *Econometrica 53(6)*, 1315-1336.

Löffler G. and P. Raupach, (2013), "Robustness and Informativeness of Systemic Risk Measures", *Deutsche Bundesbank Discution Paper 4*, 40 pages.

Lönnbark C., (2010), "Uncertainty of Multiple Period Risk Measures", *Umea Economic Studies 768*, 37 pages.

Lopez J., (1998), "Methods for Evaluating Value-at-Risk Estimates", *Federal Reserve Bank of San Francisco 98-02*, 3-17.

Lopez J., (1999), "Regulatory Evaluation of Value-at-Risk Models", *Journal of Risk 1(2)*, 37-64.

McAleer M., J.-A. Jiménez-Martín and T. Pérez-Amaral, (2013), "International Evidence on GFC-robust Forecast for Risk management under the Basel Accord", *Journal of Forecasting 32(3)*, 267-288.

McCracken M.W., (2000), "Robust Out-of-sample Inference", *Journal of Econometrics 99(5)*, 195-223.

Nieto M. and E. Ruiz, (2008), "Measuring Financial Risk: Comparison of Alternative Procedures to estimate VaR and ES", *Statistics and Econometrics Working Paper - University Carlos III*, 45 pages.

Pérignon Ch. and D. Smith, (2008), "A New Approach to Comparing VaR

Estimation Methods", *Journal of Derivatives 16(2)*, 54-66.

Pérignon Ch. and D. Smith, (2010), "The Level and Quality of Value-at-Risk Disclosure by Commercial Banks", *Journal of Banking and Finance 34(2)*, 362-377.

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

RiskMetrics, (1996), *Technical Document, Morgan Guaranty Trust Company of New York*, 296 pages.

Silber W., (2005), "What Happened to Liquidity when World War I Shut the NYSE?", *Journal of Financial Economics 78(3)*, 685-701.

Talay D. and Z. Zheng, (2002), "Worst Case Model Risk Management", *Finance and Stochastics 6(4)*, 517-537.

West K., (1996), "Asymptotic Inference about Predictive Ability", *Econometrica 64*, 1067-1084.

Wilde T., (2001), "Probing Granularity", *Risk 18(4)*, 103-106.

Wong W., (2008), "Backtesting Trading Risk of Commercial Banks using Expected Shortfall", *Journal of Banking and Finance 32(7)*, 1404-1415.

Wong W., (2010), "Backtesting Value-at-Risk based on Tail Losses", *Journal of Empirical Finance 17(3)*, 526-538.

A   Model risk when forecasting risk

Financial risk forecast models, just like any other statistical model, are thus subject to model risk. In spite of this, almost all presentations of risk forecasts focus on point estimates, omitting any mention of model risk, not even mentioning estimation risk. They are, however, subject to the same basic elements of model risk as any other model, but are also subject to unique model risk factors because of the specific application.

In order to formally identify the model risk factors, we propose a five level classification scheme:

1. **Parameter estimation error** arises from uncertainty in the parameter values of the chosen model;

2. **Specification error** refers to the model risk stemming from inappropriate assumptions about the form of the data generating process (DGP) for the random variable;

3. **Granularity error** is based on the impact of undiversified idiosyncratic risk on the portfolio VaR;

4. **Measurement error** relates to the use of erroneous data when measuring the risks and testing the models;

5. **Liquidity risk** is defined as the consequence of both infrequent quotes and the inability to conduct sometimes a transaction at current market prices because of the relatively too large size of the transaction.

The ultimate objective is to forecast VaR, where we indicate the estimate by "estimated VaR" (denoted EVaR). It is a function of the portfolio size and the true model parameters θ_0. In what follows, VaR is the $(1-\alpha)^{\text{th}}$ quantile (with $\alpha > .50$) of the profit and loss distribution, so that the VaR is negative (and expressed hereafter as a return for the sake of simplicity). We also indicate the theoretical (or true) VaR by $\text{ThVaR}(\theta_0, \alpha)$. Thus, when comparing the estimated VaR with the theoretical VaR (*i.e.* EVaR and ThVaR respectively), we present both the buffer needed to directly adjust the EVaR and the probability (or quantile) shift required. Our objective is to approximate the errors or "biases" of VaR estimates since we do not know the "true" DGP with real data. Biases defined hereafter are "errors" (that can be repeated) that come mainly from the use of a wrong model and/or the wrong specification regarding the "true" (assumed DGP). Our proposed

procedure consists of approximating these errors, based on the minimum correction needed not to reject a predefined consensual backtest. In the following sub-sections, we detail these specific model risks that impact VaR forecasts and provide some examples.

Estimation risk

Estimation risk occurs in every estimation process. Relatively small changes in the estimation procedure or in the number of data observations can change the magnitude and even the sign of some important decision variables. Thus, estimation risk is the risk associated with an inaccurate estimation of parameters, due to the estimator quality and/or limited sample of data (past and/or future), and/or noise in the data.

If PEAVaR denotes the perfect estimation adjusted VaR, $\text{EVaR}(\hat{\theta}, \alpha)$ the estimated VaR and $\text{bias}(\hat{\theta}, \theta_0, \alpha)$ the bias function, where $\hat{\theta}$ are the estimated parameters, we have:

$$\text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{EVaR}(\hat{\theta}, \alpha) + \text{bias}(\hat{\theta}, \theta_0, \alpha). \tag{15}$$

Example 1 *As an illustration, assuming an ARCH model, the estimation risk (denoted herein $ER(\cdot)$) is expressed in Gouriéroux and Zakoïan (2013), as (with the previous notations):*

$$EVaR(\hat{\theta}, \alpha) = ThVaR(\theta_0, \alpha) + ER(ThVaR(\theta_0, \alpha), \hat{\theta}, \alpha),$$

with:

$$ER\left[ThVaR(\theta_0, \alpha), \hat{\theta}, \alpha\right] = -(2T)^{-1} h\left[ThVaR(\theta_0, \alpha), \hat{\theta}, \alpha\right] + o\left(T^{-1}\right),$$

where T is the length of the estimation period, $o\left(T^{-1}\right)$ converges to a term of order T^{-1} and:

$$
\begin{aligned}
h\left[ThVaR(\theta_0, \alpha), \hat{\theta}, \alpha\right] &= \left\{\left\{\frac{\partial^2 g^{-1}}{\partial r^2}(r_{t-1}, \theta, r) - ThVaR(\theta_0, \alpha)\left[\frac{\partial g^{-1}}{\partial r}(r_{t-1}, \theta, r)\right]^2\right\}\right\} \\
&\times \frac{\partial g}{\partial \theta'}[r_{t-1}, \theta, ThVaR(\theta_0, \alpha)]\ \Omega(\theta_0)\ \frac{\partial g}{\partial \theta'}[r_{t-1}, \theta, ThVaR(\theta_0, \alpha)] \\
&+ \frac{\partial g^{-1}}{\partial r}(r_{t-1}, \theta, r)\ Tr\left\{\Omega(\theta_0)\ \frac{\partial^2 g}{\partial \theta \partial \theta'}[r_{t-1}, \theta, ThVaR(\theta_0, \alpha)]\right\},
\end{aligned}
$$

and $r = g\left[r_{t-1}, \theta, ThVaR(\theta_0, \alpha)\right]$, $\Omega(\theta)$ the variance-covariance of parameters in θ, $g\left(.\right)$ a continuous function, strictly increasing with respect to the VaR parameter and $g^{-1}\left(.\right)$ its inverse.

Specification risk

Specification error arises from using inappropriate assumptions about the form of the DGP. We propose denoting the *strong form* of specification risk as the risk from using a risk model which cannot capture the true unknown DGP. The *weak form* of specification risk then corresponds to the risk of using a risk model inadequate with the assumed, and hence known, DGP.

Consider the special case of knowing the true model parameters, but not knowing the model. In this case, we can define the perfect specification adjusted VaR (PSAVaR) as:

$$\text{PSAVaR}(\theta_0, \theta_1, \alpha) = \text{EVaR}(\theta_1, \alpha) + \text{bias}(\theta_0, \theta_1, \alpha), \tag{16}$$

where θ_1 are known parameters, defined so that we can link the misspecified model to the true model, with some mapping $\theta_0 = f(\theta_1)$.

Example 2 *A simple measure of the specification risk (denoted as $SR(\cdot)$) associated to the expansion of the unknown "true" theoretical model of VaR (denoted $ThVaR(\theta, \alpha)$), can be written as:*

$$EVaR(\hat{\theta}, \alpha) = ThVaR(\theta, \alpha) + SR\left[ThVaR(\theta, \alpha), \hat{\theta}, \alpha\right],$$

with:

$$
\begin{aligned}
SR\left[ThVaR(\theta, \alpha), \hat{\theta}, \alpha\right] &= \frac{\sigma}{6}\left\{\left[\frac{AVaR(\hat{\theta}, \alpha) - \mu}{\sigma}\right]^2 - 1\right\} Sk \\
&+ \frac{\sigma}{24}\left\{\left[\frac{AVaR(\hat{\theta}, \alpha) - \mu}{\sigma}\right]^3 - 3\left[\frac{AVaR(\hat{\theta}, \alpha) - \mu}{\sigma}\right]\right\} Ku \\
&- \frac{\sigma}{36}\left\{2\left[\frac{AVaR(\hat{\theta}, \alpha) - \mu}{\sigma}\right]^3 - 5\left[\frac{AVaR(\hat{\theta}, \alpha) - \mu}{\sigma}\right]\right\} Sk^2 \\
&+ o(T^{-1}),
\end{aligned}
$$

where $ThVaR(\theta, \alpha)$ is the "true" theoretical model of VaR, $AVaR(\hat{\theta}, \alpha)$ is the asymptotic α-quantile of the approximate model in use, $SR(\cdot)$ is the specification error associated to this specific model and parameters μ, σ, Sk and Ku stand, respectively, for the mean, the standard deviation, the skewness and the kurtosis of the return distribution.

Granularity error

Granularity error is caused by the bias resulting from a finite number of assets in portfolios and then by the resulting residual idiosyncratic risk, e.g. see Gordy (2003), Wilde (2001). The granularity principle yields a decomposition of such risk measures that highlights the different effects of systematic and non-systematic risks.

More precisely, any portfolio risk measure can be decomposed into the sum of an asymptotic risk measure corresponding to an infinite portfolio size and $1/n$ times an adjustment term where n is the portfolio size (number of assets). The asymptotic portfolio risk measure, called the cross-sectional asymptotic risk measure, captures the non-diversifiable effect of risks on the portfolio. The adjustment term, called granularity adjustment, summarizes the effect of the individual specific risks and their cross-effect with systematic risks, when the portfolio size is large, but finite.

Suppose the theoretical VaR is based on an asymptotic factorial model, valid asymptotically. In this case, we can apply a similar adjustment factor to arrive at the perfect granularity adjusted VaR (PGAVaR) so that:

$$\text{PGAVaR}(\theta_0, \alpha, n) = \text{EVaR}(\theta_0, \alpha, N) + \text{bias}(\theta_0, \alpha, n), \tag{17}$$

where n is the number of assets in the portfolio under study and N a large number of assets for which the asymptotic model is valid.

Example 3 *As an illustration, and following here Gagliardini and Gouriéroux (2013), in the special case of independent stochastic drift and volatility, the granularity risk (denoted below $GR(\cdot)$) that impacts the estimated VaR can be expressed as (with the previous notations):*

$$EVaR(\theta, \alpha, N) = ThVaR(\theta, \alpha, n) + (n^{-1})GR(\alpha) + o(n^{-1}),$$

with:

$$GR(\alpha) = -(2^{-1})E\left\{\sigma^2[q]\right\} \times \left\{\frac{d\log f(q)}{dq}\right\},$$

where n is the number of assets in the portfolio under study, N a large number of assets for which the asymptotic model is valid, $q = EVaR(\theta_0, \alpha, N)$ is the quantile of a factor G and $f(\cdot)$ is its density function.

Measurement error

Financial data are prone to measurement errors caused by various phenomena such as non-synchronous trading, rounding errors, infrequent trading, microstructure noise or insignificant volume exchanges. In addition, observed data might be subject to manipulations (smoothing, extra revenues, fraudulent exchanges, informationless trading, *etc*).

Measurement error risk can strongly distort backtesting results and significantly affects the performance of standard statistical tests used to backtest VaR models. Frésard et al. (2011) extensively document the phenomena and report that a large fraction of banks artificially boost the performance of their models by polluting their "true" profit and loss with extra revenues that cause under-estimation of the true risk.

Example 4 *Certain financial institutions report a contaminated P&L (denoted PL_t^c) with extraneous profits (denoted π_t)) such as intraday revenues, fees, commissions, net interest incomes and revenues from market making or underwriting activities such as:*

$$PL_t^c = PL_t + \pi_t,$$

with PL_t the true profit at time t.

So, the estimated VaR is impacted by a contamination risk (denoted $CR(\cdot)$) that reads:

$$EVaR(\theta, \alpha, \pi) = ThVaR(\theta, \alpha) + CR(\pi).$$

Liquidity risk

While liquidity has many meanings, from the point of view of risk forecasting, the most relevant are some aspects of market liquidity, as defined by the BCBS (2010), such as the ability to quickly trade large quantities, at a low cost, without impacting the price. These directly follow from Kyle's (1985) three dimensions of liquidity: tightness, depth and resilience.

For portfolios of illiquid securities, reported returns will tend to be smoother than true economic returns, which will understate volatility and increase

risk-adjusted performance measures such as the Sharpe ratio. As an extreme example of illiquidy, we can mention that the NY stock exchange remained shut for more than four months at the beginning of the First World War (from the 31st July, 1914 to the 12th December, 1914) and that the re-opening brought the largest one-day percentage drop in the DJIA (-24.4%).[13]

Getmansky et al. (2004) propose, for instance, an econometric model of illiquidity exposure and develop estimators for the smoothing profile as well as a smoothing-adjusted Sharpe ratio (that basically leads to the intensification of the measured smoothed volatility by a factor to recover a proxy of the true underlying volatility). Measures for gauging illiquidity exposure of several asset classes are presented in Chan et al. (2006). Liquidity aspects enter the Value-at-Risk methodology quite naturally. The VaR approach is built on the hypothesis that "market prices represent achievable transaction prices" (Jorion, 2007). In other words, the prices used to compute market returns in the VaR models have to be representative of market conditions and traded volume. Consequently, the price impact of portfolio liquidation has to be taken into account. Chordia et al. (2001) find a significant cross-sectional relation between stock returns and the variability of liquidity, which is approximated by measures of trading activity such as volume and turnover. Giot and Grammig (2005), using a weighted spread in an intraday VaR framework, show that accounting for liquidity risk becomes a crucial factor and that the traditional (frictionless) measures severely underestimate the true VaR.

Example 5 *As a simple illustration, we can formalize that risk using the following relation (with the previous notations):*

$$\hat{PL}_t = PL_t + \pi_{1,t} + 1\!I_{Le}\pi_{2,t},$$

where $\pi_{1,t}$ is a factor that contributes to the smoothing of the released prices and $\pi_{2,t}$ a liquidity risk premium that only occurs when a liquidity event happens (denoted Le, such as quotation interruption, due to large movement in the market related to an exogenous shock: war, terrorist attack, a large collapse ...), modelled here thanks to a Heaviside function (1I.) that takes the value 1 when the event happens, which leads to a biased estimated VaR with a liquidity risk (denoted $LR(\cdot)$) as:

$$EVaR(\theta, \alpha, \pi_1, \pi_2) = ThVaR(\theta, \alpha) + LR(\pi_1, \pi_2).$$

---

[13]See e.g. Silber (2005).

B Main Backtest Procedures (Web Appendices)

We present hereafter three tests proposed in the literature to gauge the accuracy of VaR estimates.

The first test for a good VaR is the so-called "traffic light" approach in the regulatory framework, related to the Kupiec (1995) Proportion of Failure Test. The Unconditional Coverage test (Kupiec, 1995) attempts to determine whether the observed frequency of exceptions is consistent with the expected frequency of exceptions according to a chosen VaR model and a confidence interval (an exception occurs when the *ex post* return is below the *ex ante* VaR)[14]. We define $I_t^{\mathrm{EVaR}(\cdot)}(\alpha)$ as the "hit variable" associated to the *ex post* observation of $\mathrm{EVaR}(\cdot)$ exceptions at the threshold α at date t, so that (with previous notations):

$$I_t^{\mathrm{EVaR}(\cdot)}(\alpha) = \begin{cases} 1 \text{ if } r_t < -\mathrm{EVaR}(\hat{\theta}, \alpha)_{t-1} \\ 0 \text{ otherwise,} \end{cases} \tag{18}$$

where r_t is the return on portfolio P at time t, with $t = [1, 2, \ldots, T]$.

If we assume that the $I_t^{\mathrm{EVaR}}(\cdot)$ variables are independently and identically distributed, then, under the Unconditional Coverage hypothesis of Kupiec (1995), the cumulated number of VaR violations follows a Binomial distribution, denoted $B(T, \alpha)$, as (see Christoffersen, 1998):

$$\mathrm{Hit}_t^{\mathrm{EVaR}(\cdot)}(\alpha) = \sum_{t=1}^{T} I_t^{\mathrm{EVaR}(\cdot)}(\alpha) \sim B(T, \alpha). \tag{19}$$

A perfect sequence of (corrected) empirical VaR in the sense of this test (not too aggressive, but not too confident), is such that it respects condition (19).

The second test for a good VaR concerns the independence of forecasting errors. The independence hypothesis is associated to the idea that if the VaR model is correct then violations associated to VaR forecasting should be independently distributed, it is also the independence of exceptions hypothesis. If the exceptions exhibit some type of "clustering", then the VaR model may

---

[14]Note that the Basel "traffic light" backtesting framework is directly inspired by this unconditional coverage test. Escanciano and Pei (2012) show, however, that this unconditional test is always inconsistent in detecting non-optimal VaR forecasts based on the historical method. In the following, nevertheless, we consider for our adjustment procedure three of the main tests (including the unconditional coverage test), as well as their bootstrapped corrected versions.

fail to capture the profit and loss variability under certain conditions, which could represent a potential problem down the road. Christoffersen (1998) supposes that, under the alternative hypothesis of VaR inefficiency, the process of $I_t^{\text{EVaR}}(\alpha)$ violations is modelled with a Markov chain whose matrix of transition probabilities is defined by:

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}, \tag{20}$$

where $\pi_{ij} = Pr\left[I_t^{\text{EVaR}}(\alpha) = j | I_{t-1}^{\text{EVaR}} = i\right]$. This Markov chain reflects the existence of an order 1 memory in the process $I_t^{\text{EVaR}}(\alpha)$. The probability of having a violation (not having one) for the current period depends on the occurrence or not of a violation (for the same level of coverage rate) in the previous period. Christoffersen (1998) shows that the likelihood ratio for the test is:

$$\text{LRind}^{I_t^{\text{EVaR}}(\alpha)} = 2\left[\log L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11}) - \log L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi)\right] \overset{d}{\sim} > \chi^2(1), \tag{21}$$

where $L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11})$ is thus the likelihood under the hypothesis of the first-order Markov dependence, and $L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi)$ is the likelihood under the hypothesis of independence $\pi_{01} = \pi_{11} = \pi$ as:

$$L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11}) = (1 - \pi_{01})^{T_{00}}\pi_{01}^{T_{10}}(1 - \pi_{11})^{T_{10}}\pi_{11}^{T_{11}},$$

and:

$$L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi) = (1 - \pi)^{T_{00} + T_{10}}\pi^{T_{01} + T_{11}},$$

with T_{ij} the number of observations in the state j for the current period and at state i for the previous period, $\pi_{01} = T_{01}/(T_{00} + T_{01})$, $\pi_{11} = T_{11}/(T_{10} + T_{11})$ and $\pi = (T_{01} + T_{11})/T$.

A perfect sequence of corrected (empirical) VaR in the sense of this test (*i.e.* not too reactive, but not too smooth) is such that it respects condition (21).

A third category of tests considers the magnitude or size of violation. This class of tests is based on the intuition that VaR exceptions are treated as continuous random variables. For this test, Berkowitz (2001) transforms the empirical series into a standard normal z_{t+1} series. He defines the observed quantile q_{t+1} with the distribution forecast f_{t+1} for the observed portfolio return r_t as:

$$q_{t+1} = \int_{-\infty}^{r_{t+1}} f_{t+1}(r)dr. \tag{22}$$

The z_{t+1} values are then compared to the normal random variables with the desired coverage level of the VaR estimates:

$$z_{t+1} = \Phi^{-1}(q_{t+1}), \tag{23}$$

where $\Phi^{-1}(\cdot)$ is the quantile function of the standard normal density.

If the VaR model generating the empirical quantiles is correct, then the γ_{t+1} series should be identically distributed with the unconditional mean and standard deviation, denoted (μ, σ) and should equal $(0, 1)$, as such:

$$\gamma_{t+1} = \begin{cases} z_{t+1} \text{ if } z_{t+1} < \Phi^{-1}(\alpha) \\ 0 \text{ otherwise,} \end{cases} \tag{24}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Finally, the corresponding test statistic is:

$$\text{LRmag}^{\gamma_{t+1}} = 2\left[L_{\text{mag}}^{\gamma_{t+1}}(\mu, \sigma) - L_{\text{mag}}^{\gamma_{t+1}}(0, 1)\right] \overset{d}{\sim} > \chi^2(2), \tag{25}$$

where:

$$\begin{aligned} L_{\text{mag}}^{\gamma_{t+1}}(\mu, \sigma) = \quad & \sum_{\{\gamma_{t+1}=0\}} \log\left\{1 - \Phi\left\{\frac{\Phi^{-1}(\alpha)-\mu}{\sigma}\right\}\right\} \\ + \quad & \sum_{\{\gamma_{t+1}\neq 0\}} \left\{-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(\gamma_{t+1}-\mu)^2}{2\sigma^2} - \log\left\{\Phi\left\{\frac{\Phi^{-1}(\alpha)-\mu}{\sigma}\right\}\right\}\right\}. \end{aligned}$$

A perfect sequence of (corrected) empirical VaR in the sense of this test (*i.e.* not too conservative, but not too over-confident) is such that it respects condition (25).

For unconditional and conditional coverage tests, Escanciano and Olmo (2009, 2010 and 2011) approximate the critical values. Thus, they propose to use robust sub-sampling techniques to approximate the true distribution of these tests. However, they also show that although the estimation risk can be diversified by choosing a large in-sample size relative to an out-of-sample one, the risk associated to the model cannot be eliminated using sub-sampling.

Indeed, let $G_x(x)$ denote the cumulative distribution function of the test statistic k for any $x \in$ IR, and, $k_{b,t} = K(t, t+1, \cdots, t+b-1)$, with $t = [1, 2, \cdots, T-b+1]$, the test statistic computed with the subsample $[1, 2, \cdots, T-b+1]$ of size b.

Hence, the approximated sampling cumulative distribution function of k, denoted $G_{k_b}(x)$, built using the distribution of the values of $k_{b,t}$ computed over the $(T-b+1)$ different consecutive subsamples of size b is given by:

$$G_{k_b}(x) = (T-b+1)^{-1} \sum_{t=1}^{T-b+1} 1\!\!I_{\{k_{b,t} < x\}}. \tag{26}$$

43

The $(1 - \tau)^{\text{th}}$ sample quantile of G_{k_b}, is given by:

$$c_{k_b, 1-\tau} = \underbrace{\inf}_{x \in \mathrm{IR}} \left\{ x \,|\, G_{k_b}(x) \geq 1 - \tau \right\}. \tag{27}$$

C Miscellaneous Complementary Results (Web Appendices)

Table A.1. Illustrations of Unconditional Simulated Errors associated to the 95%, 99% and 99.5% Annualized VaR: Gaussian *versus* t-Student Quantiles

	Panel A. Gaussian DGP and Gaussian VaR with Estimation Error					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-29.49%	.00%	.00%	-7.93%	7.24%
$\alpha = 99.00\%$	-41.88%	-41.88%	.00%	.00%	-9.92%	9.17%
$\alpha = 99.50\%$	-46.41%	-46.41%	.00%	.00%	-12.45%	10.16%

	Panel B. t-Student(5) DGP and Gaussian VaR with Specification Error					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-6.73%	-6.73%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-18.87%	-18.87%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-26.46%	-26.46%

	Panel C. t-Student(5) DGP and Gaussian VaR with Specification and Estimation Errors					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-13.97%	1.20%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-28.04%	-8.95%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-36.62%	-14.01%

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 20th September, 2011. These statistics were computed with the results of 100,000 simulated series of 250 daily returns according to a specific DGP (Gaussian for Panel A and t-Student(5) for Panel B and C) and using an annualized parametric VaR. The columns represent, respectively, the average Estimated VaR with specification or/and estimation errors, the Theoretical VaR, and the average-minimum-maximum of the adjustment terms of all samples. A positive adjustment term indicates that the Estimated VaR (negative return) should be more conservative (more negative).

Table A.2. Estimated Annualized VaR and Model-risk Errors (%)
in the Brownian Case

Three price processes of the asset returns are considered below, such as for $t = [1, \cdots, T]$ and $p = [1, 2, 3]$:

$$dS_t = S_t(\mu dt + \sigma dW_t + J_t^p dN_t),$$

with $J_t^1 = 0$ for Brownian, where S_t is the price of the asset at time t, W_t is a standard Brownian motion, independent from the Poisson process N_t, governing the jumps of various intensities J_t^p (null, constant or time-varying according to the process p).

	Panel A. Gaussian DGP and Gaussian VaR with Estimation Error					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-24.78%	-24.78%	.00%	.00%	-8.69%	10.16%
$\alpha = 99.00\%$	-35.74%	-35.74%	.00%	.00%	-14.21%	20.70%
$\alpha = 99.50\%$	-39.95%	-39.95%	.00%	.00%	-16.04%	28.92%

	Panel B. Brownian DGP and Gaussian VaR with Specification Error					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-6.73%	-6.73%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-18.87%	18.87%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-26.46%	26.46%

	Panel C. Brownian DGP and Gaussian VaR with Specification and Estimation Errors					
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	- 6.73%	-6.73%	-13.97%	1.20%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-28.04%	-8.95%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-36.62%	-14.01%

Source: simulations by the authors. Errors are defined as the differences between the "true" asymptotic simulated VaR and the Estimated VaR. These statistics were computed with a series of 250,000 simulated daily returns with specific DGP (Brownian), averaging the parameters estimated in Aït-Sahalia et al. (2013, Table 2, *i.e.* β=41.66%, λ_3=1.20% and γ=22.22%), and *ex post* recalibrated for sharing the same first two moments (*i.e.* μ=.12% and σ=1.02%) and the same mean jump intensity (for the last two processes – which leads after rescaling here for instance, to an intensity of the Lévy such as: λ_2=1.06%). *Per* convention, a negative adjustment term in the table indicates that the Estimated VaR (negative return) should be more conservative (more negative).

Table A.3. Estimated Annualized VaR and Model-risk Errors (%) in the Lévy Case

Three price processes of the asset returns are considered below, such as for $t = [1, \cdots, T]$ and $p = [1, 2, 3]$:

$$dS_t = S_t(\mu dt + \sigma dW_t + J_t^p dN_t),$$

with $J_t^2 = \lambda_2 exp(-\lambda_2 t)$ for Lévy, where S_t is the price of the asset at time t, W_t is a standard Brownian motion, independent from the Poisson process N_t, governing the jumps of various intensities J_t^p (null, constant or time–varying according to the process p), defined by parameters, λ_2, which is a positive constant.

Panel A. Gaussian DGP and Gaussian VaR with Estimation Error

Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-24.78%	-24.78%	.00%	.00%	-8.69%	10.16%
$\alpha = 99.00\%$	-35.74%	-35.74%	.00%	.00%	-14.21%	20.70%
$\alpha = 99.50\%$	-39.95%	-39.95%	.00%	.00%	-16.04%	28.92%

Panel B. Lévy DGP and Gaussian VaR with Specification Error

Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-6.73%	-6.73%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-18.87%	-18.87%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-26.46%	-26.46%

Panel C. Lévy DGP and Gaussian VaR with Specification and Estimation Errors

Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-13.97%	1.20%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-28.04%	-8.95%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-36.62%	-14.01%

Source: simulations by the authors. Errors are defined as the differences between the "true" asymptotic simulated VaR and the Estimated VaR. These statistics were computed with a series of 250,000 simulated daily returns with specific DGP (Lévy), averaging the parameters estimated in Aït-Sahalia et al. (2013, Table 2, *i.e.* β=41.66%, λ_3=1.20% and γ=22.22%), and *ex post* recalibrated for sharing the same first two moments (*i.e.* μ=.12% and σ=1.02%) and the same mean jump intensity (for the last two processes - which leads after rescaling here, for instance, to an intensity of the Lévy with λ_2=1.06%). *Per* convention, a negative adjustment term in the table indicates that the Estimated VaR (negative return) should be more conservative (more negative).

Table A.4. Estimated annualized VaR and model–risk errors (%) in the Hawkes Case

Three price processes of the asset returns are considered below, such as for $t = [1, \cdots, T]$ and $p = [1, 2, 3]$:

$$dS_t = S_t(\mu dt + \sigma dW_t + J_t^p dN_t),$$

with $J_t^3 = \lambda_3 + \beta exp[-\gamma(t-s)]$ for Hawkes, where S_t is the price of the asset at time t, W_t is a standard Brownian motion, independent from the Poisson process N_t, governing the jumps of various intensities J_t^p (null, constant or time-varying according to the process p), defined by parameters, λ_3, β and γ, which are positive constants with s the date of the last observed jump.

Panel A. Gaussian DGP and Gaussian VaR with Estimation Error						
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-24.78%	-24.78%	.00%	.00%	-8.69%	10.16%
$\alpha = 99.00\%$	-35.74%	-35.74%	.00%	.00%	-14.21%	20.70%
$\alpha = 99.50\%$	-39.95%	-39.95%	.00%	.00%	-16.04%	28.92%

Panel B. Hawkes DGP and Gaussian VaR with Specification Error						
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-6.73%	-6.73%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-18.87%	-18.87%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-26.46%	-26.46%

Panel C. Hawkes DGP and Gaussian VaR with Specification and Estimation Errors						
Probability	Mean Estimated VaR	Perfect VaR	Mean Bias	Median Bias	Min. Bias	Max. Bias
$\alpha = 95.00\%$	-29.49%	-36.22%	-6.73%	-6.73%	-13.97%	1.20%
$\alpha = 99.00\%$	-41.88%	-60.75%	-18.87%	-18.87%	-28.04%	-8.95%
$\alpha = 99.50\%$	-46.41%	-72.87%	-26.46%	-26.46%	-36.62%	-14.01%

Source: simulations by the authors. Errors are defined as the differences between the "true" asymptotic simulated VaR and the Estimated VaR. These statistics were computed with a series of 250,000 simulated daily returns with specific DGP (Hawkes), averaging the parameters estimated in Aït-Sahalia et al. (2013, Table 2, *i.e.* β=41.66%, λ_3=1.20% and γ=22.22%), and *ex post* recalibrated for sharing the same first two moments (*i.e.* μ=.12% and σ=1.02%) and the same mean jump intensity. *Per* convention, a negative adjustment term in the table indicates that the Estimated VaR (negative return) should be more conservative (more negative).

Table A.5. A Road Map of the Main Risk Model Validation Tests

---

Exception Frequency Tests

Intuition: test the violation frequency that should be equal to the probability threshold

An Unconditional Coverage Test - Kupiec (1995)
A GMM Duration Test - Candelon et al. (2011)
A Z-test - Jorion (2007)
A Multi-variate Unconditional Coverage Test - Pérignon and Smith (2008)
A D-test - Escanciano and Pei (2012)

Exception Independence Tests

Intuition: test the violations associated to the VaR forecasting that should be independent (not clustered and/or no forecasting power via a time-series model for extremes)

An Independence Test - Christoffersen (1998)
A Violation Duration-based Test - Christoffersen and Pelletier (2004)
A Discrete Violation Duration-based Test - Haas (2005)
A Dynamic Quantile Test - Engle and Manganelli (2004)
A GMM Duration Test - Candelon et al. (2011)
A Multivariate Test of Zero-autocorrelation of Violations - Hurlin and Tokpavi (2006)
An Estimation-risk adjusted Test - Escanciano and Olmo (2009, 2010 and 2011)

Exception Frequency and Independence of Violations Tests

Intuition: test jointly the hit ratio and the independence of VaR violations

A Conditional Coverage Test - Christoffersen (1998)
A GMM Duration Test - Candelon et al. (2011)
A Dynamic Binary Response Test - Dumitrescu et al. (2012)

Exception Magnitude Tests

Intuition: test the amplitude of VaR violations (that should be small)

A Magnitude Test (under normality assumption) - Berkowitz (2001)
A Test based on a Loss Function - Lopez (1998 and 1999)
A Two-stage Test (Coverage Rate and Loss Function) - Angelidis and Degiannakis (2007)
A Double-threshold Test - Colletaz et al. (2013)

Exceedances for Expected Shortfall Test

Intuition: Measure the observed ES, then compare to a local approximated value (and the difference should be small)

A Saddlepoint Technique Test for ES - Wong (2008 and 2010)

See, among others, Campbell (2007), Nieto and Ruiz (2008) and Berkowitz et al. (2011) for comprehensive surveys.

Table A.6. Dates of the Maximum Adjustment for different 95% VaRs and Backtests Models at 5% Confidence Level

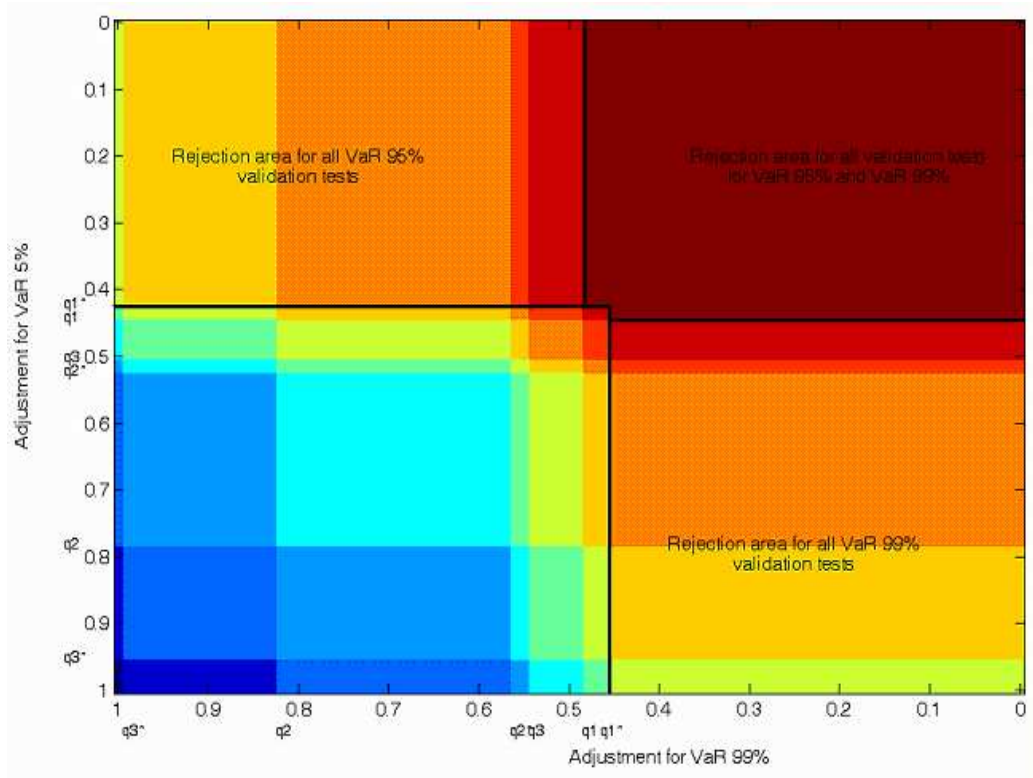| VaR Methods		Dates	$\underline{q}_1$	Dates	$\underline{q}_1^*$	Dates	$\underline{q}_2$	Dates	$\underline{q}_2^*$	Dates	$\underline{q}_3$	Dates	$\underline{q}_3^*$
Historical	1	02/09/2009	42.02%	01/16/2009	32.80%	09/08/1930	78.26%	04/06/2009	52.18%	08/24/2009	49.91%	09/08/1930	95.06%
	2	11/28/2008	40.60%	01/06/1930	31.94%	04/06/2009	46.46%	08/24/2009	42.02%	09/08/1930	48.91%	04/21/1930	78.26%
	3	10/16/1930	39.97%	01/05/1933	21.73%	07/25/1988	42.52%	04/21/1930	40.15%	04/06/2009	47.68%	12/02/1929	72.61%
	4	12/11/1929	37.75%	01/06/1931	19.87%	03/07/1988	40.82%	12/02/1929	37.79%	11/17/2008	46.46%	04/06/2009	65.07%
Normal	1	11/28/2008	42.86%	01/16/2009	30.07%	09/08/1930	74.43%	08/24/2009	44.46%	04/06/2009	44.46%	09/08/1930	88.47%
	2	12/11/1929	38.57%	01/06/1930	25.55%	07/25/1988	39.56%	11/17/2008	42.86%	11/17/2008	42.86%	12/02/1929	74.43%
	3	09/14/2009	38.07%	01/05/1933	17.88%	12/02/1929	38.57%	09/08/1930	42.45%	12/02/1929	38.86%	11/17/2008	62.83%
	4	11/11/1929	36.69%	01/06/1938	14.92%	03/07/1988	34.42%	12/02/1929	38.86%	04/21/1930	38.57%	04/21/1930	59.24%
Student	1	11/28/2008	40.16%	01/16/2009	30.02%	09/08/1930	68.50%	04/06/2009	45.93%	09/08/1930	50.18%	09/08/1930	101.52%
	2	12/11/1929	33.06%	01/06/1930	18.40%	12/02/1929	68.18%	08/24/2009	40.27%	04/06/2009	44.50%	11/17/2008	86.13%
	3	09/14/2009	32.85%	01/05/1933	14.25%	07/25/1988	35.52%	12/02/1929	35.86%	12/02/1929	33.57%	03/07/1988	71.79%
	4	11/11/1929	31.43%	01/13/1975	11.14%	03/07/1988	30.26%	09/08/1930	34.80%	04/21/1930	33.06%	12/02/1929	68.50%
Cornish Fisher	1	05/13/1915	133.65%	01/04/1916	120.56%	02/14/1916	135.43%	09/27/1915	142.87%	09/27/1915	135.43%	09/27/1915	142.87%
	2	05/07/1915	133.36%	01/04/1915	106.47%	09/27/1915	133.86%	05/10/1915	133.86%	05/10/1915	133.86%	05/10/1915	130.22%
	3	05/06/1915	131.48%	01/03/1917	82.83%	05/10/1915	133.65%	02/14/1916	114.82%	02/14/1916	117.67%	04/09/1917	111.55%
	4	05/04/1915	130.22%	01/14/1988	76.04%	09/08/1930	93.47%	07/03/1916	90.03%	03/07/1988	90.73%	09/08/1930	93.47%
Risk Metrics	1	03/28/1938	15.85%	01/04/1921	10.50%	05/10/1915	32.76%	12/02/1929	16.43%	12/02/1929	16.43%	03/21/1932	46.62%
	2	10/28/1929	15.02%	01/06/1938	9.14%	12/02/1929	31.91%	05/03/1920	16.33%	05/09/1938	16.04%	12/20/1937	32.36%
	3	03/15/1938	14.80%	01/02/1908	7.88%	04/21/1930	30.51%	09/26/1938	15.85%	12/20/1937	14.57%	03/07/1988	31.92%
	4	01/25/1938	14.57%	01/06/1930	5.64%	07/25/1988	26.69%	09/08/1930	15.02%	05/03/1920	14.35%	12/02/1929	31.91%
GARCH	1	03/24/1938	18.24%	01/06/1930	15.50%	09/08/1930	41.44%	05/09/1938	18.48%	12/02/1929	19.42%	05/09/1938	39.61%
	2	04/06/1938	18.15%	01/06/1938	9.28%	12/02/1929	34.12%	12/20/1937	18.24%	05/03/1920	18.55%	11/02/1931	35.40%
	3	10/28/1929	17.17%	01/02/1908	8.09%	07/25/1988	32.49%	09/26/1938	18.15%	05/09/1938	18.48%	12/20/1937	34.98%
	4	03/15/1938	16.78%	01/17/2008	7.01%	03/07/1988	30.73%	12/02/1929	17.17%	09/08/1930	17.17%	03/07/1988	33.59%
CAViaR	1	01/21/1994	30.10%	01/17/2008	24.96%	09/24/2007	41.75%	02/11/2008	35.73%	09/24/2007	33.48%	03/21/1932	41.32%
	2	06/06/2007	29.81%	01/14/1994	20.53%	09/08/1930	41.44%	04/25/1994	32.95%	04/25/1994	31.72%	11/30/1998	38.82%
	3	02/26/2008	28.11%	01/17/2006	15.47%	12/02/1929	34.12%	09/24/2007	32.64%	04/19/1999	23.31%	10/19/1987	33.59%
	4	03/11/2008	26.93%	01/15/1999	13.87%	07/25/1988	32.49%	09/12/1994	31.72%	07/31/2006	19.50%	04/19/1999	31.35%
GEV	1	11/28/2008	39.05%	01/06/1930	32.12%	09/08/1930	72.11%	08/24/2009	48.25%	04/06/2009	45.07%	04/21/1930	112.42%
	2	12/11/1929	36.82%	01/16/2009	29.52%	12/02/1929	61.01%	04/06/2009	41.84%	08/24/2009	41.84%	09/08/1930	75.87%
	3	11/11/1929	35.24%	01/05/1933	14.55%	03/07/1988	34.35%	04/21/1930	37.15%	04/21/1930	39.52%	12/02/1929	72.11%
	4	09/14/2009	33.32%	01/08/1947	13.62%	11/17/2008	29.52%	12/02/1929	35.24%	11/17/2008	39.05%	11/17/2008	56.53%
GPD	1	11/28/2008	37.89%	01/16/2009	26.98%	09/08/1930	71.38%	11/17/2008	42.33%	04/06/2009	43.75%	04/21/1930	105.01%
	2	09/14/2009	34.58%	01/06/1930	25.64%	04/06/2009	42.33%	04/21/1930	35.69%	08/24/2009	42.33%	09/08/1930	71.38%
	3	12/11/1929	32.86%	01/06/1931	10.30%	12/02/1929	32.86%	12/02/1929	32.86%	04/21/1930	32.86%	12/02/1929	67.67%
	4	07/18/1930	32.80%	01/05/1933	7.67%	07/25/1988	31.95%	05/09/1938	29.66%	09/08/1930	32.80%	11/17/2008	52.51%

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to re-estimate parameters dynamically for the various methods. The variable $\underline{q}_1$ refers to the hit test; $\underline{q}_2$ to the independence test; $\underline{q}_3$ to the magnitude test; and $\underline{q}_1^*$, $\underline{q}_2^*$, $\underline{q}_3^*$ correspond to their resampling versions, following Escanciano and Olmo (2009, 2010 and 2011).

Table A.7. Minimum k Ratio Model Risk for 95% Annualized Value-at-Risk Models for various Validity Tests with 5%

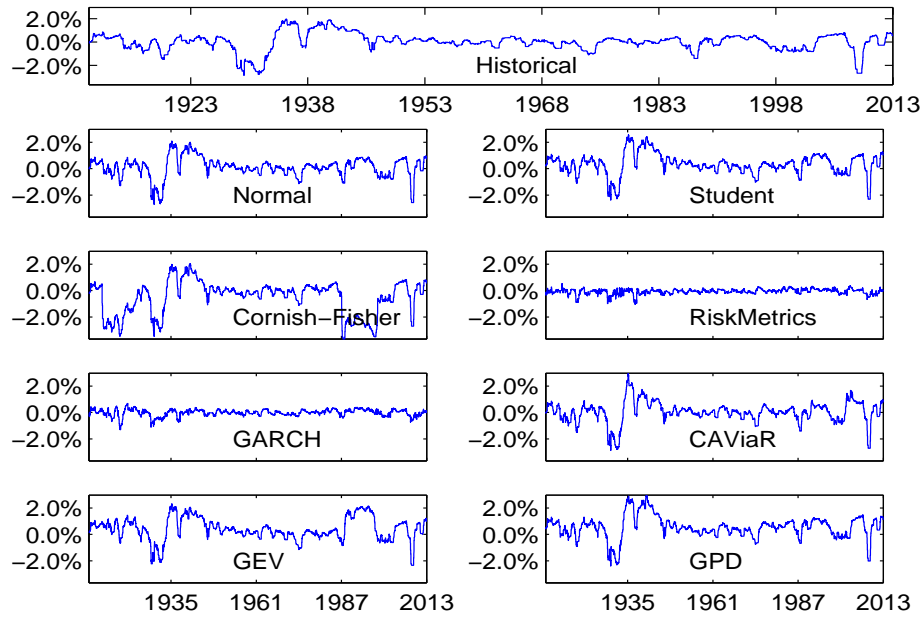| VaR Methods	Mean VaR	$\underline{q}_1$	$\underline{q}_1^*$	$\underline{q}_2$	$\underline{q}_2^*$	$\underline{q}_3$	$\underline{q}_3^*$
Historical	-25.78%	2.37	2.08	3.29	2.53	2.59	3.79
Normal	-27.09%	2.26	1.84	2.84	2.31	2.32	3.18
Student	-30.52%	2.02	1.73	2.52	2.04	2.04	3.31
Cornish-Fisher	-20.25%	3.45	3.73	4.88	4.76	3.72	4.88
RiskMetrics	-25.67%	1.71	20.77	57.18	41.47	36.84	101.69
GARCH	-25.99%	1.59	1.75	2.65	1.85	1.94	2.35
CAViaR	-26.84%	10.95	9.82	23.89	8.9	8.55	7.6
GEV	-29.71%	2.01	1.72	2.37	2.13	2.06	3.24
GPD	-33.97%	2.04	1.69	2.64	2.21	2.01	3.63

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 20th September, 2011. We use a moving window of four years (1,040 daily returns) to dynamically re-estimate parameters for the various methods. The variable $\underline{q}_1$ refers to the hit test; $\underline{q}_2$ to the independence test; $\underline{q}_3$ to the magnitude test; and $\underline{q}_1^*$, $\underline{q}_2^*$, $\underline{q}_3^*$ correspond to their resampling versions, following Escanciano and Olmo (2009, 2010 and 2011).

Figure A.1: Risk Map for Maximum Annualized Adjustment Values at 5% Confidence Levels for Tests for 95% and 99% Value-at-Risk Models (see Colletaz et al., 2013)
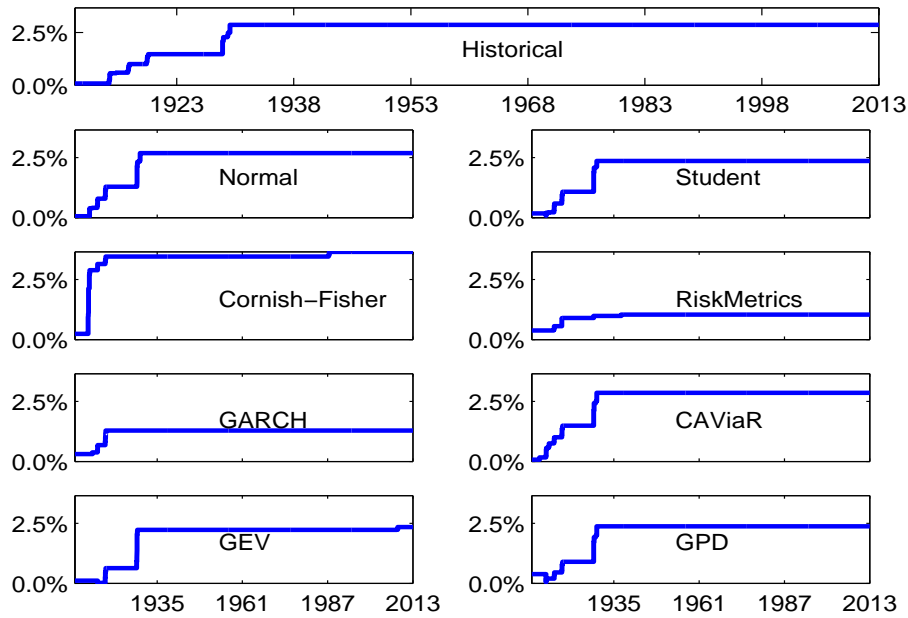

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 20th September, 2011; computations by the authors. We use a moving window of four years (1,040 daily returns) to dynamically re-estimate parameters for the various methods. The variable q_1 refers to the hit test; q_2 to the independence test; q_3 to the magnitude test; and q_1^*, q_2^*, q_3^* correspond to their resampling versions, following Escanciano and Olmo (2009, 2010 and 2011).

Figure A.2: Dynamic Optimal Adjustment on the Daily 95% VaR related to the Hit Test for 10-Year Sample Estimation Data
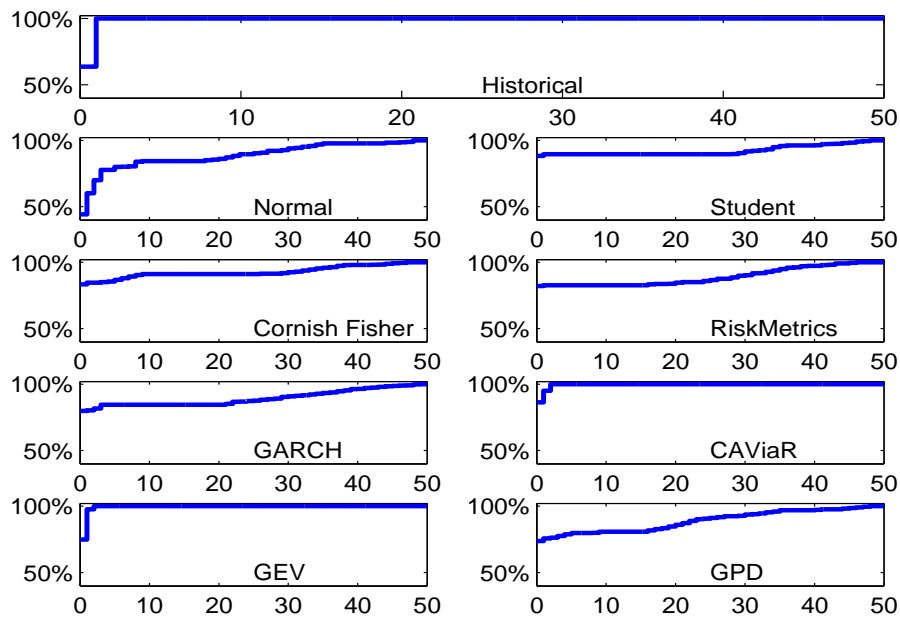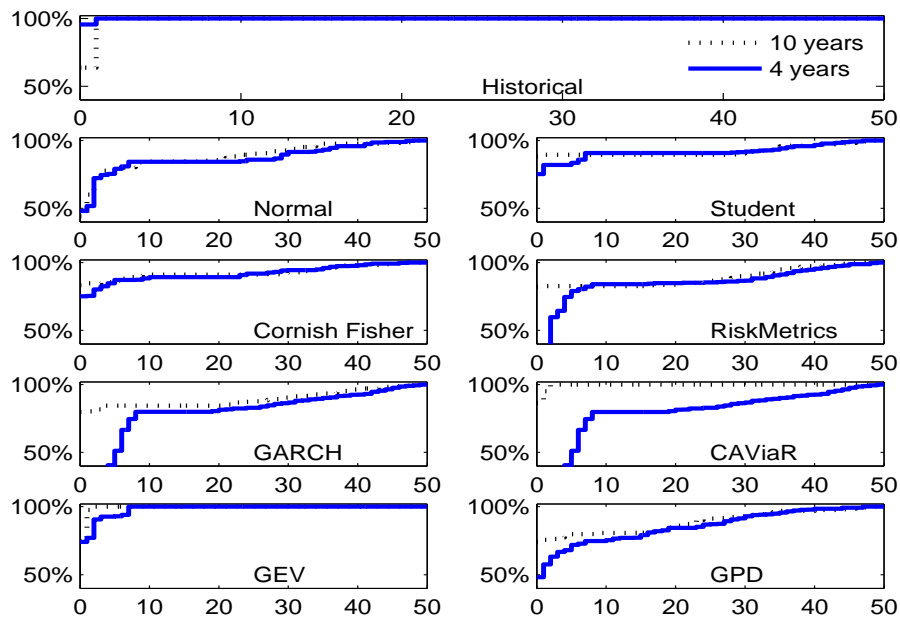

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 7th October, 2013; computations by the authors. We use a moving window of ten years (2,600 daily returns) to re-estimate parameters dynamically for the various methods.

Figure A.3: Optimal Dynamic Absolute Value of Minimum Negative Adjustments for the Hit Test for Different 95% VaR - 10-Year Sample Estimation Data

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 7th October, 2013; computations by the authors. We use a moving window of ten years (2,600 daily returns) to re-estimate parameters dynamically for the various methods.

Figure A.4: Optimal Dynamic Relative Adjustment for the Hit Test for Different Starting Dates and 95% VaR by Horizon (in years) - 10 Years Sample Estimation Data
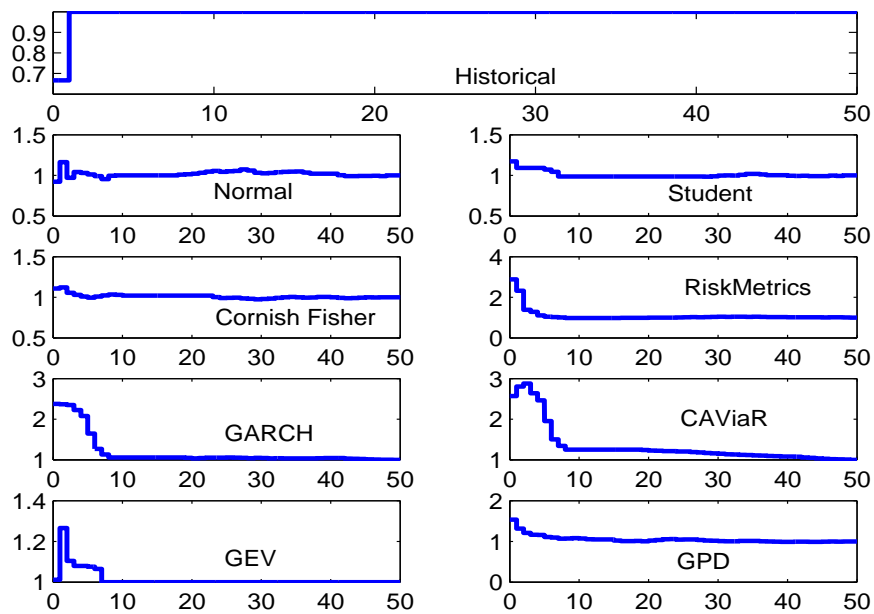

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 7th October, 2013; computations by the authors. We use a moving window of ten years (2,600 daily returns) to dynamically re–estimate parameters for the various methods. This figure illustrates the dynamic negative adjustment required for passing the hit test, having randomly chosen the first date of implementation. Optimal relative negative adjustments are here expressed in terms of the percentage of their maximum value over the whole sample.

Figure A.5: Optimal Dynamic Relative Adjustment for the Hit Test for Different Starting Dates and 95% VaR by Horizon (in years) − 4 years and 10 years VaR Estimation

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 7th October, 2013; computations by the authors. We use a moving window of four and ten years to dynamically re-estimate parameters for the various methods. This figure illustrates the dynamic negative adjustment required for passing the hit test, having randomly chosen the first date of implementation. Optimal relative negative adjustments are here expressed in terms of the percentage of their maximum value over the whole sample.

Figure A.6: Ratio Ten/Four Year Optimal Dynamic Relative Adjustment for the Hit Test for Different Starting Dates and 95% VaR by Horizon (in years)

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January, 1900 to the 7th October, 2013; computations by the authors. We use a moving window of four and ten years to dynamically re–estimate parameters for the various methods. This figure illustrates the dynamic negative adjustment required for passing the hit test, having randomly chosen the first date of implementation. Optimal relative negative adjustments are here expressed in terms of the percentage of their maximum value over the whole sample.