"Risk Model-at-Risk"*

Christophe M. Boucher[†] Jón Daníelsson[‡]
Patrick S. Kouontchou[§] Bertrand B. Maillet[¶]

March 2012
Preliminary Version

Abstract

The recent experience from the global financial crisis has raised serious doubts about the accuracy of standard risk measures as a tool to quantify extreme downward risks. Risk measures are hence subject to a "model risk" due, e.g., to the specification and estimation uncertainty. Therefore, regulators have proposed that financial institutions assess the "model risk" but, as yet, there is no accepted approach for computing such a risk. We propose a general framework to compute risk measures robust to the model risk, while focusing on the Value-at-Risk (VaR). The proposed procedure aims empirically adjusting the imperfect quantile estimate based on a backtesting framework, assessing the good quality of VaR models such as the frequency, the independence and the 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.

*We thank Carol Alexander, Arie Gozluklu, Monica Billio, Massimiliano Caporin, Rama Cont, Christophe Hurlin, Sébastien Laurent, Christophe Pérignon, Michaël Rockinger, Thierry Roncalli and Jean-Michel Zakoïan for suggestions when preparing this article, as well as Benjamin Hamidi for research assistance and joint collaborations on collateral subjects. All errors are ours and the usual disclaimer applies. The first author thanks the Banque de France Foundation and the fourth the Europlace Institute of Finance for financial supports. Some extra materials related to this article can be found on: www.riskresearch.org.

[†]A.A.Advisors-QCG (ABN AMRO), Variances and University of Paris-1 (CES/CNRS).

[‡]London School of Economics.

[§]Variances and University of Lorraine (CEREFIGE).

[¶]A.A.Advisors-QCG (ABN AMRO), Variances and University of Orléans (LEO/CNRS and EIF). Corresponding author: Dr. Bertrand B. Maillet, University of Orléans, Rue de Blois - BP 26739 - 45067 ORLEANS Cedex 2, France. Email: bertrand.maillet@univ-orleans.fr.

1 Introduction

Risk forecast methods developed over the past couple of decades where shown to have been mostly inadequate in the past crises. Not only failing to anticipate the extreme risks that were realized, but also struggling to keep up as the crisis unfolded and then unwound. It seems that the risk models got it wrong in all states of the world. Our objective in this paper is to identify explicitly how the models failed and make a proposal as how we should approach risk forecasting in the future. Our focus is on Value-at-Risk (VaR), but the analysis equally applies to other risk measures.

Model risk of risk models refers both to the range of risk estimates as well as the inability to forecast properly risk realizations. For illustration purposes, Figure 1 provides some highest and lowest VaR for a range of extreme return forecasts on the Dow Jones Index on a century, and shows how the most popular risk measure techniques fail to provide a unique and accurate risk forecast. Typically, the estimated VaR do not vary often but when they do, they vary sharply and with some delay relative to extreme returns. Moreover, as emphasized in Daníelsson et al. (2011b), the range of potential candidates of estimated VaR is large and the optimal model is never the same across times[1]. Thus facing a large variety of plausible methods and their related model risk, our main objective is to propose a general method to correct the imperfect risk estimate whatever the risk model implied.

Our main contribution consists in proposing a simple framework to compute risk measures robust to the main model risks based on an incremental buffer assessing the main valuable properties of risk models. Thus, this article can be classified into the general area of modelling the specification and model risk. As such, our approach fits with the general literature on model misspecification, (e.g., Berkowitz, 2001). Unfortunately, less work exists for risk models specifically, beyond standard backtesting. Surprisingly, financial regulators only very recently express concerns over model risk[2], even though risk models have a clear history of failure.

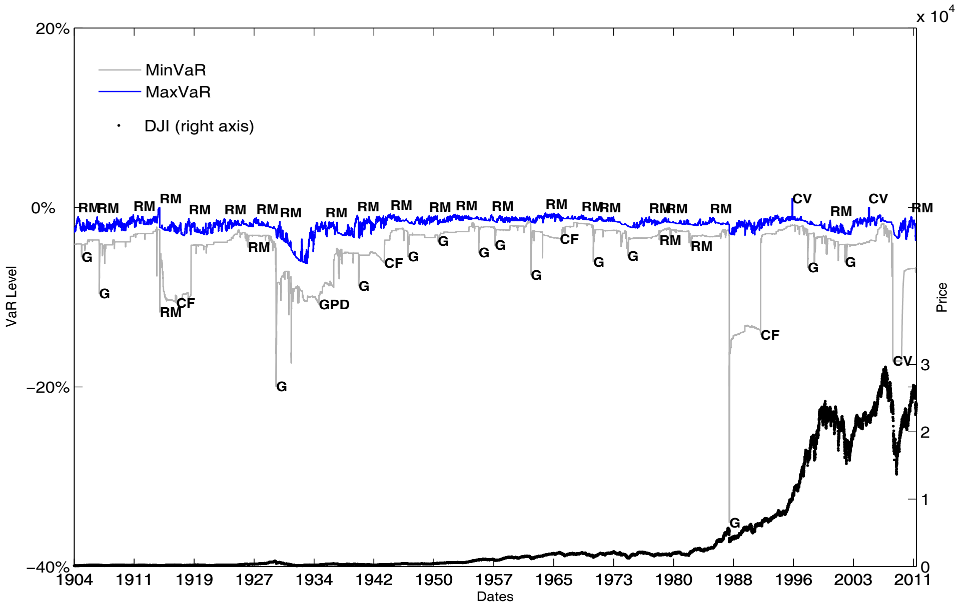
This paper aims to document the model risk of extreme risk measures, based on the effect it has on the quality of the portfolio risk quantification, and to show how to adjust the VaR to main model risks (namely estimation, specification, identification, granularity, data contamination and liquidity). The main idea consists in adjusting computed risk measures (with an incremental correction) based on a backtesting strategy assessing the main good qualities of VaR models such as the frequency, the independence and the magnitude of past violations.

---

[1]Even if the RiskMetrics VaR (denoted RM on Figure 1) tends to be one of the most aggressive, and the GARCH-VaR (denoted G) the most conservative overall.

[2]In July 2009, the Basel Committee on Banking Supervision issued a directive ("Revisions to the Basel II Market Risk Framework") requiring that financial institutions quantify model risk.

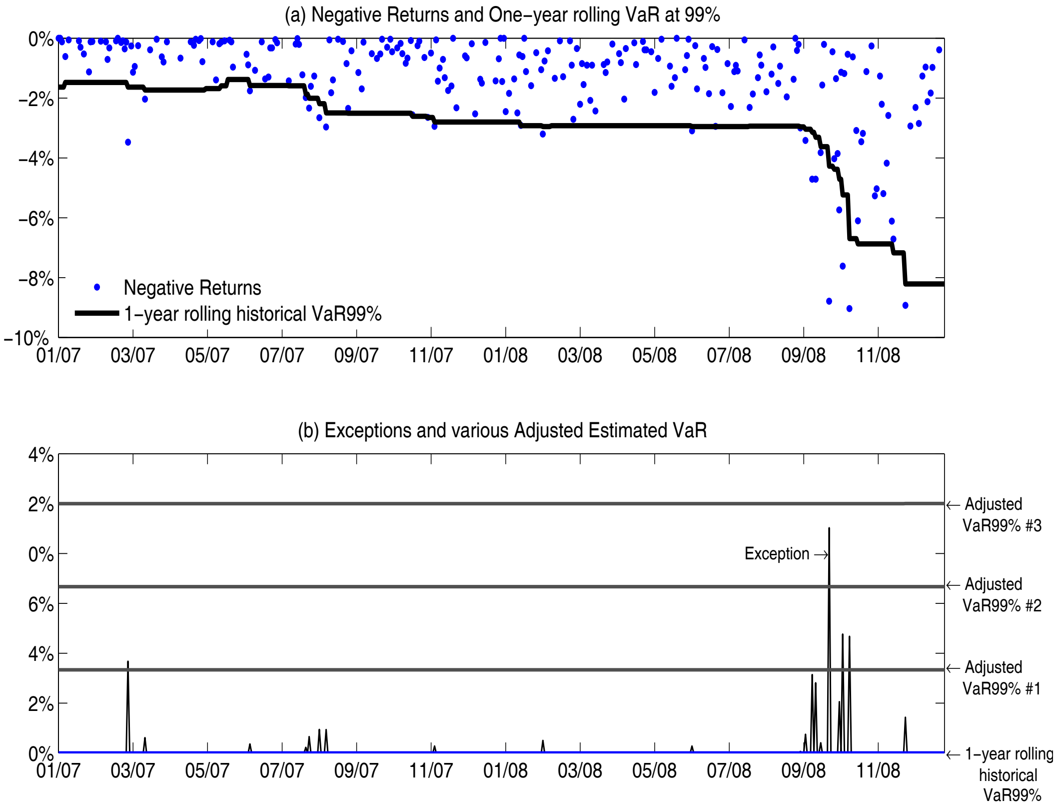
Figure 1: DJIA and Range for a Set of Daily VaR99% Forecasts



Source: Bloomberg; daily data of the DJIA index in USD from 1st January 1900 to the 20th of 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 letters "H", "N", "t", "CF", "RM", "G", "CV", "GEV, "GPD" stand for, respectively, Historical, Normal, Student, Cornish-Fisher, RiskMetrics, GARCH, CAViaR, GEV and GPD method for VaR calculation, and appear on the Figure for some *extrema*, when they happen to be the min or the max of the set of VaR computed on the same date.

First grasp the intuition of our approach with the following realistic illustration based on market events around the Lehman Brothers's collapse. Figure 2 presents on the period from 01/01/2007 to 01/01/2009 the Peaks-over-VaR based on the 1-year rolling daily historical VaR 99% on the S&P 500 index. These points, called "exceptions" or hits, represent the difference between the estimated VaR and the daily returns. This figure shows that hits (returns below the estimated VaR) are too numerous, highly autocorrelated (clusters of violations) and, around October 2008, far from the estimated VaR, even if it is progressively adjusted after the hits. Then, adding a buffer can make the VaR guess more robust to market accidents.

Figure 2: DJIA and Range for a Set of Daily VaR99% Forecasts around the 2008 Lehman Brothers's Event



Source: Bloomberg; daily data of the S&P500 index in USD from the 1st January 2003 to the 1st January 2009; computations by the authors. The figure presents peak-over-VaR based on the 4-year rolling daily historical VaR 99% on the S&P 500 index, as well as corrected VaR estimates with various *ad hoc* incremental buffers (numbered #1 to #3).

Depending on the size of this buffer for model risks considered (from rather prudent to highly cautious ones), the properties of hits are significantly different in terms of frequency, dependence and size. But the question for the risk manager is how to *ex ante* fix the size of this buffer (correction #1, #2 or #3 on the righthand side y-axis in Figure 2)? This paper proposes a method to calibrate the buffer based on a backtesting framework. The corrected-VaR is then equal to an imperfect estimated VaR minus the minimum VaR adjustment - no more and no less - to be validated by backtests.

In this set-up, we show, for some assets and some realistic simulations, how fragile VaR estimates are, not only with respect to the pure choice of the risk model, but also regarding the choice of the underlying econometric model and of the appropriate sampling period.

Our work relates to the recent work of Gagliardini et al. (2010), who propose Estimation and Granularity adjustments for VaR, and very similar to that of Lönnbark (2010), who derives adjustments of interval forecasts to account for parameter estimation. In the context of extreme risk measure, the objective of this paper is similar to Kerkhov et al. (2010), who first propose a incremental market risk capital charge calibrated on the backtesting framework of the regulators, as well as, Alexander and Sarabia (2011) who deal explicitly with VaR model risk by quantifying VaR model risk and proposing an adjustment to regulatory capital based on an maximum entropy *criterion*.

A number of papers have considered estimation risk for risk models. This type of model risk is probably the most frequently discussed in the literature (see for instance Gibson et al., 1999; and Talay and Zheng, 2002). The issue of estimation risk for VaR has been considered previously in the identically and independently distributed return case by, for example, Jorion (1996) and Pritsker (1997). Estimation risk in dynamic models has also been considered by several authors. Berkowitz and O'Brien (2002) observe that the usual VaR estimates are too conservative (at the time of publication). 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 errors 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 (*i.e.* an empirical quantile) 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 error, 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 errors distributions, Chan et al. (2007) derive the asymptotic distributions of extremal quantiles. Escanciano and Olmo (2009, 2010-a and 2010-b) 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 (2010) 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.

The outline of the paper is as follows. The next section exposes and discusses the model risks of VaR. Section 4 presents simple illustrations of elementary model risks with realistic simulations. Section 5 proposes an economic valuation of model risks. Section 6 concludes the body of the paper, but the appendix follows, outlining the main back testing methods used in the paper.

2 Model risk when forecasting risk

Financial risk forecast models, just like any other statistical models, are thus subject to model risk. In spite of this, almost all presentations of risk forecasts focus on point estimate, omitting any mention of model risk, not to even mention 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 5 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 onto 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 availability and the inability to conduct sometimes a transaction at current market prices because of the too large relative 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 model parameters θ_0. In what follows, VaR is the $(1-\alpha)-$quantile (with $\alpha > .50$) of the profit and loss distribution so that the VaR is negative. We also indicate the theoretical (or true) VaR by ThVaR(θ_0, α). Thus, when comparing the estimated VaR with the theoretical VaR (i.e. EVaR and ThVaR respectively), we present both the buffer needed to adjust directly the EVaR and the probability (or quantile) shift required. In the following sub-sections, we detail these specific model risks that impact VaR forecasts.

2.1 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 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.

While one could investigate the effect of parameter variations on risk measures that are computed in a particular parametric framework, our aim here is to explicitly consider model risk as a separate risk factor. Our work is similar in spirit to the work of West (1996), who discusses when and how to adjust critical values for tests of predictive ability in order to take parameter estimation uncertainty into account

Consider the best case scenario where we know the DGP but only observe a finite sample. In this case, the estimated VaR will be a biased estimates of the theoretical VaR. We hereafter indicate the bias by the function $\text{bias}(\hat{\theta}, \theta_0, \alpha)$. In this best case scenario, we know bias function and can therefore obtain the perfect estimation adjusted VaR (PEAVaR) by:

$$\text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{EVaR}(\hat{\theta}, \alpha) + \text{bias}(\hat{\theta}, \theta_0, \alpha). \tag{1}$$

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 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, and hence it is harder to model the shape of the tail than the shape of the interior distribution. As a consequence, it might be tempting tempted to forecast VaR closer to the center of the distribution, perhaps at $\alpha = 95\%$, and then use those estimation results to get at the VaR or the bias for more extreme probability levels, like $\alpha = 99\%$ or $\alpha = 99.9\%$. This is often referred to as *probability shifting*.

Therefore, we can Alternatively analyze the impact of such endeavors within our framework in defining two probabilities, α^* and α^{**}, such that:

$$\begin{cases} \text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{EVaR}(\hat{\theta}, \alpha^*) = \text{ThVaR}(\theta_0, \alpha^*) \\ \text{EVaR}(\hat{\theta}, \alpha^{**}) = \text{PEAVaR}(\hat{\theta}, \theta_0, \alpha) = \text{ThVaR}(\theta_0, \alpha), \end{cases} \tag{2}$$

or equivalently:

$$\begin{cases} \alpha^* = F^{-1}\left[\hat{F}^{-1}(\alpha)\right] \\ \alpha^{**} = \hat{F}^{-1}\left[F^{-1}(\alpha)\right]. \end{cases} \tag{3}$$

If one were to use α^* instead of α, the bias adjusted VaR would result, whilst α^{**} does the opposite, mapping the probability corresponding to the biased VaR, to the theoretical VaR. It follows that if $\alpha^* > \alpha > \alpha^{**}$ the estimated VaR is biased towards zero, whilst if $\alpha^* < \alpha < \alpha^{**}$ it is biased towards minus infinity.

This leaves the question, if we can identify the probability shifts could one then obtain more accurate VaR forecasts? In the best case scenario considered here, it might be possible since we know the true DGP. However, even in this case, it might not be feasible. Suppose the forecasted VaR is biased towards zero, where $\alpha = .99$

and we get $\alpha^{**} = .9999$. If the sample size is smaller than 100,000 and we employee the Historical Simulation method, there is just no empirical quantile to use.

Generally, the tendency of some the users, encouraged by the supervisors, to use relatively extreme probabilities, such as 99.5% in the European Solvency II framework may not be warranted on grounds of probabilities shift type arguments because it likely would lead to bigger model risk that if the probability levels were more moderate.

2.2 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 can not 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 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) such as:

$$\text{PSAVaR}(\theta_0, \theta_1, \alpha) = \text{EVaR}(\theta_1, \alpha) + \text{bias}(\theta_0, \theta_1, \alpha), \tag{4}$$

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

We can also define the relationship in terms of the probability shifts (as in (2)) such as:

$$\begin{cases} \text{PSAVaR}(\theta_0, \theta_1, \alpha^*) = \text{EVaR}(\hat{\theta}, \alpha) = \text{ThVaR}(\theta_0, \alpha^*) \\ \text{EVaR}(\theta_1, \alpha^{**}) = \text{PSAVaR}(\theta_0, \theta_1, \alpha) = \text{ThVaR}(\theta_0, \alpha). \end{cases} \tag{5}$$

2.3 Granularity error

Granularity error is caused by bias resulting from finite number of assets portfolios and then the resulting residual idiosyncratic risk, see e.g. 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, here 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 onto 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):

$$\text{PGAVaR}(\theta_0, \alpha, n) = \text{EVaR}(\theta_0, \alpha, N) + \text{bias}(\theta_0, \alpha, n), \tag{6}$$

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

We can also define the relationship in terms of the probability shifts (as in (2)) such as:

$$\begin{cases} \text{PGAVaR}(\theta_0, \alpha, n) = \text{EVaR}(\hat{\theta}, \alpha) = \text{ThVaR}(\theta_0, \alpha^*) \\ \text{EVaR}(\theta_1, \alpha^{**}) = \text{PGAVaR}(\theta_0, \alpha, n) = \text{ThVaR}(\theta_0, \alpha). \end{cases} \tag{7}$$

2.4 Measurement error

Financial data are prone to measurement errors caused by various phenomenon such as non-synchronous trading, rounding errors, infrequent trading, microstructure noise or insignificant volume exchanges. In addition, observed data is subject to manipulation (smoothing, extra revenues, fraudulent exchange, information less 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. (2010) extensively document the phenomenon 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 causes under-estimation of the true risk. Certain financial institutions report a contaminated P&L with extraneous profits such as intraday revenues, fees, commissions, net interest incomes and revenues from market making or underwriting activities.

2.5 Liquidity risk

While liquidity has many meanings, from the point of view of the risk forecasting, the most relevant is some aspects of market liquidity, as defined by the BCBS, such as the ability to quickly trade large quantities, at a low cost, and without impacting the price. These directly follow from Kyle's (1985) three dimensions of liquidity: tightness, depth, and resiliency[3]. As an extreme example, we can mention that NY stock exchange remained shuttered for more than four months at the beginning of

[3] 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.

the First World War (from the 31st of July 1914 to the 12th of December 1914) and that the re-opening brings the largest one-day percentage drop in the DJIA (24.4%).[4]

Measures for gauging illiquidity exposure of several asset classes are presented in Chan et al. (2006). 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. Giot and Grammig (2005), using weighted spread in intraday VaR-framework, show that accounting for liquidity risk becomes a crucial factor the traditional (frictionless) measures severely underestimate the true VaR.

3 Analysis of estimation and specification errors

Many of the potential sources of error discussed above can have a significant 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, is estimation and specification error. For this reason, we first investigate those two in detail why means of a Monte Carlo experiment, whereby we specify a general model (the DGP), and used this model to generate data. Secondly, 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 $\widehat{\theta}_1$. In this case, we can extend (1) and (4) and indicate the estimated VaR by $\mathrm{EVaR}(\widehat{\theta}_1, \alpha)$ and define the perfect model risk adjusted VaR (denoted herein PMAVaR) by:

$$\mathrm{PMAVaR}(\widehat{\theta}_1, \alpha) = \mathrm{ThVaR}(\theta_0, \alpha) - \mathrm{bias}(\theta_0, \widehat{\theta}_1, \alpha). \tag{8}$$

3.1 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 back in small samples. The DGP is a second order Markov-Switching Generalized AutoRegressive Conditionally Heteroskedastic with t-student disturbances[5] (hereafter

[4] See e.g., Silber (2005).

[5] As a complement (not reported here for space reasons, but available on demand in a web appendix), we also used in preliminary tests other alternatives frameworks: a Student *versus* a Normal density, as well as a Brownian, Lévy and Hawkes processes, with the same qualitative response with a relative model error for VaR ranging form 5-15% in the simplest cases (Gaussian estimation risk with 250 observations) to as large as 200% when the process in complicated and the

denoted MS(2)-GARCH(1,1)-t).[6]

We first simulate a long artificial series of 360,000 daily returns with estimated parameters on the daily DJIA from 01/01/1990 to 09/20/2011[7]. We then forecast various VaRs using 1,000 observations, and finally, computing statistical indications of the forecast error, measured by the difference between the asymptotic VaR (computed with the true simulated DGP on 360,000 observations) and the empirical ones recovered from limited samples.

The DGP is:

$$r_t = \mu_{s_t} + \sigma_t z_t, \tag{9}$$

with: $z_t \rightsquigarrow \text{iidSt}(0,1,\upsilon)$ and $\sigma_t^2 = \omega_{s_t} + \alpha_{s_t}\varepsilon_{t-1}^2 + \beta_{s_t}^2\sigma_{t-1}^2$, where $s_t \in \{1,2\}$ characterizes the state of the market, $\mu_{s_t}$is the mean return and the unscaled return innovation return factor following a centered standard Student 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 fat tails of a Student 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 surely strictly positive.

3.2 Simulation results

We will compute in what follows, the annualized daily 95%, 99% and 99.5% VaR, assuming a particular DGP for the returns.

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 a normal GARCH(1,1) and a MS(2)-GARCH(1,1)-t simulations and estimates. The columns represent respectively the average adjusted VaR according to specification and/or estimation errors, the theoretical VaR, the average of the adjustment term, the minimum value of the adjustment term, the maximum value of the adjustment term. Note that a positive adjustment term indicates that the estimated VaR (negative return) should be more conservative (more negative).

We present the estimation bias (denoted bias$(\theta_0, \widehat{\theta}, \alpha)$ as previously introduced in equation (1)), in Panel A of Table 1, when we simulate a simple model (Normal

sample small (case of Hawkes processes).

[6]see Frésard et al. (2011) in a VaR context, we hereafter use a MS(2)-GARCH(1,1)-t (see Hamilton and Susmel, 1994; Gray, 1996; Klaassen, 2002; Haas et al., 2004) for more details on the process.

[7]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. (2007) and Billio et al. (2010) obtain similar results on the S&P.

GARCH(1,1)) and use the appropriate methodology for computing the VaR (Normal GARCH-VaR). This bias arises only due to the limited number of observations (1000), and is zero for the 360,000 simulations. But the dispersion of this estimation bias is quite important since the minimum and the maximum values of the bias (or adjustment term) represent about 50% of the "true" or "perfect" VaR. For example, with $\alpha = 99\%$, the minimum and maximum biases are respectively equal to -33% and +32% for a perfect VaR of -60%.

The specification bias (denoted 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 important here, positive and increases 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.

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 add up in an independent manner. We compute the global error - denoted, in its most general formulation, bias$(\theta_0, \theta_1, \widehat{\theta}_1, \alpha)$ - as the difference between the "true" theoretical 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) is used to estimate a MS(2)-GARCH(1,1)-t, the average bias is positive and increases with α. The mean errors are thus equivalent to the specification bias component, but the dispersion of the "model risk"realizations is increased by the estimation bias.

Table 1: Illustrations of Conditional Simulated Errors associated to the 95%, 99% and 99.5% VaR: GARCH(1,1) *versus* MS(2)-GARCH(1,1)-t

	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%	-19.60%	19.53%
$\alpha = 99.00\%$	-59.70%	-59.70%	.00%	-.04%	-32.02%	32.66%
$\alpha = 99.50\%$	-70.99%	-70.99%	.00%	-.06%	-38.03%	38.35%

	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%

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January 1900 to the 20th September 2011; computations by the authors. This statistics were computed with the results on 360,000 simulated series of 1000 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 of the adjustment term, the minimum value of the adjustment term, the maximum value of the adjustment term. A*Per* convention, a negative adjustment term in the table indicates that the Estimated VaR (negative return) should be more conservative (more negative). Panel A relates GARCH(1,1) DGP with estimated GARCH-VaR results; Panel B presents true MS(2)-GARCH(1,1) DGP with estimated GARCH-VaR ones; Panel C reports estimated MS(2)-GARCH(1,1) DGP with estimated GARCH-VaR results.

Regarding now the probability shifts, the impact of model risk is captured in Table 2, which presents the two modified probability levels α^* and α^{**}. The former is associated to the "true" density and corresponds to the (mis-)estimated α-VaR, whilst the latter, associated to the estimated VaR, corresponds to the α-VaR without model error.

The gap between α^* and α can be interpreted as a measure of the model risk of the risk model, which should be founded in a context of parameter uncertainty. The gap between α^{**} and α can also be analyzed as the probability shift that we should apply using a specific model of VaR to reach the "true" theoretical VaR.

This alternative representation of the model risk of risk models shows that α^{**} is often unreachable to allow us to correct the estimated VaR. For instance, the maximum associated to the 99.5% VaR in Panel C has to be superior to 100% which can not be in practice discriminated from the maximum (*i.e.* associated to the 100% probability). More generally, α^{**} is very often superior to α, (and α^* 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 tendency of regulators to recommend more extreme quantiles, *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, surprising to us, that the mean bias is not an 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 however 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 unknown by definition. The proposed adjustments are thus impossible to be accurately quantified outside a pure academic simulation exercise. But we might have an idea of the minimum needed adjustment for meeting the regulation requirements. In other words, errors on pricing or return modelling and biases in the empirical estimation lead to a model error on the VaR that we can try to approximate thanks to its empirical economic consequences as proposed in the following section.

Table 2: Illustrations of Probability Shifts associated to the 95%, 99% and 99.5% Annualized VaR: GARCH(1,1) *versus* MS(2)-GARCH(1,1) Quantiles

	Probability α^* associated to the true density corresponding to the (mis-)estimated VaR				*Probability α^{**} 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.

Source: Bloomberg; daily data of the DJIA index in USD from the 1st January 1900 to the 20th September 2011; computations by the authors. This statistics were computed with the results on 360,000 simulated series of 1000 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 parametric VaR. The columns represent respectively the average Estimated VaR according with specification or/and estimation errors, the mean-minimum-maximum of the modified probability level α^*, the mean-minimum-maximum of the modified probability level α^{**}. The letters N.R. stands for "Not Reached", *i.e.* condition on bounds is not met even for 100.00%. Panel A relates GARCH(1,1) DGP with estimated GARCH-VaR results; Panel B presents true MS(2)-GARCH(1,1) DGP with estimated GARCH-VaR ones; Panel C reports estimated MS(2)-GARCH(1,1) DGP with estimated GARCH-VaR results.

4 A simple economic valuation of model risks

While it is not possible to optimally adjust for biases such as those caused by estimation and specification, we can approximate these biases by adjusting the VaR forecasts by the performance of the same model historically. In other words, historical errors are used to adjust future forecasts.

To this end, we define the imperfect model adjusted VaR (IMAVaR) as:

$$\text{IMAVaR}(\hat{\theta}_1, \alpha, n) = \text{EVaR}(\hat{\theta}_1, \alpha, N) + \text{adj}(\theta_0, \theta_1, \hat{\theta}_1, \alpha, n), \tag{10}$$

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

$$\text{IMAVaR}(\hat{\theta}_1, \alpha, n) = \underbrace{\sup}_{\text{VaR} \in \text{IR}} \{\text{VaR}(\alpha, n)^*\}, \tag{11}$$

where VaR$(\cdot)^*$ is a set of VaRs, from a model approved by the supervisor, and IMAVaR$(\cdot)$ is the limit highest VaR (the less conservative VaR) such that the supervisor still validates the model.

Generally speaking, the better the VaR model and the lower the minimum required adjustment and *vice-versa*. We now have to explicit the limit VaR that bounds the IMAVaR.

4.1 General backtest procedures

A variety of tests have been proposed in the literature to gauge the accuracy of VaR estimates, based on the three good properties that we should expect from a risk model such as the right frequency of violations, the independence of hits and the restricted magnitude of exceptions. We will use some of the best knowns for the adjustments without loss of generality. We briefly mentioned them in this section, and provide details in Appendices A.

The first VaR test for a good VaR is the unconditional coverage test (Kupiec, 1995), based on observed number of violations of VaR, compared to that expected. If we assume that the $I_t^{\text{EVaR}}(\cdot)$ variables are independently and identically distributed, then, under the unconditional coverage hypothesis (Kupiec, 1995), the total number of VaR exceptions follows a Binomial distribution (Christoffersen, 1998), denoted $B(T, \alpha)$, such as:

$$\text{Hit}_t^{\text{EVaR}(\cdot)}(\alpha) = \sum_{t=1}^{T} I_t^{\text{EVaR}(\cdot)}(\alpha) \sim> B(T, \alpha). \tag{12}$$

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 called independence of exceptions hypothesis. In particular the Christoffersen (1998) test can be written such as:

$$\text{LRind}^{I_t^{\text{EVaR}}(\alpha)} = 2\left[\log L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11}) - \log L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi)\right] \sim> \chi^2(1), \qquad (13)$$

where $\pi_{ij} = \Pr\left[I_t^{\text{EVaR}}(\alpha) = j | I_{t-1}^{\text{EVaR}} = i\right]$ is Markov chain reflects the existence of an order 1 memory in the process $I_t^{\text{EVaR}}(\alpha)$, $L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11})$ 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 $\pi_{01} = \pi_{11} = \pi$.

A third class of tests focuses on the magnitude of the losses experienced when VaR estimates are exceeded. The underlying idea is that a small violation might acceptable but that a large one can lead to bankruptcy. Berkowitz (2001) for instance proposes a hypothesis test for determining whether the magnitudes of observed VaR exceptions (γ_{t+1}) are consistent with the underlying VaR model, such as:

$$\text{LRmag}^{\gamma_{t+1}} = 2\left[L_{\text{mag}}^{\gamma_{t+1}}(\mu, \sigma) - L_{\text{mag}}^{\gamma_{t+1}}(0, 1)\right] \sim> \chi^2(2), \qquad (14)$$

where $L_{\text{mag}}^{\gamma_{t+1}}(\cdot)$ is the likelihood function assoiciated to the parameters μ and σ.

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

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

4.2 A good VaR and the backtests

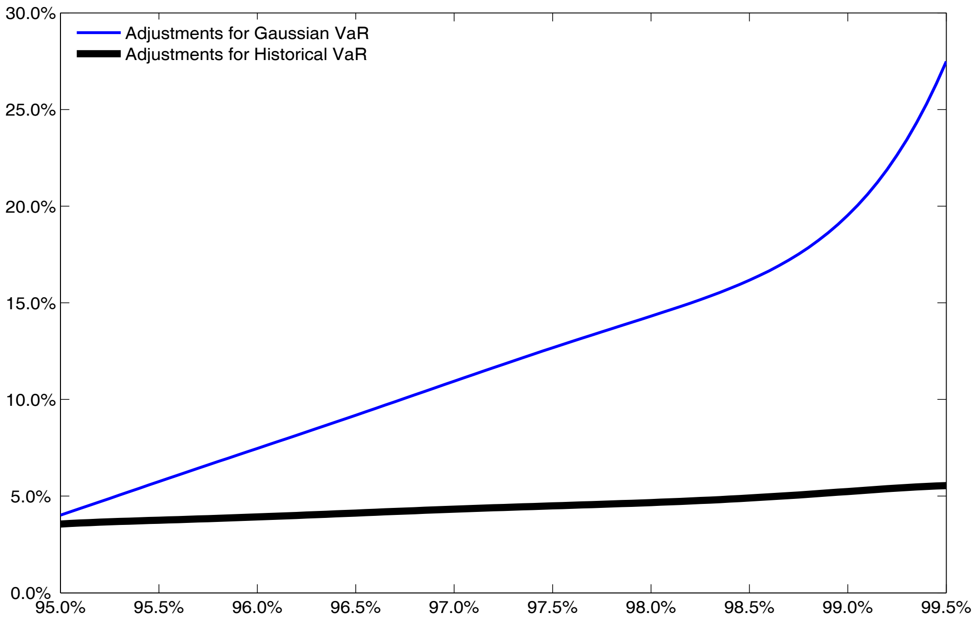
For a given VaR forecast and the bounding range for the tests above, we can obtain the IMAVaR that respects conditions (12), (13) and/or (14) (or their subsampled versions) such as (with previous notations):

$$\text{adj}(\hat{\theta}_1, \theta_1, \theta_0, n) = \underline{q}^* = \underbrace{\max}_{q^* \in \text{IR}} \{\text{VaR}(\alpha, n)_t^*\}$$
$$\text{s.t.:}$$
$$\begin{cases} \text{Hit}_t^{\text{VaR}(\cdot)^*}(\alpha) \sim> B(T, \alpha) \text{ for the “Hit” test} \\ \text{(and/or)} \\ \text{LRind}_t^{\text{VaR}(\cdot)^*}(\alpha) \sim> \chi^2(1) \text{ for the “Independence” test} \\ \text{(and/or)} \\ \text{LRmag}_t^{\gamma_{t+1}}(\alpha) \sim> \chi^2(2) \text{ for the “Exception Magnitude” test,} \end{cases} \qquad (15)$$

with: $\text{VaR}(\alpha, n)_t^* = \text{EVaR}(\hat{\theta}_1, \alpha, N)_t + q^*$.

As a first illustration, Figure 3 represent the minimum adjustments (errors), denoted $\underline{q}^*$ as solutions of the program (15), when first only considering the hit test, for both one-year historical and Gaussian VaR computed on the DJIA on one century of daily data.

Figure 3: Minimum Model Risk Absolute Adjustment for the Hit Test associated to Historical and Gaussian VaR on the DJIA according to the Level of Confidence



Source: Bloomberg; daily data of the DJIA index in USD from the 1st January 1900 to the 20th September 2011; computations by the authors. This figure represents on the y-axis the minimal absolute adjustment (in percent of the underlying VaR) necessary to respect the hit ratio *criterion* according 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 VaR is here computed on a weekly horizon as a parametric Gaussian quantile using 5 years of past returns to estimate the parameters.

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 time (with full information at time T). We can see here that the comparison of the two methods is in favour of the historical method since the error is lower (around 4%) for all quantiles and thus rather independent of the confidence level.

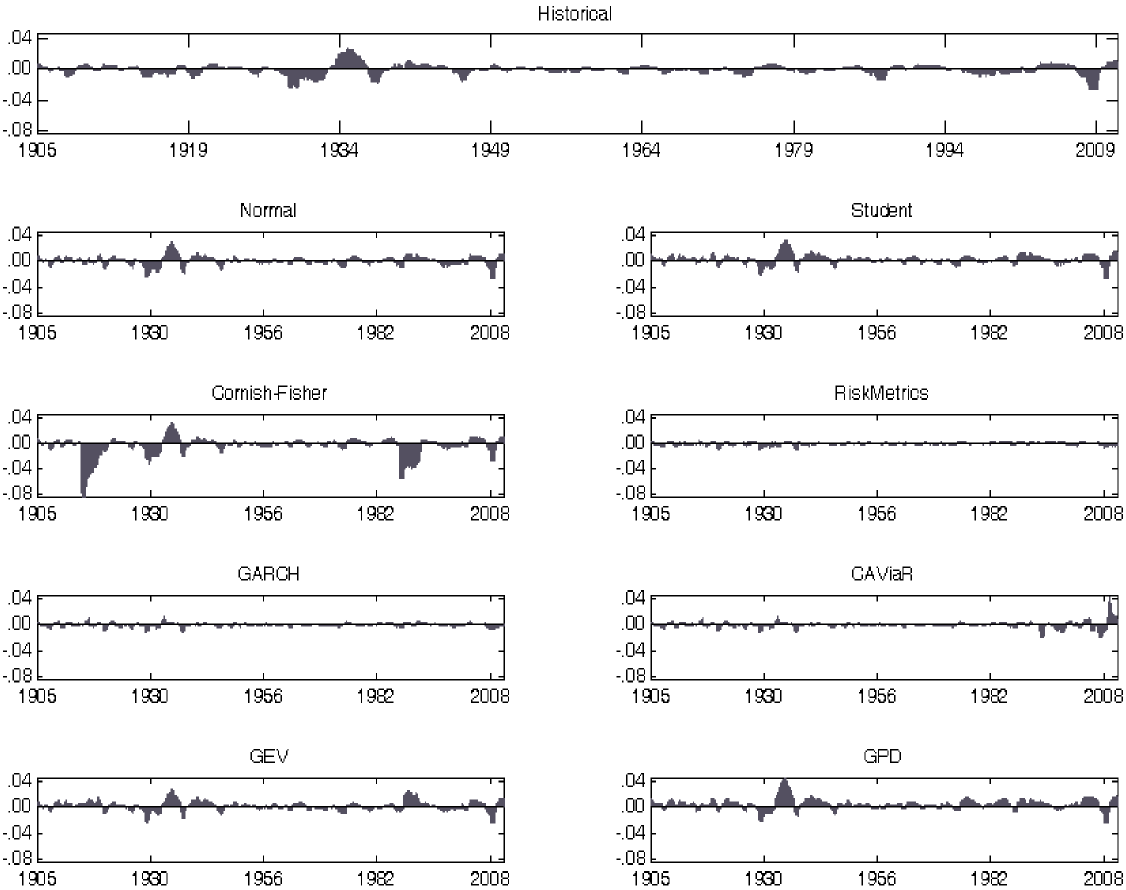
4.3 VaR model comparisons

We apply the general adjustment method presented above, obtained for the daily DJIA index from the 1st of January 1900 until the 2nd of March 2011, or 29,002 daily returns. We use a moving window of four years (1,040 daily returns) to re-estimate dynamically parameters for the various methods. Forecasted VaR are computed for each method for the final 29,957 days (about 108 years). This comparison considers daily estimation of the 95%, 99% and 99.5% conditional VaR.

This leaves the choice of VaR forecast method. While there is an almost infinite number of techniques that could be used, we restrict ourselves to the most common in practice. Historical simulation, parametric approaches, based on Gaussian or t-Student return distributions, with a compromise with the modified VaR using a statistical expansion (Cornish-Fisher VaR). We also employ three dynamic methods, namely RiskMetrics, GARCH(1,1) and CAViaR (Engle and Manganelli, 2004). Finally, we complement the these methods by using two extreme densities for the returns such as the GEV distribution and the GPD.

Figure 4 represents the dynamic required corrections corresponding to the Hit test for the various risk models on the DJIA. These corrections are the daily correction to pass the hit test over the past one-year of daily returns (on the period from $t-250$ to t). The magnitude can be sometimes large (specifically around the 1929 and 2008 crises), ranging from 0 to 15% (for RiskMetrics) 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. In other words, except for the CAViaR method, the final adjustment is obtained just after 1934 (and does not vary after). Dynamic measures, such as RiskMetrics, GARCH and CAViaR, also demonstrate some superiority over unconditional parametric methodologies.

Figure 4: Dynamic Optimal Adjustment on the Daily VaR 95% related to the Hit Test for different VaR models

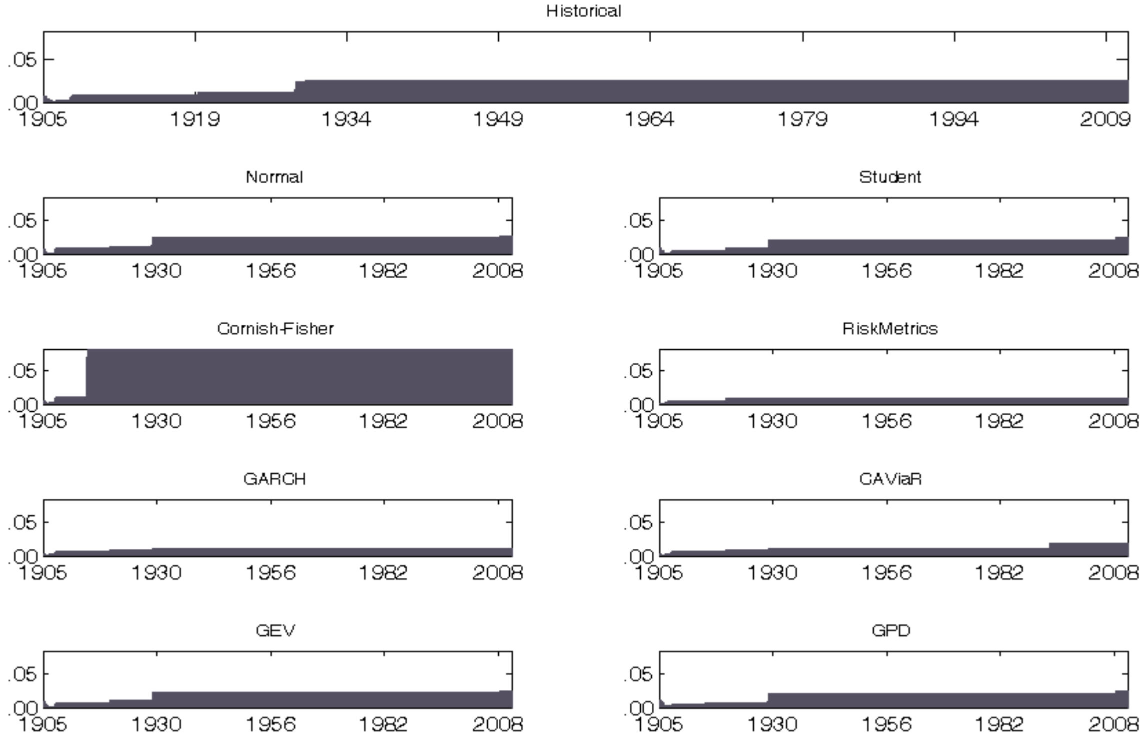


Source: Bloomberg; daily data of the DJIA index in USD from 1st January 1900 to the 20th of September 2011; computations by the authors. We use a moving window of four years (1,040 daily returns) to dynamically re-estimate parameters for the various methods.

Figure 5 illustrates the evolution of the maximum required corrections for all VaR methods under consideration (maxima of correction needed from 0 to t represented by Figure 4). This is for the Hit test, from the general program aiming to correct today's VaR with the maximum correction that has been necessary since the beginning of the series (expressed here in relative terms compared to the level of VaR).

The magnitude of the correction is at the end large in general, ranging from 15% (RiskMetrics) to more than 100% (for the Cornish-Fisher VaR). We also see that, except for the CAViaR method, the final adjustment is obtained just after 1934 and do not vary a lot after. Corrections for dynamic measures once again quite rapidly reach their required maximum and are thus quite stable overall.

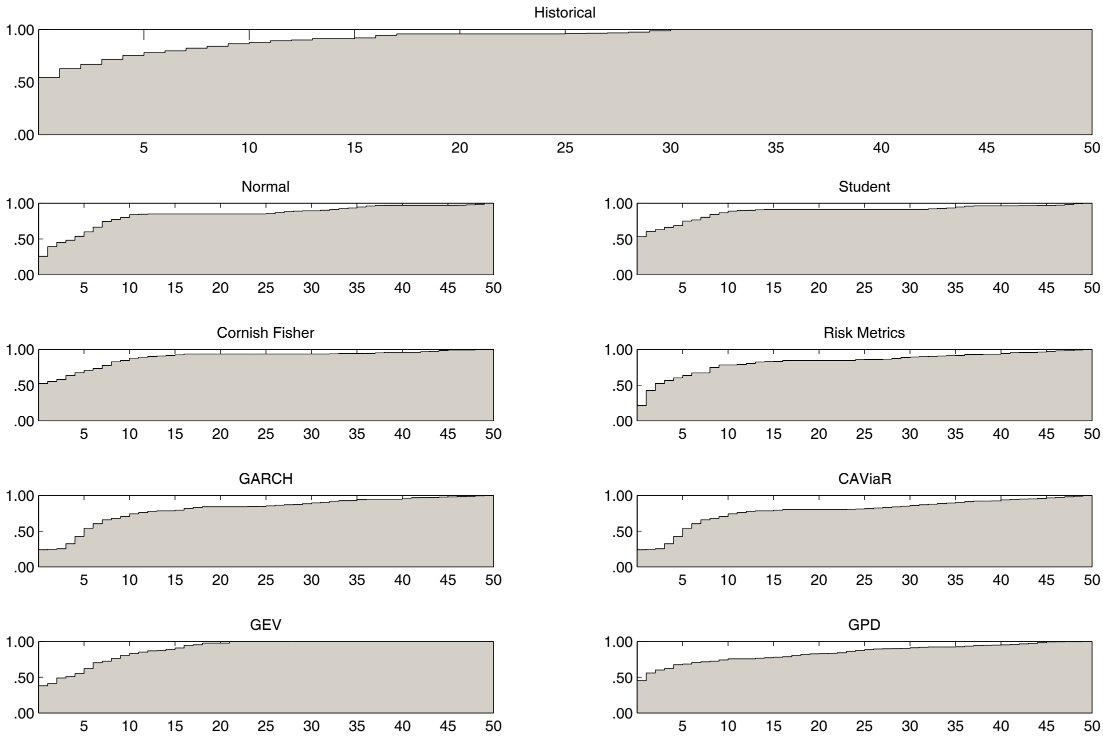
Figure 5: Optimal Dynamic Absolute Value of Minimum Negative Adjustments for the Hit Test for different VaR Measures at 95%



Source: Bloomberg; daily data of the DJIA index in USD from 1st January 1900 to the 20th of 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.

Figure 6 illustrates the minimum dynamic adjustment required for passing the hit test for a randomly chosen the first date of implementation. More precisely the exercise consists in choosing a first date and then computing the dynamic adjustment until the end of the sample and redoing the exercise 30,000 time whilst keeping, for each horizon, the minimum correction obtained. The optimal adjustments are here expressed in terms of percentage of their maximum value over the whole sample. The figure shows that depending on the VaR method, the horizon for having the major part of the maximum correction varies from 18 years (GEV) to 46 years (CAVIAR). Moreover, regardless of the model, the major part (80% or so) of the correction is reached after 10 years. That means that, whatever the VaR model, the most of the highest surprises have been faced after a decade of history (even in the worst scenario when the sample is amongst the less turbulent ones). In other words, at least ten years are needed to have a fairly good idea of the magnitude of the required correction.

Figure 6: Optimal Dynamic Relative Adjustment for the Hit Test for different Starting Dates and VaR Measures at 95% by Horizon (in years)



Source: Bloomberg; daily data of the DJIA index in USD from 1st January 1900 to the 20th of 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. 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 (and not only the hit test) as a generalization of the approach by Kerkhof et al. (2010) taking into account several qualities of VaR models. Table 3 reports the various maximum required corrections related to the three main categories tests (altogether with their Escanciano and Olmo, 2010, bootstrapped corrected versions). We first note that the hit test is less permissive when the corrections implied by the bootstrapped critical values are made, whilst the tests of independence and normality imposed very severe corrections (of order of 100% in relative terms for some tests).

The exception frequency test, *i.e.* the unconditional coverage test by Kupiec 1995 at a 5% level): the RiskMetrics is the best model for estimating the DJIA Index 95% VaR. Following the same argument, GARCH VaR, then GEV VaR model come just after. When the dynamics of hit are considered, the conditional methods perform better for the second correction (with the best result for the GARCH model). Finally, when we are interested in the magnitude of the violations - the most severe test - once again the dynamic measures show some superiority, whilst extreme density VaR exhibit some weaknesses.

Table 3: Maximum Negative Adjustment Values for 95% Daily Value-at-Risk Models

VaR Methods	Mean VaR	q_1	q_1^*	q_2	q_2^*	q_3	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%
Cornish-Fisher	-1.26%	-8.29%	-7.48%	-8.40%	-8.86%	-8.40%	-8.86%
RiskMetrics	-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%

Source: Bloomberg; daily data of the DJIA index in USD from the 01/01/1900 to the 09/20/2011; 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 Hit test; q_1^* to Escanciano and Olmo (2010) unconditional test; q_2 to independence test; q_2^* to Escanciano and Olmo (2010) independence test, q_3 to the magnitude test and q_3^* lies to the Bootstrap sampled magnitude test.

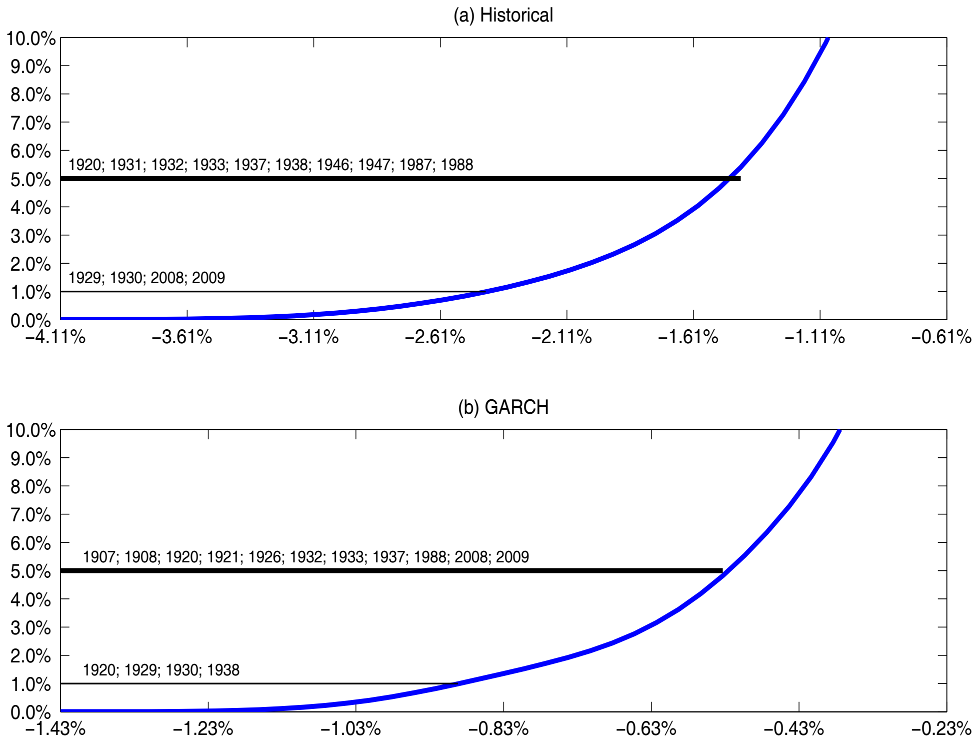
4.4 Generalized model risk of model risk

These results have particular implications on how one might approach adjustments. First, the link between the three implicit levels of confidence required to be defined; the confidence 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 affect to the solidity of the buffer.

Typically, a high incremental buffer leads a high protection against the model risk and the extreme events on the market. But a reduced buffer decreases the insurance against these major turbulent episodes and then the failures of risk models.

Figures 7 below illustrates this link between the level of the buffer, translated the protection against the more severe crises and the degree of confidence associated to the buffer. These Figures represent the cumulative density functions of required adjustments (on the last century on 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 95%. We can see that if we allow a 5% model risk, we are, unsurprisingly, not protected against the 5% biggest shocks on the data (such as 1929, 1930, 2008 and 2009).

Figure 7: Empirical Cumulative Density Function of Optimal Adjustment Values for the Hit Test of a 95% Daily Historical and GARCH Value-at-Risk



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) for computing the VaR. The threshold for the Hit test us fixed at 95%. We have used a Gaussian Kernel Smoothing Density (see Bowman and Azzalini, 1997).

The second generalization concerns the recommended stress VaR on various profiles of investment portfolios. We can express the confidence level related to model risk using a stress-VaR methodology (*i.e.* imposing VaR levels for each investment class and a variance-covariance matrix on the underlying assets), as well as the implicit 100% model risk free k ratio - that is to be applied to the VaR for computing the capital requirement for financial institutions. In other words, one can evaluate the accuracy of the k factor used by regulator for fixing a safe capital (k being between 3 and 5, applied to the VaR for calculating the capital requirement) when compared to the worst forecast realizations (measured by the ratio of the largest adjustment out of the VaR forecast on a specific day), when following the recommended VaR methodology based on defined stresses on sub-components of various portfolios.

Table 4: Minimum Annualized Model Risk for a 95% GARCH-VaR for all 5% Validity Tests on various Portfolios

Portfolio	q_1	q_1^*	q_2	q_2^*	q_3	q_3^*
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%
Bonds	-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%

Source: DataStream and Bloomberg; daily data from the 31$^{\text{th}}$ December 1986 to the 28$^{\text{th}}$ November 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. "Defensive Profile" refers to 10% Bonds+90% Liquidity; "Balanced Profile" refers to 30% Equity+10% Real Estate +10% Commodity + 40% Bonds + 10% Liquidity; and "Aggressive Profile" refers to 70% Equity + 15% Real Estate + 15% Commodity. The variable q_1 refers to Hit test; q_1^* to Escanciano and Olmo (2010) unconditional test; q_2 to independence test; q_2^* to Escanciano and Olmo (2010) independence test, q_3 to the magnitude test and q_3^* lies to the Bootstrap sampled magnitude test.

Table 5: Maximum k ratio Model Risk Confidence Levels for a 95% GARCH-VaR for all 5% Validity Tests on various Portfolios

Portfolio	q_1	q_1^*	q_2	q_2^*	q_3	q_3^*
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
Bonds	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
Equilibrate Profile	1.45	1.42	1.57	1.63	2.52	1.84
Aggresive Profile	1.42	1.38	1.38	1.52	2.58	1.57

Source: DataStream and Bloomberg; daily data from the 31$^{\text{th}}$ December 1986 to the 28$^{\text{th}}$ November 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. "Defensive Profile" refers to 10% Bonds+90% Liquidity; "Balanced Profile" refers to 30% Equity+10% Real Estate +10% Commodity + 40% Bonds + 10% Liquidity; and "Aggressive Profile" refers to 70% Equity + 15% Real Estate + 15% Commodity. The variable q_1 refers to Hit test; q_1^* to Escanciano and Olmo (2010) unconditional test; q_2 to independence test; q_2^* to Escanciano and Olmo (2010) independence test, q_3 to the magnitude test and q_3^* lies to the Bootstrap sampled magnitude test.

Table 6: Maximum k ratio Model Risk Confidence Levels of 95% Stress-test for 5% Validity Tests on various Portfolios -Annualized Values

Portfolio	q_1	q_1^*	q_2	q_2^*	q_3	q_3^*
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	.89	.89	1.27	.92	1.83	1.19
Bonds	.51	.50	.52	.56	1.04	.76
Defensive Profile	.11	.11	.11	.12	.20	.16
Equilibrate Profile	2.63	2.50	1.54	2.63	5.10	3.81
Aggresive Profile	3.08	2.94	1.83	3.78	6.10	4.94

Source: DataStream and Bloomberg; daily data from the 31$^{\text{th}}$ December 1986 to the 28$^{\text{th}}$ November 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 stress-test consists of a shock on all asset class: -30% for Equity; -40% for Real Estate; -30% for Commodity and -20%. "Defensive Profile" refers to 10% Bonds+90% Liquidity; "Balanced Profile" refers to 30% Equity+10% Real Estate +10% Commodity + 40% Bonds + 10% Liquidity; and "Aggressive Profile" refers to 70% Equity + 15% Real Estate + 15% Commodity. The variable q_1 refers to Hit test; q_1^* to Escanciano and Olmo (2010) unconditional test; q_2 to independence test; q_2^* to Escanciano and Olmo (2010) independence test, q_3 to the magnitude test and q_3^* lies to the Bootstrap sampled magnitude test.

5 Conclusion

We propose incorporating model risk into risk measure calculations by constructing classes of models on the basis of standard econometric procedures in a backtesting framework. We distinguish between several stages of modelling which each contributes to the model risk that leads to different adjustments of a chosen risk measure. In this way, we define several components of model risk which we refer to as estimation risk, specification risk, granularity, data contamination and liquidity risk. We then evaluate their effect by their indirect impact on the properties of a good VaR model, while focusing on the estimation and specification uncertainties.

Our main objective is to account for some dimensions of the riskiness of risk models and adjust consequently computed risk measures. We show, for some assets and some realistic simulations, how fragile VaR estimates are, not only with respect to the pure choice of the risk model, but also regarding the choice of the underlying econometric model and of the appropriate sampling period.

Thus, our aim is to show the extent to which (quantile) risk measures are affected by model risk and to propose a practical method to account for this risk of risk measures, explicitly recognizing that identifying and measuring model risk lead to better measurement methods. This adjustment, depending on the efficiency of the risk model, allows the risk manager to jointly treat theoretical and estimation risks, taking into account their possible dependence, and can be seen as a measure of the model risk of a risk model called in Basel III regulation.

We, finally, use our framework for a simple comparison of main parametric, semi-parametric and non-parametric VaR models following the simple principle that the lower the required adjustment, the lower the model risk and, therefore, the better the model[8]. We also complement our approach in identifying the model risk of model risk of risk model, by associating a degree of confidence on the correction based on the distribution of past violations. Three levels of confidence are here needed: the level of VaR, the severity of tests and the trust we want to put into the buffer.

Our approach, however, only considers the model risk at a micro level since we do not consider the endogeneity of risk at the macro level. Indeed, external and uniform regulatory risk constraints may lead to the amplification of a crisis by reducing liquidity (Daníelsson, 2002). This systemic endogenous risk reflects the fact that the measurement of risk has not fully taken into account the possible domino macro-effect of a crisis in violent market turbulence.

[8] Note that the model with the lower needed adjustment (i.e, the better model), is more profitable for a financial institution since it implies to take more risk in good times and less risk in bad times.

Our work can be extended in several ways. Our general correction framework can be used when comparing the various tests on a good VaR quality proposed in the literature (Berkowitz et al., 2010). The second extension could be to apply some specific VaR model 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 (Hoogerheide et al., 2011) or several quantile levels (Colletaz et al., 2011). Another research proposal would be in adopting the same approach leading to an estimated multi-VaR, built as a portfolio of various VaR models, directly aiming to minimize the model risk (McAleer et al., 2011). Finally, using the same metric of corrections, the quality of other VaR-based measures in a context of systemic risk measures (such as the Marginal Expected Shortfall or the CoVaR) would be worth considering (e.g. Daníelsson et al., 2011b; Benoit et al., 2011; Löffler and Raupach, 2011).

References

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

Alexander C. and J.-M. Sarabia, (2011), "Model Risk of VaR", *SSRN Working Paper*, 24 pages.

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

Basel Committee on Banking Supervision, (1988), *International Convergence of Capital Measurement and Capital Standard*, Bank for International Settlements, 30 pages.

Basel Committee on Banking Supervision, (1996), *Amendment to the Capital Accord to Incorporate Market Risks*, Bank for International Settlements, 63 pages.

Basel Committee on Banking Supervision, (2001), *The New Basel Capital Accord*, Consultative Document of the Bank for International Settlements, 139 pages.

Basel Committee on Bank Supervision, (2006), *Core Principles for Effective Banking Supervision*, Bank for International Settlements, 13 pages

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.

Berkes I. and L. Horváth, (2004), "The Efficiency of the Estimators of the Parameters in GARCH Processes", *The Annals of Statistics 32(2)*, 633-655.

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, (2010), "Evaluating Value-at-Risk Models with Desk-Level Data", forthcoming in *Management Science*, 38 pages.

Benoit S., G. Colletaz and C. Hurlin, (2011), "A Theoretical and Empirical Comparison od Systemic Risk Measures: MES *versus* CoVaR", *SSRN Working Paper.*

Boucher C. and B. Maillet, (2011), "The Riskiness of Risk Models", Working Paper, University of Paris-1, 8 pages.

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

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

Campbell J.Y., A.W. Lo and C. MacKinlay (1997), *The Econometrics of Financial Markets,* Princeton University Press, 632 pages.

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

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

Colletaz G., Ch. Hurlin and Ch. Pérignon, (2010), "Backtesting Risk Models using both Number and the Magnitude of VaR Exceptions", *mimeo*, 25 pages.

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

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

Daníelsson J., (2011), *Financial Risk Forecasting*, John Wiley, 274 pages.

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.

Engle R. and S. Manganelli, (1999), "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 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-a), "Backtesting Parametric Value-at-Risk with Estimation Risk", *Journal of Business and Economic Statistics 28(1)*, 36-51.

Escanciano J. and J. Olmo, (2010-b), "Robust Backtesting Test for Value-at-Risk Models", forthcoming in *Journal of Financial Econometrics*, 39 pages.

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, (2010), "The Pernicious Effects of Contaminated Data in Risk Management", *Journal of Banking and Finance 35(10),* 2569-2583.

Gagliardini P. and Ch. Gouriéroux, (2010), "Granularity Adjustment for Risk Measures: Systematic *versus* Unsystematic Risks", *Proceedings of CREDIT_2010 Conference*, 60 pages.

Gagliardini P., Ch. Gouriéroux and A. Monfort, (2010), "Micro-information, Non-linear Filtering and Granularity", *Proceedings of CREDIT_2010 Conference*, 56 pages.

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.

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, (2010), "Estimation Adjusted VaR", *mimeo CREST*, 39 pages.

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)*, 17-38.

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.

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.

Löffler G. and P. Raupach, (2011), "Robustness and Informativeness of Systemic Risk Measures", *mimeo*, 28 pages.

Longin F., (2000), "From VaR to Stress Testing: The Extreme Value Approach", *Journal of Banking and Finance 24(7)*, 1097-1130.

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

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

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

Martins-Filho C. and F. Yao, (2006), "Estimation of Value-at-Risk and Expected Shortfall based on Non-linear Models of Return Dynamics and Extreme Value Theory", *Studies in Nonlinear Dynamics & Econometrics 10(2)*, 41 pages.

McAleer M., J.-A. Jiménez-Martín and T. Pérez-Amaral, (2011), "International Evidence on GFC-robust Forecasts for Risk Management under the Basel Accord", *KIER Working Papers 757*, Kyoto University, Institute of Economic Research, 40 pages.

McAleer M., J.-A. Jiménez-Martín and T. Pérez-Amaral, (2009), "The Ten Commandments for Managing Value-at-Risk under the Basel II Accord", *Journal of Economic Surveys 23(5)*, 850-855.

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

Pérignon Ch. and D. Smith, (2008), "A New Approach to Comparing VaR Estimation Methods", *Journal of Derivatives 16(2)*, 54-66.

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.

Schulte-Herbrüggen W. and G. Becker, (2005), "From Basel II to Basel III", *Risk 18(1)*, 58-61.

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

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

A Appendix: Main Backtest Procedures

A variety of tests have been proposed in the literature to gauge the accuracy of VaR estimates, based on the three good properties that we should expect from a risk model such as the right frequency of violations, the independence of hits and the restricted magnitude of exceptions. The first test for a good VaR is the so-called "traffic light" approach in the regulatory framework is 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 of the *ex ante* VaR). The so-called "Hit variable" associated to the *ex post* observation of EVaR(·) violations at the threshold α and time t, denoted $I_t^{\text{EVaR}}(\alpha)$, is defined such as (with previous notations):

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

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

If we assume that the $I_t^{\text{EVaR}}(\cdot)$ variables are Independently and Identically Distributed, then, under the Unconditional Coverage hypothesis (Kupiec, 1995), the total number of VaR exceptions (Cumulated Hits) follows a Binomial distribution (Christoffersen, 1998), denoted $B(T, \alpha)$, such as:

$$\text{Hit}_t^{\text{EVaR}(\cdot)}(\alpha) = \sum_{t=1}^{T} I_t^{\text{EVaR}(\cdot)}(\alpha) \sim> B(T, \alpha). \tag{A.2}$$

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 (A.2).

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 called independence of exceptions hypothesis. If the exceptions exhibited some type of "clustering", then the VaR model may 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:

$$\Pi = \begin{pmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{pmatrix}, \tag{A.3}$$

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:

$$\text{LRind}^{I_t^{\text{EVaR}}(\alpha)} = 2\left[\log L^{I_t^{\text{EVaR}}(\alpha)}(\pi_{01}, \pi_{11}) - \log L^{I_t^{\text{EVaR}}(\alpha)}(\pi, \pi)\right] \sim> \chi^2(1), \quad \text{(A.4)}$$

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 $\pi_{01} = \pi_{11} = \pi$ such 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 (A.4).

A third class of tests focuses on the magnitude of the losses experienced when VaR estimates are exceeded. The underlying idea is that a small violation might acceptable but that a large one can lead to bankruptcy. In other words, not only the number of violations should be under scrutiny, but also super-exceptions. Berkowitz (2001) proposes a hypothesis test for determining whether the magnitudes of observed VaR exceptions are consistent with the underlying VaR model. The key intuition is that VaR exceptions are treated as continuous random variables and not converted into the indicator variable used for the coverage tests. For this test, Berkowitz (2001) transforms the empirical series into standard normal z_{t+1} series. If the observed quantile q_{t+1} with the distribution forecast f_{t+1} for the observed portfolio return r_t, is defined as:

$$q_{t+1} = \int_{-\infty}^{r_{t+1}} f_{t+1}(r)dr. \quad \text{(A.5)}$$

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

$$z_{t+1} = \Phi^{-1}(q_{t+1}), \quad \text{(A.6)}$$

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), should equal $(0, 1)$, such as:

$$\gamma_{t+1} = \begin{cases} z_{t+1} \text{ if } z_{t+1} < \Phi^{-1}(\alpha) \\ 0 \text{ otherwise,} \end{cases} \tag{A.7}$$

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

Finally, the corresponding test statistic is:

$$\text{LRmag}^{\gamma_{t+1}} = 2\left[L_{\text{mag}}^{\gamma_{t+1}}(\mu,\sigma) - L_{\text{mag}}^{\gamma_{t+1}}(0,1)\right] \sim> \chi^2(2), \tag{A.8}$$

where:

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

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 (A.8).

For both unconditional and conditional coverage tests[9], Escanciano and Olmo (2009, 2010a and 2010b) 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. 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 out-of-sample one, the risk associated to the model cannot be eliminated using sub-sampling.

Indeed, let $G_x(x)$ denotes 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:

$$G_{k_b}(x) = (T-b+1)^{-1} \sum_{t=1}^{T-b+1} 1\text{I}_{\{k_{b,t}<x\}}. \tag{A.9}$$

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

$$c_{k_b,1-\tau} = \underbrace{\inf}_{x\in \text{IR}} \{x \mid G_{k_b}(x) \geq 1-\tau\}. \tag{A.10}$$

[9]The conditional coverage test proposed by Christoffersen (1998) combines an unconditional coverage test (the frequency corresponds to the probability) and the independence test (see above).

B Miscaleneous Complementary Results (Web Appendixes)

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%	-7.24%	7.93%
$\alpha = 99.00\%$	-41.88%	-41.88%	.00%	.00%	-9.17%	9.92%
$\alpha = 99.50\%$	-46.41%	-46.41%	.00%	.00%	-10.16%	12.45%

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 1^{st} January 1900 to the 20^{th} September 2011; computations by the authors. This statistics were computed with the results on 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 term 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
α = 95.00%	-24.78%	-24.78%	-.06%	-.06%	-10.16%	8.69%
α = 99.00%	-35.74%	-35.74%	-.11%	-.11%	-20.70%	14.21%
α = 99.50%	-39.95%	-39.95%	.09%	.09%	-28.92%	16.04%

Panel B. Brownian DGP and Gaussian VaR with Specification Error

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

Source: simulations by the authors. Errors are defined as the difference 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. (2010, Table 5, *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 two last processes such as - 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 some 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%	-.06%	-.06%	-10.16%	8.69%
$\alpha = 99.00\%$	-35.74%	-35.74%	-.11%	-.11%	-20.70%	14.21%
$\alpha = 99.50\%$	-39.95%	-39.95%	.09%	.09%	-28.92%	16.04%

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 difference 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. (2010, Table 5, *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 two last processes such as - 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.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 some 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%	-.06%	-.06%	-10.16%	8.69%
$\alpha = 99.00\%$	-35.74%	-35.74%	-.11%	-.11%	-20.70%	14.21%
$\alpha = 99.50\%$	-39.95%	-39.95%	.09%	.09%	-28.92%	16.04%

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 difference 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. (2010, Table 5, *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 two last processes such as - 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.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. (2010)
A Z-test - Jorion (2007)
A Multi-variate Unconditional Coverage Test - Prignon and Smith (2008)
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. (2010)
A Multivariate Test of Zero-autocorrelation of Violations - Hurlin and Tokpavi (2007)
An Estimation-risk adjusted Test - Escanciano and Olmo (2009, 2010-a and 2010-b)
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. (2010)
A Dynamic Binary Response Test - Dumitrescu et al. (2011)
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 and1999)
A Two-stage Test (Coverage Rate and Loss Function) - Angelidis and Gegiannakis (2007)
A Double-threshold Test Colletaz et al. (2010)
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. (2010) for comprehensive surveys.

Table A.6. Dates of the Maximum Adjustment for different VaRs Measures and Backtests Models for

	VaR Methods		Dates	q_1	Dates	q_1^*	Dates	q_2	Dates	q_2^*	Dates	q_3	Dates	q_3^*
95%	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%

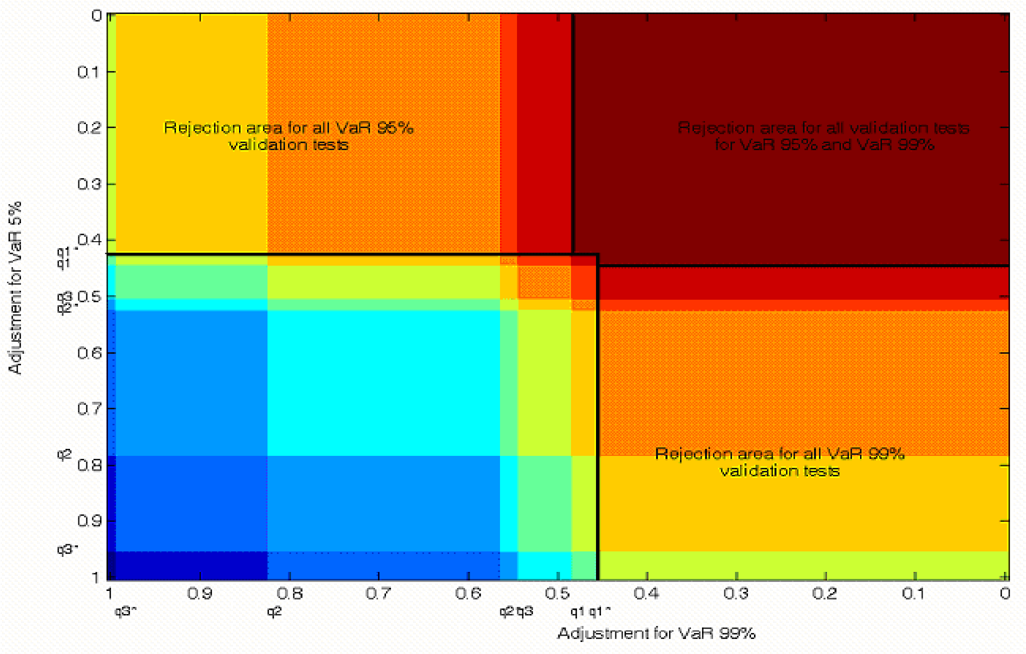
Source: Bloomberg; daily data of the DJIA index in USD from the 1st January 1900 to the 20th September 2011; computations by the authors. We use a moving window of four years (1,040 daily returns) to dynamically re-estimate parameters for the various methods. The variable q_1 refers to the Hit test;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

Table A.7. Ratio k of Maximum Annualized Adjustment Values for Annualized Value-at-Risk Models

VaR Methods	Mean VaR	q_1	q_1^*	q_2	q_2^*	q_3	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; 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 Hit test; q_1^* to Escanciano and Olmo (2010) unconditional test; q_2 to independence test; q_2^* to Escanciano and Olmo (2010) independence test, q_3 to the magnitude test and q_3^* lies to the Bootstrap sampled magnitude test.

Figure A.1: Risk Map for Maximum Annualized Adjustment Values at 5% Confidence Levels for Tests for 95% and 99% Value-at-Risk Models



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_1^* to the Escanciano and Olmo (2010) unconditional test; q_2 to the independence test; q_2^* to the Escanciano and Olmo (2010) independence test, q_3 to the magnitude test and q_3^* lies to a Bootstrap-sample version of the magnitude test.