Risk Analysis, Vol. xx, No. x, 2012 DOI: xxxx

Quantile Uncertainty and Value-at-Risk Model Risk

Carol Alexander,[1*] José María Sarabia[2]

This paper develops a methodology for quantifying model risk in quantile risk estimates. The application of quantile estimates to risk assessment has become common practice in many disciplines, including hydrology, climate change, statistical process control, insurance and actuarial science and the uncertainty surrounding these estimates has long been recognized. Our work is particularly important in finance, where quantile estimates (called Value-at-Risk) have been the cornerstone of banking risk management since the mid 1980's. A recent amendment to the Basel II Accord recommends additional market risk capital to cover all sources of 'model risk' in the estimation of these quantiles. We provide a novel and elegant framework whereby quantile estimates are adjusted for model risk, relative to a benchmark which represents the state of knowledge of the authority that is responsible for model risk. A simulation experiment in which the degree of model risk is controlled illustrates how to quantify model risk and compute the required regulatory capital add-on for banks. An empirical example based on real data shows how the methodology can be put into practice, using only two time series (daily Value-at-Risk and daily profit and loss) from a large bank. We conclude with a discussion of potential applications to non-financial risks.

KEY WORDS: Basel II, maximum entropy, model risk, quantile, risk capital, value-at-risk

1. INTRODUCTION

This paper focuses on the model risk of quantile risk assessments with particular reference to 'Value-at-Risk' (VaR) estimates, which are derived from quantiles of portfolio profit and loss (P&L) distributions. VaR corresponds to an amount that could be lost, with a specified probability, if the portfolio remains unmanaged over a specified time horizon. It has become the global standard for assessing risk in all types of financial firms: in fund management, where portfolios with long-term VaR objectives are actively marketed; in the treasury divisions of large corporations, where VaR is used to assess position risk; and in insurance companies, who measure underwriting and asset management risks in a VaR framework. But most of all, banking regulators remain so confident in VaR that its application to computing market risk capital for banks, used since the 1996 amendment to the Basel I Accord,[3] will soon be extended to include stressed VaR under an amended Basel II and the new Basel III Accords.[4]

The finance industry's reliance on VaR has been supported by decades of academic research. Especially during the last ten years there has been an explosion of articles published on this subject. Popular topics include the introduction of new VaR

[1]ICMA Centre, Henley Business School at the University of Reading, Reading, RG6 6BA, UK; c.alexander@icmacentre.rdg.ac.uk

[2]Department of Economics, University of Cantabria, Avda. de los Castros s/n, 39005-Santander, Spain; sarabiaj@unican.es

*Address correspondence to Carol Alexander, Chair of Risk Management, ICMA Centre, Henley Business School at the University of Reading, Reading, RG6 6BA, UK; Phone: +44 118 3786431

[3]See Basel Committee on Banking Supervision.[(1)]

[4]See Basel Committee on Banking Supervision.[(2,3)]

 0272-4332/xx/0100-0001$22.00/1

models,[5] and methods for testing their accuracy.[6] However, the stark failure of many banks to set aside sufficient capital reserves during the banking crisis of 2008 sparked an intense debate on using VaR models for the purpose of computing the market risk capital requirements of banks. Turner (18) is critical of the manner in which VaR models have been applied and Taleb (19) even questions the very idea of using statistical models for risk assessment. Despite the warnings of Turner, Taleb and early critics of VaR models such as Beder, (20) most financial institutions continue to employ them as their primary tool for market risk assessment and economic capital allocation.

For internal, economic capital allocation purposes VaR models are commonly built using a 'bottom-up' approach. That is, VaR is first assessed at an elemental level, e.g. for each individual trader's positions, then is it progressively aggregated into desk-level VaR, and VaR for larger and larger portfolios, until a final VaR figure for a portfolio that encompasses all the positions in the firm is derived. This way the traders' limits and risk budgets for desks and broader classes of activities can be allocated within a unified framework. However, this bottom-up approach introduces considerable complexity to the VaR model for a large bank. Indeed, it could take more than a day to compute the full (often numerical) valuation models for each product over all the simulations in a VaR model. Yet, for regulatory purposes VaR must be computed at least daily, and for internal management intra-day VaR computations are frequently required.

To reduce complexity in the internal VaR system simplifying assumptions are commonly used, in the data generation processes assumed for financial asset returns and interest rates and in the valuation models used to mark complex products to market every day. For instance, it is very common to apply normality assumptions in VaR models, along with lognormal, constant volatility approximations for exotic options prices and sensitivities.[7] Of course, there is conclusive evidence that financial asset returns are not well represented by normal distributions. However, the risk analyst in a large bank may be forced to employ this assumption for pragmatic reasons.

Another common choice is to base VaR calculations on simple historical simulation. Many large commercial banks have legacy systems that are only able to compute VaR using this approach, commonly basing calculations on at least 3 years of daily data for all traders' positions. Thus, some years after the credit and banking crisis, vastly over-inflated VaR estimates were produced by these models long after the markets returned to normal. The implicit and simplistic assumption that history will repeat itself with certainty – that the banking crisis will recur within the risk horizon of the VaR model – may well seem absurd to the analyst, yet he is constrained by the legacy system to compute VaR using simple historical simulation. Thus, financial risk analysts are often required to employ a model that does not comply with their views on the data generation processes for financial returns, and data that they believe are inappropriate.[8]

Given some sources of uncertainty a Bayesian methodology (21,22) provides an alternative framework to make probabilistic inferences about VaR, assuming that VaR is described in terms of a set of unknown parameters. Bayesian estimates may be derived from posterior parameter densities and posterior model probabilities which are obtained from the prior densities via Bayes theorem, assuming that both the model and its parameters are uncertain. Our method shares ideas with the Bayesian approach, in the sense that we use a 'prior' distribution for $\hat{\alpha}$, in order to obtain a posterior distribution for the quantile.

The problem of quantile estimation under model and parameter uncertainty has also been studied from a classical (i.e. non-Bayesian) point of view. Modarres, Nayak and Gastwirth (23) considered the accuracy of upper and extreme tail estimates of

[5] Historical simulation (4) is the most popular approach amongst banks (5) but data-intensive and prone to pitfalls. (6) Other popular VaR models assume normal risk factor returns with the RiskMetrics covariance matrix estimates. (7) More complex VaR models are proposed by Hull and White, (8) Mittnik and Paolella, (9), Ventner and de Jongh (10), Angelidis et al. (11), Hartz et al. (12), Kuan et al. (13) and many others.

[6] The coverage tests introduced by Kupiec (14) are favoured by banking regulators, and these are refined by Christoffersen. (15) However Berkowitz et al. (16) demonstrate that more sophisticated tests such as the conditional autoregressive test of Engle and Manganelli (17) may perform better.

[7] Indeed, model risk frequently spills over from one business line to another, e.g. normal VaR models are often employed in large banks simply because they are consistent with the geometric Brownian motion assumption that is commonly applied for option pricing and hedging.

[8] Banking regulators recommend 3-5 years of data for historical simulation and require at least 1 year of data for constructing the covariance matrices used in other VaR models.

three right skewed distributions (log-normal, log-logistic and log-double exponential) under model and parameter uncertainty. These authors examined and compared performances of the maximum likelihood and non-parametric estimators based on the empirical or a quasi-empirical quantile function, assuming four different scenarios: the model is correctly specified, the model is mis-specified, the best model is selected using the data and no form is assumed for the model. Giorgi and Narduzzi [(24)] have studied quantile estimation for a self-similar time series and uncertainty that affects their estimates. Figlewski [(25)] deals with estimation error in the assessment of financial risk exposure. This author finds that, under stochastic volatility, estimation error can increase the probabilities of multi-day events such as three 1% tail events in a row, by several orders of magnitude. Empirical findings are also reported using 40 years of daily S&P 500 returns.

The term 'model risk' is commonly applied to encompass various sources of uncertainty in statistical models, including model choice and parameter uncertainty. In July 2009, revisions to the Basel II market risk framework added the requirement that banks set aside additional reserves to cover all sources of model risk in the internal models used to compute the market risk capital charge.[9] Thus, the issue of model risk in internal risk models has recently become very important to banks. Financial risk research has long recognized the importance of model risk. However, following some early work [(26,27,28,29,30,31)] surprizingly few papers deal explicitly with VaR model risk. Early work [(32,33)] investigated sampling error and Kerkhof et al. [(34)] quantify the adjustment to VaR that is necessary for some econometric models to pass regulatory backtests. Quantile-based risk assessment has also been applied to numerous problems in insurance and actuarial science: see Reiss and Thomas, [(35)] Cairns, [(36)] Matthys et al., [(37)] Dowd and Blake [(38)] and many others. However, a general methodology for assessing quantile model risk in finance has yet to emerge.

This paper introduces a new framework for measuring quantile model risk and derives an elegant, intuitive and practical method for computing the risk capital add-on to cover VaR model risk. In addition to the computation of a model risk 'add-on' for a given VaR model and given portfolio, our approach can be used to assess which, of the available VaR models, has the least model risk relative to a given portfolio. Similarly, given a specific VaR model, our approach can assess which portfolio has the least model risk. However, outside of a simulation environment, the concept of a 'true' model against which one might assess model risk is meaningless. All we have is some observable data and our beliefs about the conditional and/or unconditional distribution of the random variable in question. As a result, model risk can only be assessed relative to some benchmark model, which itself is a matter for subjective choice.

In the following: the definition of model risk and a benchmark for assessing model risk is discussed in Section 2; Section 3 gives a formal definition of quantile model risk and outlines a framework for its quantification. We present a statistical model for the probability $\hat{\alpha}$ that is assigned, under the benchmark distribution, to the α quantile of the model distribution. Our idea is to endogenize model risk by using a distribution for $\hat{\alpha}$ to generate a distribution for the quantile. The mean of this model-risk-adjusted quantile distribution detects any systematic bias in the model's α quantile, relative to the α quantile of the benchmark distribution. A suitable quantile of the model-risk-adjusted distribution determines an uncertainty buffer which, when added to the bias-adjusted quantile gives a model-risk-adjusted quantile that is no less than the α quantile of the benchmark distribution at a pre-determined confidence level, this confidence level corresponding to a penalty imposed for model risk; Section 4 presents a numerical example on the application of our framework to VaR model risk, in which the degree of model risk is controlled by simulation; Section 5 illustrates how the methodology could be implemented by a manager or regulator having access to only two time series from the bank: its aggregate daily trading P&L and its corresponding 1% VaR estimates, derived from the usual 'bottom up' VaR aggregation framework; Section 6 discusses the relevance of the methodology to non-financial problems; and Section 7 summarizes and concludes.

2. MODEL RISK AND THE BENCHMARK

We distinguish two sources of model risk: *model choice*, i.e. inappropriate assumptions about the form of the statistical model for the random variable; and *parameter uncertainty*, i.e. estimation error in the parameters of the chosen model. One never knows the

[9] See Basel Committee on Banking Supervision, Section IV. [(3)]

'true' model except in a simulation environment, so assumptions about the form of statistical model must be made. Parameter uncertainty includes sampling error (parameter values can never be estimated exactly because only a finite set of observations on a random variable are available) and optimization error (e.g. different numerical algorithms typically produce slightly different estimates based on the same model and the same data). We remark that there is no consensus on the sources of model risk. For instance, Cont [39] points out that both these sources could be encompassed within a universal model, and Kerkhof et al. [34] distinguish 'identification risk' as an additional source.

Model risk in finance has been approached in two different ways: examining all feasible models and evaluating the discrepancy in their results, or specifying a benchmark model against which model risk is assessed. Papers on the quantification of valuation model risk in the risk-neutral measure exemplify each approach: Cont [39] quantifies the model risk of a complex product by the range of prices obtained under all possible valuation models that are calibrated to market prices of liquid (e.g. vanilla) options; Hull and Suo [40] define model risk relative to the implied price distribution, i.e. a benchmark distribution implied by market prices of vanilla options. In the context of VaR model risk the benchmark approach, which we choose to follow, is more practical than the former.

Some authors identify model risk with the departure of a model from a 'true' dynamic process: see Branger and Schlag [41] for instance. Yet, outside of an experimental or simulation environment, we never know the 'true' model for sure. In practice, all we can observe are realizations of the data generation processes for the random variables in our model. It is futile to propose the existence of a unique and measurable 'true' process because such an exercise is beyond our realm of knowledge.

However, we can observe a maximum entropy distribution (MED). This is based on a 'state of knowledge', i.e. no more and no less than the information available regarding the random variable's behaviour. This information includes the observable data that are thought to be relevant plus any subjective beliefs. Since neither the choice of sample nor the beliefs of the modeller can be regarded as objective, the MED is subjective. For our application to VaR we consider two perspectives on the MED, the internal perspective where the MED would be set by the risk analyst himself, or else by the Chief Risk Officer of the bank, and the external perspective where the MED would be set by the regulator.

Shannon [43] defined the entropy of a probability density function $g(x)$, $x \in \mathcal{R}$ as

$$H(g) = -\mathrm{E}_g[\log g(x)] = -\int_{\mathcal{R}} g(x) \log g(x) dx.$$

This is a measure of the uncertainty in a probability distribution and its negative is a measure of information.[10] The maximum entropy density is the function $f(x)$ that maximizes $H(g)$, subject to a set of conditions on $g(x)$ which capture the testable information.[11] The criterion here is to be as vague as possible (i.e. to maximize uncertainty) given the constraints imposed by the state of knowledge. This way, the MED represents no more (and no less) than the information available. If this information consists only of a historical sample on X of size n then, in addition to the normalization condition, there are n conditions on $g(x)$, one for each data point. In this case, the MED is just the empirical distribution based on that sample. Otherwise, the testable information consists of fewer conditions, which capture only that sample information which is thought to be relevant, and any other conditions imposed by subjective beliefs.

Our recommendation is that banks assess their VaR model risk by comparing their aggregate VaR figure, which is typically computed using the bottom-up approach, with the VaR obtained using the MED in a 'top-down' approach, i.e. calibrated directly to the bank's aggregate daily trading P&L. Typically this P&L contains marked-to-model prices for illiquid products, in which case their valuation model risk is not quantified in our framework.

From the banking regulator's perspective what matters is not the ability to aggregate and disaggregate VaR in a bottom-up framework, but the adequacy of a bank's total market risk capital

[10] For instance, if g is normal with variance σ^2, $H(g) = \frac{1}{2}(1+\log(2\pi)+\log(\sigma))$, so the entropy increases as σ increases and there is more uncertainty and less information in the distribution. As $\sigma \to 0$ and the density collapses the Dirac function at 0, there is no uncertainty but $-H(g) \to \infty$ and there is maximum information. However, there is no universal relationship between variance and entropy and where their orderings differ entropy is the superior measure of information. See Ebrahimi, Maasoumi and Soofi [42] for further insight.

[11] A piece of information is *testable* if it can be determined whether F is consistent with it. One of piece of information is always a normalization condition.

reserves, which are derived from the aggregate market VaR. Therefore, regulators only need to define a benchmark VaR model to apply to the bank's aggregate daily P&L. This model will be the MED of the regulator, i.e. the model that best represents the regulator's state of knowledge regarding the accuracy of VaR models.

Following the theoretical work of Shannon,[43] Zellner,[44] Jaynes[45] and many others it is common to assume the testable information is given by a set of moment functions derived from a sample, in addition to the normalization condition. When only the first two standard moments (mean and variance) are deemed relevant, the MED is a normal distribution.[43] More generally, when the testable information contains the first N sample moments, $f(x)$ takes an exponential form. This is found by maximizing entropy subject to the conditions

$$\mu_n = \int_{\mathcal{R}} x^n g(x)dx, \qquad n = 0, \ldots, N,$$

where $\mu_0 = 1$ and $\mu_n,\ n = 1, ..., N$ are the moments of the distribution. The solution is

$$f(x) = \exp\left(-\sum_{n=0}^{n=N} \lambda_n x^n\right),$$

where the parameters $\lambda_0, \ldots \lambda_n$ are obtained by solving the system of non-linear equations

$$\mu_n = \int x^n \exp\left(-\sum_{n=0}^{n=N} \lambda_n x^n\right) dx, \qquad n = 0, \ldots, N.$$

Rockinger and Jondeau,[46] Wu,[47] Chan[48,49] and others have applied a simple four-moment MED to various econometric and risk management problems. Perhaps surprizingly, since tail weight is an important aspect of financial asset returns distributions, none consider the tail weight that is implicit in the use of an MED based on standard sample moments. But simple moment-based MEDs are only well-defined when N is even. For any odd value of N there will be an increasing probability weight in one of the tails. Also, the four-moment MED has lighter tails than a normal distribution, due to the presence of the term $\exp[-\lambda_4 x^4]$ with non-zero λ_4 in $f(x)$. Indeed, the more moments included in the conditions, the thinner the tail of the MED. Because financial asset returns are typically heavy-tailed it is likely that this property will carry over to a bank's aggregate daily P&L, in which case we would not advocate the use of simple moment-based MEDs.

Park and Bera[50] address the issue of heavy tails in financial data by introducing additional parameters into the moment functions, thus extending the family of moment-based MEDs. Even with just two (generalized) moment conditions based on one additional parameter they show that many heavy-tailed distributions are MEDs, including the Student t and generalized error distributions that are commonly applied to VaR analysis – see Jorion[32] and Lee et al.[51] for example. Since our paper concerns the estimation of low-probability quantiles we shall utilize these distributions as MEDs in our empirical study of Section 5.

There are advantages in choosing a parametric MED for the benchmark. VaR is a quantile of a forward-looking P&L distribution, but to base parameter estimates entirely on historical data limits beliefs about the future to experiences from the past. Parametric distributions are frequently advocated for VaR estimation, and stress testing in particular, because the parameters estimated from historical data may be changed subjectively to accommodate beliefs about the future P&L distribution. We distinguish two types of parametric MEDs. Unconditional MEDs are based on the independent and identically distributed (i.i.d.) assumption. However, since Mandelbrot[52] it has been observed that financial asset returns typically exhibit a 'volatility clustering' effect, thus violating the i.i.d. assumption. Therefore it may be preferable to assume the stochastic process for returns has time-varying conditional distributions that are MEDs.

Volatility clustering is effectively captured by the flexible and popular class of generalized conditional heteroskedastic models (GARCH) models introduced by Bollerslev[53] and since extended in numerous ways by many authors. Berkowitz and O'Brien[54] found that most bottom-up internal VaR models produced VaR estimates that were too large, and insufficiently risk-sensitive, compared with top-down GARCH VaR estimates derived directly from aggregate daily P&L. Thus, from the regulator's perspective, a benchmark for VaR model risk based on a GARCH process for aggregate daily P&L with conditional MEDs would seem appropriate. Filtered historical simulation of aggregate daily P&L would be another popular alternative, especially when applied with a volatility filtering that increases its risk sensitivity: see Barone-Adesi et al.[55] and Hull and White.[56] Alexander and Sheedy[57] demonstrated empirically that GARCH volatility filtering combined with historical simulation can produce very accurate VaR estimates, even at extreme quantiles. By contrast, the standard historical simulation approach, which is based on the i.i.d. assumption, failed many of their backtests.

3. MODELLING QUANTILE MODEL RISK

The α quantile of a continuous distribution F of a real-valued random variable X with range $\mathcal{R}$ is denoted

$$q_\alpha^F = F^{-1}(\alpha). \quad (1)$$

In financial applications the probability α is often predetermined. Frequently it will be set by senior managers or regulators and small or large values corresponding to extreme quantiles are very commonly used. For instance, regulatory market risk capital is based on VaR models with $\alpha = 1\%$ and a risk horizon of 10 trading days.

In our statistical framework F is identified with the unique MED based on a state of knowledge $\mathcal{K}$ which contains all testable information on F. We characterise a statistical model as a pair $\{\hat{F}, \hat{\mathcal{K}}\}$ where $\hat{F}$ is a distribution and $\hat{\mathcal{K}}$ is a filtration which encompasses both the model choice and its parameter values. The model provides an estimate $\hat{F}$ of F, and uses this to compute the α quantile. That is, instead of (1) we use

$$q_\alpha^{\hat{F}} = \hat{F}^{-1}(\alpha). \quad (2)$$

Quantile model risk arises because $\{\hat{F}, \hat{\mathcal{K}}\} \neq \{F, \mathcal{K}\}$. Firstly, $\hat{\mathcal{K}} \neq \mathcal{K}$, e.g. $\mathcal{K}$ may include the belief that only the last six months of data are relevant to the quantile today; yet $\hat{\mathcal{K}}$ may be derived from an industry standard that must use at least one year of observed data in $\hat{\mathcal{K}}$;[12] and secondly, $\hat{F}$ is not, typically, the MED even based on $\hat{\mathcal{K}}$, e.g. the execution of firm-wide VaR models for a large commercial bank may present such a formidable time challenge that $\hat{F}$ is based on simplified data generation processes, as discussed in the introduction.

In the presence of model risk the α quantile of the model is not the α quantile of the MED, i.e. $q_\alpha^{\hat{F}} \neq q_\alpha^F$. The model's α quantile $q_\alpha^{\hat{F}}$ is at a different quantile of F and we use the notation $\hat{\alpha}$ for this quantile, i.e. $q_\alpha^{\hat{F}} = q_{\hat{\alpha}}^F$, or equivalently,

$$\hat{\alpha} = F(\hat{F}^{-1}(\alpha)). \quad (3)$$

In the absence of model risk $\hat{\alpha} = \alpha$ for every α. Otherwise, we can quantify the extent of model risk by the deviation of $\hat{\alpha}$ from α, i.e. the distribution of the quantile probability errors

$$e(\alpha|F, \hat{F}) = \hat{\alpha} - \alpha. \quad (4)$$

If the model suffers from a systematic, measurable bias at the α quantile then the mean error $\bar{e}(\alpha|F, \hat{F})$ should be significantly different from zero. A significant and positive (negative) mean indicates a systematic over (under) estimation of the α quantile of the MED. Even if the model is unbiased it may still lack efficiency, i.e. the dispersion of $e(\alpha|F, \hat{F})$ may be high. Several measures of dispersion may be used to quantify the efficiency of the model, including the root mean squared error (RMSE), the mean absolute error (MAE) and the range.

We now regard $\hat{\alpha} = F(\hat{F}^{-1}(\alpha))$ as a random variable with a distribution that is generated by our two sources of model risk, i.e. model choice and parameter uncertainty. Because $\hat{\alpha}$ is a probability it has range $[0, 1]$, so the α quantile of our model, adjusted for model risk, falls into the category of generated random variables. For instance, if $\hat{\alpha}$ is parameterized by a beta distribution $\mathcal{B}(a, b)$ with density $(0 < u < 1)$

$$f_{\mathcal{B}}(u; a, b) = B(a, b)^{-1}[u^{a-1}(1-u)^{b-1}], \quad (5)$$

$a, b \geq 0$, where $B(a, b)$ is the beta function, then the α quantile of our model, adjusted for model risk, is a beta-generated random variable:

$$Q(\alpha|F, \hat{F}) = F^{-1}(\hat{\alpha}), \quad \hat{\alpha} \sim \mathcal{B}(a, b).$$

Beta generated distributions were introduction by Eugene et al.[(58)] and Jones.[(59)] They may be characterized by their density function $(-\infty < x < \infty)$

$$g_F(x) = B(a, b)^{-1} f(x)[F(x)]^{a-1}[1 - F(x)]^{b-1},$$

where $f(x) = F'(x)$. Several other distributions $\mathcal{D}[0, 1]$ with range the unit interval are available for generating the model-risk-adjusted quantile distribution; see Zografos and Balakrishnan[(59)] for example. Hence, in the most general terms the model-risk-adjusted VaR is a random variable with distribution:

$$Q(\alpha|F, \hat{F}) = F^{-1}(\hat{\alpha}), \quad \hat{\alpha} \sim \mathcal{D}[0, 1]. \quad (6)$$

The mean $E[Q(\alpha|F, \hat{F})]$ of $Q(\alpha|F, \hat{F})$ quantifies any systematic bias in the quantile estimates: e.g. if the MED has heavier tails than the model then

[12] As is the case under current banking regulations for the use of VaR to estimate risk capital reserves - see Basel Committee on Banking Supervision.[(1)]

extreme quantiles $q_\alpha^{\hat{F}}$ will be biased: if α is close to zero then $E[Q(\alpha|F,\hat{F})] > q_\alpha^F$ and if α is close to one then $E[Q(\alpha|F,\hat{F})] < q_\alpha^F$. This bias can be removed by adding the difference $q_\alpha^F - E[Q(\alpha|F,\hat{F})]$ to the model's α quantile $q_\alpha^{\hat{F}}$ so that the bias-adjusted quantile has expectation q_α^F.

The bias-adjusted α quantile estimate could still be far away from the maximum entropy α quantile: the more dispersed the distribution of $Q(\alpha|F,\hat{F})$, the greater the potential for $q_\alpha^{\hat{F}}$ to deviate from q_α^F. Because financial regulators require VaR estimates to be conservative, we adjust for the inefficiency of the VaR model by introducing an uncertainty buffer to the bias-adjusted α quantile by adding a quantity equal to the difference between the mean of $Q(\alpha|F,\hat{F})$ and $G_F^{-1}(y)$, the y quantile of $Q(\alpha|F,\hat{F})$, to the bias-adjusted α quantile estimate. This way, we become $(1-y)\%$ confident that the model-risk-adjusted α quantile is no less than q_α^F.

Finally, our point estimate for the model-risk-adjusted α quantile becomes:

$$\begin{aligned} q_\alpha^{\hat{F}} &+ \{q_\alpha^F - E[Q(\alpha|F,\hat{F})]\} + \{E[Q(\alpha|F,\hat{F})] - G_F^{-1}(y)\} \\ &= q_\alpha^{\hat{F}} + q_\alpha^F - G_F^{-1}(y), \end{aligned} \tag{7}$$

where $\{q_\alpha^F - E[Q(\alpha|F,\hat{F})]\}$ is the 'bias adjustment' and $\{E[Q(\alpha|F,\hat{F})] - G_F^{-1}(y)\}$ is the 'uncertainty buffer'.

The total model-risk adjustment to the quantile estimate is thus $q_\alpha^F - G_F^{-1}(y)$, and the computation of $E[Q(\alpha|F,\hat{F})]$ could be circumvented if the decomposition into bias and uncertainty components is not required. The confidence level $1-y$ reflects a penalty for model risk which could be set by the regulator. When X denotes daily P&L and α is small (e.g. 1%), typically all three terms on the right hand side of (7) will be negative. But the $\alpha\%$ daily VaR is $-q_\alpha^{\hat{F}}$, so the model-risk-adjusted VaR estimate becomes $-q_\alpha^{\hat{F}} - q_\alpha^F + G_F^{-1}(y)$. The add-on to the daily VaR estimate, $G_F^{-1}(y) - q_\alpha^F$, will be positive unless VaR estimates are typically much greater than the benchmark VaR. In that case there should be a negative bias adjustment, and this could be large enough to outweigh the uncertainty buffer, especially when y is large, i.e. when we require only a low degree of confidence for the model-risk-adjusted VaR to exceed the benchmark VaR.

4. NUMERICAL EXAMPLE

We now describe an experiment in which a portfolio's returns are simulated based on a known data generation process. This allows us to control the degree of VaR model risk and to demonstrate that our framework yields intuitive and sensible results for the bias and inefficiency adjustments described above.

Recalling that the popular and flexible class of GARCH models was advocated by Berkowitz and O'Brien[(54)] for top-down VaR estimation we assume that our conditional MED for the returns X_t at time t is $\mathcal{N}(0,\sigma_t^2)$, where σ_t^2 follows an asymmetric GARCH process. The model falls into the category of maximum entropy ARCH models introduced by Park and Bera,[(49)] where the conditional distribution is normal. Thus it has only two constraints, on the conditional mean and variance.

First the return x_t from time t to $t+1$ and its variance σ_t^2 are simulated using:

$$\sigma_t^2 = \omega + \alpha(x_{t-1}-\lambda)^2 + \beta\sigma_{t-1}^2, \quad x_t|\mathcal{I}_t \sim \mathcal{N}(0,\sigma_t^2), \tag{8}$$

where $\omega > 0, \alpha, \beta \geq 0, \alpha+\beta \leq 1$ and $\mathcal{I}_t = (x_{t-1}, x_{t-2}, \ldots)$.[13] For the simulated returns the parameters of (8) are assumed to be:

$$\omega = 1.5\times10^{-6}, \alpha = 0.04, \lambda = 0.005, \beta = 0.95, \tag{9}$$

and so the steady-state annualized volatility of the portfolio return is 25%.[14] Then the MED at time t is $F_t = F(X_t|\mathcal{K}_t)$, i.e. the conditional distribution of the return X_t given the state of knowledge $\mathcal{K}_t$, which comprises the observed returns $\mathcal{I}_t$ and the knowledge that $X_t|\mathcal{I}_t \sim \mathcal{N}(0,\sigma_t^2)$.

At time t, a VaR model provides a forecast $\hat{F}_t = \hat{F}(X_t|\hat{\mathcal{K}}_t)$ where $\hat{\mathcal{K}}$ comprises $\mathcal{I}_t$ plus the model $X_t|\mathcal{I}_t \sim \mathcal{N}(0,\hat{\sigma}_t^2)$. We now consider three different models for $\hat{\sigma}_t^2$. The first model has the correct choice of model but uses incorrect parameter values: instead of (9) the fitted model is:

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}(x_{t-1}-\hat{\lambda})^2 + \hat{\beta}\hat{\sigma}_{t-1}^2, \tag{10}$$

with

$$\hat{\omega} = 2\times10^{-6}, \hat{\alpha} = 0.0515, \hat{\lambda} = 0.01, \hat{\beta} = 0.92. \tag{11}$$

The steady-state volatility estimate is therefore correct, but since $\hat{\alpha} > \alpha$ and $\hat{\beta} < \beta$ the fitted volatility process is more 'jumpy' than the simulated variance

[13] We employ the standard notation α for the GARCH return parameter here; this should not be confused with the notation α for the quantile of the returns distribution, which is also standard notation in the VaR model literature.

[14] The steady-state variance is $\bar{\sigma}^2 = (\omega+\alpha\lambda^2)/(1-\alpha-\beta)$ and for the annualization we have assumed returns are daily, and that there are 250 business days per year.

generation process. In other words, compared with σ_t, $\hat{\sigma}_t$ has a greater reaction but less persistence to innovations in the returns, and especially to negative returns since $\hat{\lambda} > \lambda$.

The other two models are chosen because they are commonly adopted by financial institutions, having been popularized by the 'RiskMetrics' methodology introduced by JP Morgan in the mid-1990's – see RiskMetrics.[(7)] The second model uses a simplified version of (8) with:

$$\hat{\omega} = \hat{\lambda} = 0, \hat{\alpha} = 0.06, \hat{\beta} = 0.94. \tag{12}$$

This is the RiskMetrics exponentially weighted moving average (EWMA) estimator in which a steady-state volatility is not defined. The third model is the RiskMetrics 'Regulatory' estimator in which:

$$\hat{\alpha} = \hat{\lambda} = \hat{\beta} = 0, \hat{\omega} = \frac{1}{250}\sum_{i=1}^{250} x_{t-i}^2. \tag{13}$$

A time series of 10,000 returns $\{x_t\}_{t=1}^{10,000}$ is simulated from the 'true' model (8) with parameters (9). Then, for each of the three models defined above we use this time series to (a) estimate the daily VaR, which when expressed as a percentage of the portfolio value is given by $\Phi^{-1}(\alpha)\hat{\sigma}_t$, and (b) compute the probability $\hat{\alpha}_t$ associated with this quantile under the simulated returns distribution $F_t = F(X_t|\mathcal{K}_t)$. Because $\Phi^{-1}(\hat{\alpha}_t)\sigma_t = \Phi^{-1}(\alpha)\hat{\sigma}_t$, this is given by

$$\hat{\alpha}_t = \Phi\left[\Phi^{-1}(\alpha)\frac{\hat{\sigma}_t}{\sigma_t}\right]. \tag{14}$$

Now for each VaR model we use the simulated distribution to estimate $\hat{\alpha}$ at every time point, using (14). For $\alpha = 0.1\%, 1\%$ and 5%, Table I reports the mean of $\hat{\alpha}$ and the RMSE between $\hat{\alpha}$ and α. The closer $\hat{\alpha}$ is to α, the smaller the RMSE and the less model risk there is in the VaR model. The Regulatory model yields an $\hat{\alpha}$ with the highest RMSE, for every α, so this has the greatest degree of model risk. The AGARCH model, which we already know has the least model risk of the three, produces a distribution for $\hat{\alpha}$ that has mean closest to the true α and the smallest RMSE. These observations are supported by Figure 1, which depicts the empirical distribution of $\hat{\alpha}$ and Figure 2, which shows the empirical densities of the model-risk-adjusted VaR estimates $F^{-1}(\hat{\alpha})$, taking $\alpha = 1\%$ for illustration. Here and henceforth VaR is stated as a percentage of the portfolio value, multiplied by 100.

A point estimate for model-risk-adjusted VaR (RaVaR, for short) is computed using (7). Because we have a conditional MED the benchmark VaR (BVaR, for short) depends on the time it is measured, and so does the RaVaR. For illustration, we select a point when the simulated volatility is at its steady-state value of 25% – so the BVaR is 4.886, 3.678 and 2.601 at the 0.1%, 1% and 5% levels, respectively. Drawing at random from the points when the simulated volatility was 25%, we obtain AGARCH, EWMA and Regulatory volatility forecasts of 27.02%, 23.94% and 28.19% respectively. These volatilities determine the VaR estimates that we shall now adjust for model risk.

Table II summarizes the bias and the uncertainty buffer, for different levels of α, based on the empirical distribution of $Q(\alpha|F,\hat{F})$.[15] It reveals a general tendency for the EWMA model to slightly underestimate VaR and the other models to slightly overestimate VaR. Yet the bias is relatively small, since all models assume the same normal form as the MED and the only difference between them is their volatility forecast. Although the bias tends to increase as α decreases it is not significant for any model.[16] Beneath the bias we report the 5% quantile of the model-risk-adjusted VaR distributions, since we shall first compute the RaVaR so that it is no less than the BVaR with 95% confidence.

Following the framework introduced in the previous section we now define:

$$\begin{aligned}\text{RaVaR}(y) &= \text{VaR} + (\text{BVaR} - E[Q(\alpha|F,\hat{F})]) + \\ &+ (E[Q(\alpha|F,\hat{F})] - G_F^{-1}(y)),\end{aligned}$$

where $\{\text{BVaR} - E[Q(\alpha|F,\hat{F})]\}$ is the 'bias adjustment' and $\{E[Q(\alpha|F,\hat{F})] - G_F^{-1}(y)\}$ is the 'uncertainty buffer'.

Table III sets out the RaVaR computation for $y = 5\%$. The model's volatility forecasts are in the first row and the corresponding VaR estimates are in the first row of each cell, for $\alpha = 0.1\%$, 1% and 5% respectively. The (small) bias is corrected by adding the bias from Table II to each VaR estimate. The main source of model risk here concerns the potential for a large (positive or negative) errors in the quantile probabilities, i.e. the dispersion of the densities in Figure 2. To adjust for this we add to the bias-

[15] Similar results based on the fitted distributions are not reported for brevity.

[16] Standard errors of $Q(\alpha|F,\hat{F})$ are not reported, for brevity. They range between 0.157 for the AGARCH at 5% to 0.891 for the Regulatory model at 0.1%, and are directly proportional to the degree of model risk just like the standard errors on the quantile probabilities given in Table I .

adjusted VaR the uncertainty buffer given in Table II . This gives the RaVaR estimates shown in the third row of each cell.

Since risk capital is a multiple of VaR, the percentage increase resulting from replacing VaR by RaVaR(y) is:

$$\% \text{ risk capital increase} = \frac{\text{BVaR} - G_F^{-1}(y)}{\text{VaR}} \quad . \qquad (15)$$

The penalty (15) for model risk depends on α, except in the case that both the MED and VaR model are normal, and on the confidence level $(1-y)\%$. Table IV reports the percentage increase in risk capital due to model risk when RaVaR is no less than the BVaR with $(1-y)\%$ confidence. We consider $y = 5\%, 15\%$ and 25%, with smaller values of y corresponding to a stronger condition on the model-risk adjustment. We also set $\alpha = 1\%$ because risk capital is based on the VaR at this level of significance under the Basel Accords. We also take the opportunity here to consider two further scenarios, in order to verify the robustness of our qualitative conclusions.

The first row of each section of Table IV reports the volatilities estimated by each VaR model at a point in time when the benchmark model has volatility 25%. Thus for scenario 1, upon which the results have been based up to now, we have the volatilities 27.02%, 23.94% and 28.19% respectively. For scenario 2 the three volatilities are 28.32%, 27.34% and 22.40%, i.e. the AGARCH and EWMA models over-estimate and the Regulatory model under-estimates the benchmark model's volatility. For scenario 3 the AGARCH model slightly under-estimates the benchmark's volatility and the other two models over-estimate it.

The three rows in each section of the Table IV give the percentage increase in risk capital that would be required were the regulator to choose 95%, 85% or 75% confidence levels for the RaVaR. Clearly, for each model and each scenario, the add-on for VaR model risk increases with the degree of confidence that the regulator requires for the RaVaR to be at least as great as the BVaR. At the 95% confidence level, a comparison of the first row of each section of the table shows that risk capital would be increased by roughly 8–10% when based on the AGARCH model, whereas it would be increased by about 13–14.5% under the EWMA model and by roughly 21–27% under the Regulatory model. The same ordering of the RaVaRs applies to each scenario, and at every confidence level. That is, the model-risk adjustment results in an increase in risk capital that is positively related to the degree of model risk, as it should be.

Table I . Sample statistics for quantile probabilities. The mean of $\hat{\alpha}$ and the RMSE between $\hat{\alpha}$ and α. The closer $\hat{\alpha}$ is to α, the smaller the RMSE and the less model risk there is in the VaR model.

α		AGARCH	EWMA	Regulatory
0.10%	Mean	0.11%	0.16%	0.23%
	RMSE	0.07%	0.14%	0.37%
1%	Mean	1.03%	1.25%	1.34%
	RMSE	0.42%	0.64%	1.22%
5%	Mean	4.97%	5.44%	5.27%
	RMSE	1.03%	1.31%	2.66%

Finally, comparison between the three scenarios shows that the add-on will be greater on days when the model under-estimates the VaR than it is on days when it over-estimates VaR, relative to the benchmark. Yet even when the model VaR is greater than the benchmark VaR the add-on is still positive. This is because the uncertainty buffer remains large relative to the bias adjustment, even at the 75% level of confidence. However, if regulators were to require a lower confidence for the uncertainty buffer, such as only 50% in this example, then it could happen that the model-risk add-on becomes negative.

5. EMPIRICAL ILLUSTRATION

How could the authority responsible for model risk, such as a bank's local regulator or its Chief Risk Officer, implement the proposed adjustment for model risk in practice? The required inputs to a model-risk-adjusted VaR calculation are two daily time series that the bank will have already been recording to comply with Basel regulations: one series is the aggregate daily trading P&L and the other is the aggregated 1% daily VaR estimates corresponding to this trading activity. From the regional office of a large international banking corporation we have obtained data on aggregate daily P&L and the corresponding aggregate VaR for each day, the VaR being computed in a bottom-up framework based on standard (un-filtered) historical simulation. The data span the period 3 Sept 2003 to 18 March 2009, thus including the banking crisis during the last quarter of 2008. In this section the bank's daily VaR will be compared with a top-down VaR estimate based on a benchmark VaR model tuned to the aggregate daily trading P&L, and a

Table II . Components of the model-risk adjustment. The bias and the 95% uncertainty buffer, for different levels of α, derived from the mean and 5% quantile of the empirical distribution of $Q(\alpha|F, \hat{F})$. The bias is the difference between the benchmark VaR (which is 4.886, 3.678 and 2.601 at the 0.1%, 1% and 5% levels) and the mean. The uncertainty buffer is the difference between the mean and the 5% quantile. UB means 'Uncertainty Buffer' and Q is the quantile.

α		AGARCH	EWMA	Regulatory
0.1%	Mean	4.919	4.793	4.961
	5% Q	4.447	4.177	3.912
	Bias	-0.033	0.093	-0.075
	UB	0.472	0.615	1.049
1%	Mean	3.703	3.608	3.735
	5% Q	3.348	3.145	2.945
	Bias	-0.025	0.070	-0.056
	UB	0.355	0.463	0.789
5%	Mean	2.618	2.551	2.641
	5% Q	2.366	2.224	2.082
	Bias	-0.017	0.050	-0.040
	UB	0.252	0.327	0.559

Table III . Computation of 95% RaVaR. The volatility $\hat{\sigma}_t$ determines the VaR estimates for $\alpha = 0.1\%$, 1% and 5% respectively, as $\Phi^{-1}(\alpha)\,\hat{\sigma}_t$. Adding the bias shown in Table II gives the bias-adjusted VaR. Adding the uncertainty buffer given in Table II to the bias-adjusted VaR yields the model-risk-adjusted VaR estimates (RaVAR) shown in the third row of each cell. BA means 'Bias-Adjusted'.

		AGARCH	EWMA	Regulatory
α	Volatility	27.02%	23.94%	28.19%
0.10%	VaR	5.277	4.678	5.509
	BA VaR	5.244	4.772	5.434
	RaVaR	5.716	5.387	6.483
1%	VaR	3.972	3.522	4.147
	BA VaR	3.948	3.592	4.091
	RaVaR	4.303	4.055	4.880
5%	VaR	2.809	2.490	2.932
	BA VaR	2.791	2.540	2.892
	RaVaR	3.043	2.867	3.451

model-risk-adjusted VaR will be derived for each day between 3 September 2007 and 18 March 2009.

When the data are not i.i.d. the benchmark should be a conditional MED rather than an unconditional MED. To illustrate this we compute the time series of 1% quantile estimates based on alternative benchmarks. First we employ the Student

Table IV . Percentage increase in risk capital from model-risk adjustment of VaR. The percentage increase from VaR to RaVaR based on three scenarios for each model's volatility estimate at time t. In each case the benchmark model's conditional volatility was 25%.

	AGARCH	EWMA	Regulatory
Scenario 1	27.02%	23.94%	28.19%
95%	8.40%	14.46%	21.29%
85%	5.23%	9.45%	13.93%
75%	3.08%	6.53%	9.14%
Scenario 2	28.32%	27.34%	22.40%
95%	8.02%	12.66%	26.79%
85%	4.98%	8.27%	17.53%
75%	2.94%	5.71%	11.50%
Scenario 3	23.18%	26.34%	28.66%
95%	9.80%	13.14%	20.94%
85%	6.09%	8.59%	13.70%
75%	3.59%	5.93%	8.99%

t distribution, which maximizes the Shannon entropy subject to the moment constraint[17]

$$E[\log(\nu^2 + (X/\lambda)^2)] = \log(\nu^2) + \psi\left(\frac{1+\nu^2}{2}\right) - \psi\left(\frac{\nu^2}{2}\right).$$

Secondly we consider the AGARCH process (8) which has a normal conditional distribution for the errors. We also considered taking the generalized error distribution (GED), introduced by Nelson,[(61)] as an unconditional MED benchmark. The GED has the probability density $(-\infty < x < \infty)$

$$f(x;\lambda,\nu) = \frac{\nu^{-1/\nu}}{2\Gamma(1+1/\nu)\lambda}\exp\left(-\frac{1}{\nu}|x/\lambda|^{\nu}\right),$$

and maximizes the Shannon entropy subject to the moment constraint $E[\nu^{-1}|X/\lambda|^{\nu}] = \frac{\nu^{-1/\nu}}{2\Gamma(1+1/\nu)}$. This is more flexible than the (always heavy-tailed) Student t because when $\nu < 2$ ($\nu > 2$) the GED has heavier (lighter) tails than the normal distribution. We also considered using a Student t conditional MED with the AGARCH process, and a symmetric GARCH process, where $\lambda = 0$ in (8), with both Student t and normal conditional distributions. However, the unconditional GED produced results similar to (and just as bad as) the unconditional

[17]Here $\psi(z)$ denotes the digamma function, λ is a scale parameter and ν the corresponding shape parameter. See Park and Bera,[(49)] Table 1 for the moment condition.

Student t. Also all four conditional MEDs produced quite similar results. Our choice of the AGARCH process with normal errors was based on the successful results of the conditional and unconditional coverage tests that are commonly applied to test for VaR model specification – see Christoffersen.[15] For reasons of space, none of these results are reported but they are available from the authors on request.

We estimate the two selected benchmark model parameters using a 'rolling window' framework that is standard practice for VaR estimation. Each sample contains n consecutive observations on the bank's aggregate daily P&L, and a sample is rolled forward one day at a time, each time re-estimating the model parameters. Figure 3 compares the 1% quantile of the Student t distribution with the 1% quantile of the normal AGARCH process on the last day of each rolling window. Also shown is the bank's aggregate P&L for the day corresponding to the quantile estimate, between 3 September 2007 and 18 March 2008. The effect of the banking crisis is evidenced by the increase in volatility of daily P&L which began with the shock collapse of Lehmann Brothers in mid September 2008. Before this time the 1% quantiles of the unconditional Student t distribution were very conservative predictors of daily losses, because the rolling windows included the commodity crisis of 2006. Yet at the time of the banking crisis the Student t model clearly underestimated the losses that were being experienced. Even worse, from the bank's point of view, the Student t model vastly over-estimated the losses made during the aftermath of the crisis in early 2009 and would have led to crippling levels of risk capital reserves. Even though we used $n = 200$ for fitting the unconditional Student t distribution and a much larger sample, with $n = 800$, for fitting the normal AGARCH process it is clear that the GARCH process captures the strong volatility clustering of daily P&L far better than the unconditional MED. True, the AGARCH process often just misses a large unexpected loss, but because it has the flexibility to respond the very next day, the AGARCH process rapidly adapts to changing market conditions just as a VaR model should.

In an extensive study of the aggregate P&L of several large commercial banks, Berkowitz and O'Brien[54] found that GARCH models estimated on aggregate P&L are far more accurate predictors of aggregate losses than the bottom-up VaR figures that most banks use for regulatory capital calculations. Figure 3 verifies this finding by also depicting the 1% daily VaR reported by the bank, multiplied by -1 since it is losses rather than profits that VaR is supposed to cover. This time series has many features in common with the 1% quantiles derived from the Student t distribution. The substantial losses of up to \$80m per day during the last quarter of 2008 were not predicted by the bank's VaR estimates, yet following the banking crisis the bank's VaR was far too conservative. We conclude, like Berkowitz and O'Brien,[54] that unconditional approaches are much less risk sensitive than GARCH models and for this reason we choose the normal AGARCH rather than the Student t as our benchmark for model risk assessment below.

Figure 4 again depicts $-1\times$ the bank's aggregate daily VaR, denoted $-\mathrm{VaR}_t$ in the formula below. Now our bias adjustment is computed daily using an empirical model-risk-adjusted VaR distribution based on the observations in each rolling window. Under the normal AGARCH benchmark, for a sample starting at T and ending at $T + n$, the daily P&L distribution at time t, with $T \leq t \leq T + n$ is $\mathrm{N}(0, \hat{\sigma}_t^2)$ where $\hat{\sigma}_t$ is the time-varying standard deviation of the AGARCH model. Following (3) we set $\hat{\alpha}_t = \Phi\left(-\hat{\sigma}_t^{-1}\mathrm{VaR}_t\right)$ for each day in the window and then we use the empirical distribution of $\hat{\alpha}_t$ for $T \leq t \leq T + n$ to generate the model-risk-adjusted VaR distribution (6). Then, following (7), the bias adjustment at $T + n$ is the difference between the mean of the model-risk-adjusted quantile distribution and the benchmark VaR at $T + n$.

Since the bank's aggregate VaR is very conservative at the beginning of the period but not large enough during the crisis, in Figure 4 a positive bias reduces the VaR prior to the crisis but during the crisis a negative bias increases the VaR. Having applied the bias adjustment we then set $y = 25\%$ in (7) to derive the uncertainty buffer corresponding to a 75% confidence that the RaVaR will be no less than the BVaR. This is the difference between the mean and the 25%-quantile of the model-risk-adjusted VaR distribution. Adding this uncertainty buffer to the bias-adjusted VaR we obtain the 75% RaVaR depicted in Figure 4 which is given by (7). This is more variable than the bank's original aggregate VaR, but risk capital is based on an average VaR figure over the last 60 days (or the previous VaR, if this is greater) so the adjustment need not induce excessive variation in risk capital, which would be difficult to budget.

Fig. 1. Density of quantile probabilities. Empirical distribution of $\hat{\alpha}_t$ derived from (14) with $\alpha = 1\%$, based on 10,000 daily returns simulated using (8) with parameters (9).

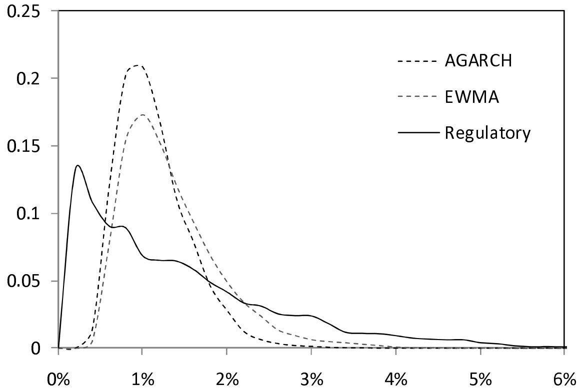
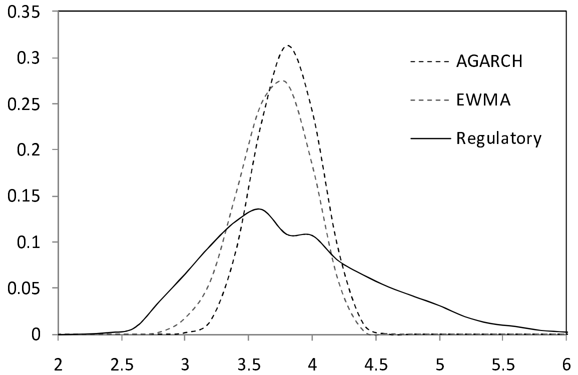


Fig. 2. Density of model-risk-adjusted daily VaR. Empirical densities of the model-risk-adjusted VaR estimates $F^{-1}(\hat{\alpha})$ with $\alpha = 1\%$, based on the 10,000 observations on $\hat{\alpha}$ whose density is shown in Figure 1.



6. APPLICATION TO NON-FINANCIAL PROBLEMS

Quantile-based risk assessment has become standard practice in a wide variety of non-financial disciplines, especially in environmental risk assessment and in statistical process control. For instance, applications to hydrology are studied by Arsenault and Ashkar[(62)] and Chebana and Ouarda,[(63)] and other environmental applications of quantile risk assessments include climate change (Katz et al.,[(64)] and Diodato and Bellocchi,[(65)]) wind power[(66)] and nuclear power[(67)]. In statistical process control, quantiles are used for computing capability indices,[(68)] for measuring efficiency[(69)] and for reliability analysis.[(70)]

In these contexts the uncertainty surrounding quantile-based risk assessments has been the subject of many papers[(36,71,72,73,74,75)]. Both model choice and parameter uncertainty has been considered. For instance, Vermaat and Steerneman[(76)] discuss modified quantiles based on extreme value distributions in reliability analysis, and Sveinsson et al.[(77)] examine the errors induced by using a sample limited to a single site in a regional frequency analysis.

As in banking, regulations can be a key driver for the accurate assessment of environmental risks such as radiation from nuclear power plants. Nevertheless, health or safety regulations are unlikely to extend as far as requirements for regular monitoring and reporting of quantile risks in the foreseeable future. The main concern about the uncertainty surrounding quantile risk assessment is more likely to come from senior management, in recognition that inaccurate risk assessment could jeopardize the reputation of the firm, profits to shareholders and/or the safety of the public. The question then arises: if it is a senior manager's knowledge that specifies the benchmark distribution for model risk assessment, why should this benchmark distribution not be utilized in practice?

As exemplified by the work of Sveinsson et al.,[(77)] the time and expense of utilizing a complete sample of data may not be feasible except on a few occasions where a more detailed risk analysis is performed, possibly by an external consultant. In this case the most significant source of model risk in regular risk assessments would stem from parameter uncertainty. Model choice might also be a source of risk when realistic model assumptions would lead to systems that are too costly and time-consuming to employ on a daily basis. For instance, Merrik et al. (2005) point out that the use of Bayesian simulation for modelling large and complex maritime risk systems should be considered state-of-the-art, rather than standard practice. Also in reliability modelling, individual risk assessments for various components are typically aggregated to derive the total risk for the system. A full account of component default codependencies may require a lengthy scenario analyses based on a complex model (e.g. multivariate copulas with non-standard marginal distributions). This type of risk assessment might not be feasible every day, but if it could be performed on an occasional basis then it could be used as a benchmark for adjusting everyday quantile estimates for model risk.

Generalizations and extensions to higher dimensions of the benchmark model could be implemented. A multivariate parametric MED for the benchmark model can be obtained using similar arguments to

Fig. 3. Daily P&L, daily VaR and two potential benchmark VaRs. The bank's daily P&L is depicted by the grey line and it's 'bottom-up' daily VaR estimates by the black line. The dotted and dashed lines are the Student t (unconditional MED) benchmark VaR and the normal AGARCH (conditional MED) benchmark VaR.

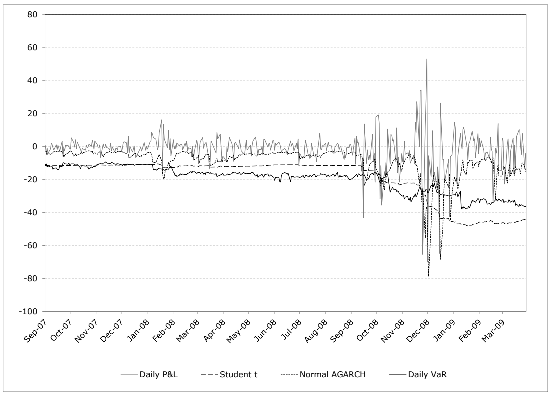
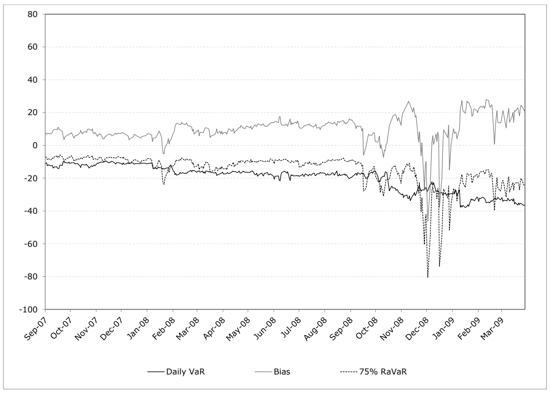


Fig. 4. Aggregate VaR, bias adjustment and 75% RaVaR. The bank's daily VaR estimates are repeated (black line) and compared with the bias-adjustment (grey line) and the final model-risk-adjusted VaR at the 75% confidence level (dotted line) based on the normal AGARCH benchmark model.



those used in the univariate case. In an engineering context, Kapur [78] have considered several classes of multivariate MED. Zografos [79] characterized Pearson's type II and VII multivariate distributions, Aulogiaris and Zografos [80] the symmetric Kotz and Burr multivariate families and Bhattacharya [81] the class of multivariate Liouville distributions. Closed expressions for entropies in several multivariate distributions have been provided by Ahmed and Gokhale [82] and Zografos and Nadarajah. [83]

A major difference between financial and non-financial risk assessment is the availability of data. For instance, in the example described in the previous section the empirical distributions for model-risk-adjusted quantiles were derived from several years of regular output from the benchmark model. Clearly, the ability to generate the adjusted quantile distribution from a parametric distribution for $\hat{\alpha}$, such as the beta distribution (5), opens the methodology to applications where relatively few observations on the benchmark quantile are available, but there are enough to estimate the parameters of a distribution for $\hat{\alpha}$.

7. SUMMARY

This paper concerns the model risk of quantile-based risk assessments, with a focus on the risk of producing inaccurate VaR estimates because of an inappropriate choice of VaR model and/or inaccuracy in the VaR model parameter estimates. We develop a statistical methodology that provides a practical solution to the problem of quantifying the regulatory capital that should be set aside to cover this type of model risk, under the July 2009 Basel II proposals. We argue that there is no better choice of model risk benchmark than a maximum entropy distribution since, by definition, this embodies the entirety of information and beliefs, no more and no less. In the context of the model risk capital charge under the Basel II Accord the benchmark could be specified by the local regulator; more generally it should be specified by any authority that is concerned with model risk, such as the Chief Risk Officer. Then VaR model risk is assessed using a top-down approach to compute the benchmark VaR from the bank's total daily P&L, and comparing this with the bank's aggregate daily VaR, which is typically derived using a computationally intensive bottom-up approach that necessitates many approximations and simplifications.

The main ideas are as follows: in the presence of model risk an α quantile is at a different quantile of the benchmark model, and has an associated tail probability under the benchmark that is stochastic. Thus, the model-risk-adjusted quantile becomes a generated random variable and its distribution quantifies the bias and uncertainty due to model risk. A significant bias arises if the aggregate VaR estimates tend to be consistently above or below the benchmark VaR, and this is reflected in a significant difference between the mean of the model-risk-adjusted VaR distribution and the benchmark VaR. Even when the bank's VaR estimates have an insignificant bias, an adjustment for uncertainty is required because the difference between the bank's VaR and the benchmark VaR could vary considerably over time. The bias and uncertainty in the VaR model, relative to the benchmark, determine a risk capital adjustment for model risk whose size will also depend on the confidence level regulators require for the adjusted risk capital to be no less than the risk capital based on the benchmark model.

Our framework was validated and illustrated by a numerical example which considers three common VaR models in a simulation experiment where the degree of model risk has been controlled. A further empirical example describes how the model-risk adjustment could be implemented in practice given only two time series, on the bank's aggregate VaR and its aggregate daily P&L, which are in any case reported daily under banking regulations.

Further research would be useful on backtesting the model-risk-adjusted estimates relative to commonly-used VaR models, such as the RiskMetrics models considered in this paper. Where VaR models are failing regulatory backtests and thus being heavily penalized or even disallowed, the top-down model-risk-adjustment advocated in this paper would be very much more cost effective than implementing a new or substantially modified bottom-up VaR system.

There is potential for extending the methodology to the quantile-based metrics that are commonly used to assess non-financial risks in hydrology, climate change, statistical process control and reliability analysis. In the case that relatively few observations on the model and benchmark quantiles are available, the approach should include a parameterization the model-risk-adjusted quantile distribution, for instance as a beta-generated distribution.

ACKNOWLEDGMENTS

The authors would like to thank to the associate editor and two anonymous reviewers for many constructive comments that improved the original version. The second author thanks Ministerio de Economía y Competitividad, Spain (ECO2010-15455) for partial support.

REFERENCES

1. Basel Committee on Banking Supervision. Amendment to the capital accord to incorporate market risks. Bank for International Settlements, Basel, 1996.
2. Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards: A revised framework. Bank for International Settlements, Basel, 2006.
3. Basel Committee on Banking Supervision. Revisions to the Basel II market risk framework. Bank for International Settlements, Basel, 2009.
4. Hendricks D. Evaluation of Value-at-Risk models using historical data. FRBNY Economic Policy Review, 1996; April:39-69.
5. Perignon C, Smith D. The level and quality of Value-at-Risk disclosure by commercial banks. Journal of Banking and Finance, 2010; 34(2):362-377.
6. Pritsker M. The hidden dangers of historical simulation. Journal of Banking and Finance, 2006; 30:561-582.
7. RiskMetrics. Technical Document. Available from the RiskMetrics website, 1997.
8. Hull J, White A. Value at risk when daily changes in market variables are not normally distributed. Journal of Derivatives, 1998; 5(3):9-19.
9. Mittnik S, Paolella M. Conditional density and Value-at-Risk: prediction of Asian currency exchange rates. Journal of Forecasting, 2000; 19:313-333.
10. Venter J, de Jongh P. Risk estimation using the normal inverse Gaussian distribution. Journal of Risk, 2002; 4:1-23.
11. Angelidis T, Benos A, Degiannakis S. The use of GARCH models in VaR estimation. Statistical Methodology, 2004; 1(2):105-128.
12. Hartz C, Mittnik S, Paolella M. Accurate Value-at-Risk forecasting based on the normal-GARCH model. Computational Statistics and Data Analysis, 2006; 51(4):2295-2312.
13. Kuan C-M, Yeh J-H, Hsu Y-C. Assessing Value-at-Risk with CARE: the conditional autoregressive expectile models. Journal of Econometrics, 2009; 150(2):261-270.
14. Kupiec P. Techniques for verifying the accuracy of risk measurement models. Journal of Derivatives, 1995; 3(2):73-84.
15. Christoffersen P. Evaluating interval forecasts. International Economic Review, 1998; 39:841-862.
16. Berkowitz J, Christoffersen P, Pelletier D. Evaluating VaR models with desk-level data. Management Science, Management Science, 2011; 57:2213-2227.
17. Engle R, Manganelli S. CAViaR: Conditional autoregressive Value at Risk by regression quantile. Journal of Business and Economic Statistics, 2004; 22:367-381.
18. Turner L. The Turner review: A regulatory response to the global banking crisis. Financial Services Authority, London, 2009.
19. Taleb N. The black swan: The impact of the highly improbable. Penguin, 2007.
20. Beder T. VAR: Seductive but dangerous. Financial Analysts Journal, 1995; 51(5):12-24.
21. Bernardo JM, Smith AFM. Bayesian Theory. Wiley, Chichester, UK, 1994.
22. O'Hagan A. Kendall's Advanced Theory of Statistics. Vol 2B: Bayesian Inference. Edward Arnold, London, 1994.
23. Modarres R, Nayak TK, Gastwirth JL. Estimation of upper quantiles model and parameter uncertainty. Computational Statistics and Data Analysis, 2002; 39:529-554.
24. Giorgi G, Narduzzi C. Uncertainty of quantile estimates in the measurement of self-similar processes. In: International Workshop on Advanced Methods for Uncertainty Estimation Measurements, AMUEM 2008, Sardagna, Trento, Italy, 2008.
25. Figlewski S. Estimation error in the assessment of financial risk exposure. SSRN-id424363, version of June 29, 2003.
26. Derman E. Model risk. Risk, 1996; 9(5):139-145.
27. Simons K. Model error - evaluation of various finance

models. New England Economic Review, 1997; Nov-Dec:17-28.
28. Crouhy M, Galai D, Mark R. Model risk. Journal of Financial Engineering, 1998; 7:267-288.
29. Green T, Figlewski S. Market risk and model risk for a financial institution writing options. Journal of Finance, 1999; 54:1465-1499.
30. Kato T, Yoshiba T. Model risk and its control. Monetary and Economic Studies, 2000; December:129-157.
31. Rebonato R. Managing Model Risk. In Volume 2 of Mastering Risk, C. Alexander (ed.), Pearson UK, 2001.
32. Jorion P. Measuring the risk in value at risk. Financial Analysts Journal, 1996; 52:47-56.
33. Talay D, Zheng Z. Worst case model risk management. Finance and Stochastics, 2002; 6:517-537.
34. Kerkhof J, Melenberg B, Schumacher H. Model risk and capital reserves. Journal of Banking and Finance, 2010; 34:267-279.
35. Reiss R, Thomas M. Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields. Birkhauser Verlag, Basel, 1997.
36. Cairns A. A discussion of parameter and model uncertainty in insurance. Insurance: Mathematics and Economics, 2000; 27:313-330.
37. Matthys G, Delafosse E, Guillou A, Beirlant J. Estimating catastrophic quantile levels for heavy-tailed distributions. Insurance: Mathematics and Economics, 2004; 34:517-537.
38. Dowd K, Blake D. After VaR: The theory, estimation and insurance applications of quantile-based risk measures. Journal of Risk and Insurance, 2009; 73(2):193-229.
39. Cont R. Model uncertainty and its impact on the pricing of derivative instruments. Mathematical Finance, 2006; 16(3):519-547.
40. Hull J, Suo W. A methodology for assessing model risk and its application to the implied volatility function model. Journal of Financial and Quantitative Analysis, 2002; 37(2):297-318.
41. Branger N, Schlag C. Model risk: A conceptual framework for risk measurement and hedging. Working Paper, EFMA Basel Meetings. Available from SSRN, 2004.
42. Ebrahimi N, Maasoumi E, Soofi E. Ordering univariate distributions by entropy and variance. Journal of Econometrics, 1999; 90:317-336.
43. Shannon C. The mathematical theory of communication, Bell System Technical Journal July-Oct. Reprinted in: C.E. Shannon and W. Weaver, The Mathematical Theory of Communication University of Illinois Press, 1948; Urbana, IL 3-91.
44. Zellner A. Maximal data information prior distributions. In: A. Aykac and C. Brumat, eds., New methods in the applications of Bayesian methods. North-Holland, 1977.
45. Jaynes E. Papers on Probability, Statistics and Statistical Physics. Edited by R.D. Rosenkrantz, Reidel Publishing Company, Dordrecht, 1983.
46. Rockinger M, Jondeau E. Entropy densities with an application to autoregressive conditional skewness and kurtosis. Journal of Econometrics, 2002; 106:119-142.
47. Wu X. Calculation of maximum entropy densities with application to income distribution. Journal of Econometrics, 2003; 115(2):347-354
48. Chan F. Modelling time-varying higher moments with maximum entropy density. Mathematics and Computers in Simulation, 2009; 79(9):2767-2778.
49. Chan F. Forecasting value-at-risk using maximum entropy. 18th World IMACS / MODSIM Congress, Cairns, Australia, 2009.
50. Park S, Bera A. Maximum entropy autoregressive conditional heteroskedasticity model. Journal of Econometrics, 2009; 150(2):219-230.
51. Lee M-C, Su J-B, Liu H-C. Value-at-risk in US stock indices with skewed generalized error distribution. Applied Financial Economics Letters, 2008; 4(6):425-431.
52. Mandelbrot B. The variation of certain speculative prices. Journal of Business, 1963; 36:394-419.
53. Bollerslev T. A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics, 1987; 69(3):542-547.
54. Berkowitz J, O'Brien J. How accurate are Value-at-Risk models at commercial banks? Journal of Finance, 2002; 55:1093-1111.
55. Barone-Adesi G, Bourgoin F, Giannopoulos K. Don't look back. Risk, 1998; 11(8):100-103.
56. Hull J, White A. Incorporating volatility updating into the historical simulation method for Value-at-Risk. Journal of Risk, 1998; 1:5-19.
57. Alexander C, Sheedy E. Developing a stress testing framework based on market risk models. Journal of Banking and Finance, 2008; 32(10):2220-2236.
58. Eugene N, Lee C, Famoye F. The beta normal distribution and its applications. Communications in Statistics, Theory and Methods, 2002; 31:497-512.
59. Jones M. Families of distributions arising from distributions of order statistics. Test, 2004; 13(1):1-43.
60. Zografos K, Balakrishnan, N. On families of beta- and generalized gamma-generated distributions and associated inference. Statistical Methodology, 2009; 6(4):344-362.
61. Nelson DB. Conditional heteroskedasticity in asset returns: A new approach. Econometrica, 1991; 59:347-370.
62. Arsenault M, Ashkar F. Regional flood quantile estimation under linear transformation of the data. Water Resources Research, 2000; 36(6):1605-1610.
63. Chebana F, Ouarda T. Multivariate quantiles in hydrological frequency analysis. Environmetrics, 2011; 22(1):63-78.
64. Katz R, Parlange M, Naveau P. Statistics of extremes in hydrology. Advances in Water Resources, 2002; 25:1287-1304.
65. Diodato N, Bellocchi G. Assessing and modelling changes in rainfall erosivity at different climate scales. Earth Surface Processes and Landforms, 2009; 34(7):969-980.
66. Bremnes J. Probabilistic wind power forecasts using local quantile regression. Wind Energy, 2004; 7(1):47-54.
67. Rumyantsev A. Safety prediction in nuclear power. Atomic Energy, 2007; 102(2):94-99.
68. Anghel C. Statistical process control methods from the viewpoint of industrial application. Economic Quality Control, 2001; 16(1):49-63.
69. Wheelock D, Wilson P. Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations. Journal of Econometrics, 200; 145:209-225.
70. Unnikrishnan Nair N, Sankaran P. Quantile-based reliability analysis. Communications in Statistics - Theory and Methods, 2009; 38(2):222-232.
71. Mkhandi S, Kachroo R, Guo S. Uncertainty analysis of flood quantile estimates with reference to Tanzania. Journal of Hydrology, 1996; 185, 317-333.
72. Pinson P, Kariniotakis G. On-line assessment of prediction risk for wind power production forecasts. Wind Energy, 2004; 7(2):119-132.
73. Noon B, Cade B, Flather C. Quantile regression reveals hidden bias and uncertainty in habitat models. Ecology, 2009; 86(3):786-800.
74. Merrick J, Van Dorp R, Dinesh V. Assessing uncertainty in simulation-based maritime risk assessment. Risk Analysis, 2005; 25(3):731-743.
75. Peng L, Qi Y. Confidence regions for high quantiles of a heavy tailed distribution. Annals of Statistics, 2006; 34(4):1964-1986.

76. Vermaat M, Does R, Steerneman A. A modified quantile estimator using extreme-value theory with applications. Economic Quality Control, 2005; 20(1):31-39.
77. Sveinsson O, Salas J, Boes D. Uncertainty of quantile estimators using the population index flood method. Water Resources Research, 2003; 39(8), 1206 doi:10.1029/2002WR001594
78. Kapur JN. Maximum Entropy Models in Engineering. John Wiley, New York, 1989.
79. Zografos K. On maximum entropy characterization of Pearson's type II and VII multivariate distributions. Journal of Multivariate Analysis, 1999; 71:67-75.
80. Aulogiaris G, Zografos K. A maximum entropy characterization of symmetric Kotz type and multivariate Burr distributions. Test, 2004; 13:65-83.
81. Bhattacharya B. Maximum entropy characterizations of the multivariate Liouville distributions. Journal of Multivariate Analysis, 2006; 97:1272-1283.
82. Ahmed NA, Gokhale DV. Entropy expressions and their estimators for multivariate distributions. IEEE Transactions on Information Theory, 1989; 35:688-692.
83. Zografos K, Nadarajah S. Expressions for Renyi and Shannon entropies for multivariate distributions. Statistics and Probability Letters, 2005; 71:71-84.