Endogenizing Model Risk to Quantile Estimates

Carol Alexander
ICMA Centre, Henley Business School at Reading, UK

José María Sarabia
Department of Economics, University of Cantabria, Spain

July 9, 2010

ABSTRACT

We quantify and endogenize the model risk associated with quantile estimates using a maximum entropy distribution (MED) as benchmark. Moment-based MEDs cannot have heavy tails, however generalized beta generated distributions have attractive properties for popular applications of quantiles. These are MEDs under three simple constraints on the parameters that explicitly control tail weight and peakness. Model risk arises because analysts are constrained to use a model distribution that is not the MED. Then the model's α quantile differs from the α quantile of the MED so the tail probability under the MED associated with the model's α quantile is not α, it is a random variable, $\hat{\alpha}$. Model risk is endogenized by parameterizing the uncertainty about $\hat{\alpha}$, whence the quantile becomes a generated random variable. To obtain a point model-risk-adjusted quantile, the generated distribution is used to adjust the model's α quantile for any systematic bias and uncertainty due to model risk. An illustration based on Value-at-Risk (VaR) computes a model-risk-adjusted VaR for risk capital reserves which encompass both portfolio and VaR model risk.

Key Words: Quantile risk measures; model risk; maximum entropy; generalized beta normal (GBN) distributions; generalized beta generated (GBG) distributions; Value-at-Risk (VaR); risk capital; S&P 500 index; GARCH; RiskMetrics

JEL Codes: C1, C19, C51, G17, G28

Carol Alexander
Chair of Risk Management,
ICMA Centre, Henley Business School at the University of Reading, Reading, RG6 6BA, UK.
Email: c.alexander@icmacentre.rdg.ac.uk

José María Sarabia
Professor of Statistics, Department of Economics, University of Cantabria, Spain
Email: jose.sarabia@unican.es

1 Introduction

Quantile estimation has long been applied to a variety of disciplines.[1] The probability of the occurrence of extreme events is of prime interest for actuaries, where heavy-tailed distributions are used to model large claims and losses (for instance, see Matthys et al., 2004). Heavy-tailed distributions are frequently used for quantile estimation in hydrology and climate change (Katz et al., 2002) and in finance, where the quantiles of loss distributions are called 'Value-at-Risk' (VaR) assessments.[2] Another area where quantile risk assessments are frequently applied is statistical process control, for computing capability indices (Anghel, 2001), for measuring efficiency (Wheelocker and Wilson, 2008) and for reliability analysis (Unnikrishnan Nair and Sankaran, 2009). Several papers acknowledge the uncertainty surrounding quantile-based risk assessments: see Mkhandi et al. (1996), Derman (1996), Jorion (1996), Green and Figlewski (1999), Brooks and Persand (2002), Peng and Qi (2006), Kerkhof et al. (2010) and others.

The term 'model risk' is frequently applied to encompass various sources of uncertainty in statistical models, but there is no general concensus on its precise definition.[3] For our purposes, model risk stems from two sources: model choice and parameter estimation error. These are defined as follows:

- *Model choice* refers to the model risk stemming from inappropriate assumptions about the form of the statistical model for the random variable;

- *Parameter estimation error* refers to the model risk stemming from uncertainty in the parameter values of the chosen model. We can never estimate parameter values exactly because we are constrained by a finite set of observations on the variable.

This purpose of this paper is to derive an explicit distribution for the quantile, which is generated by the uncertainty surrounding model choice and parameter estimation error, and from which a point estimate for the model-risk-adjusted quantile may be obtained, if desired. Our work presents a completely new research strand which fuses two original ideas: the use of maximum entropy as a benchmark for model risk and the application of generated distributions to model quantiles when tail probabilities are uncertain.

---

[1] A survey of their appliction to insurance, actuarial science, finance, hydrology and several other fields is given by Reiss and Thomas (1997).

[2] The literature on VaR is vast: see Christoffersen (2009) for a survey.

[3] For instance: in insurance, though not in finance, it is common to separate model choice from parameter uncertainty, as in Cairns (2000) for instance; and Kerkhof et al. (2010) include 'identification risk' which arises when observationally indistinguishable models have different consequences for capital reserves.

Some authors identify model risk with the departure of a model from a 'true' dynamic process: see Branger and Schlag (2004) 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 realisations of the data generation process of a random variable. It is futile to propose the existence of a unique and measurable 'true' data generation process because such an exercise is beyond our realm of knowledge. Each analyst's knowledge is subjective, hence no process that encapsulates the state of knowledge of the world and which could therefore be labelled the 'true' process exists.

However, an analyst *is* endowed with a maximum entropy distribution (MED), i.e. a distribution based on no more and no less than the information available to him regarding the random variable's behaviour, including any beliefs. In short, the MED is the distribution which best represents the analyst's *state of knowledge.* Under our definition, model risk arises because this MED is not an available model. That is, the analyst is constrained – by external regulations, internal management, computational complexity and other practical features of a large organization or system – to use a distribution that is *not* his MED. Specific practical examples of these constraints will be discussed below.

In the following: Section 2 gives a formal definition of quantile model risk and outlines a framework for its quantification; The innovative use of maximum entropy as a benchmark for assessing model risk is discussed in Section 3, aiming to provide advice for senior managers, regulators, policy makers and any other parties responsible for model risk assessment. We show that the popular moment-based MEDs are not appropriate if heavy tails are included in the state of knowledge and instead we recommended the very flexible class of generalized beta generated distributions, which are MEDs under three intuitive shape constraints; A model for the probability $\hat{\alpha}$ that is assigned under the MED to the α quantile of the model distribution is introduced in Section 4. 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 distribution detects any systematic bias in the model's α quantile, relative to the α quantile of the MED, and 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 MED at a pre-determined confidence level; Section 5 illustrates the application of our framework to VaR model risk, deriving a model-risk-adjusted VaR, with separate bias and uncertainty components, that encompasses both portfolio risk and model risk; Section 6 summarizes and concludes with recommendations for further research.

2 Quantile Model Risk

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

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

Typically the 'true' distribution for X in (1) is unknown, except in simulation or experimental environments. Statistical models provide an estimate $\hat{F}$ of F, and to use this to compute the α quantile. That is, instead of (1) we use $q_\alpha^{\hat{F}} = \hat{F}^{-1}(\alpha)$. We have argued that a 'true' distribution for X, known or unknown, is beyond our realm of knowledge. We either accept this fact or reject the concept of model risk entirely: and if we accept it, then the benchmark for model risk assessment must be that F which embodies no more and no less than our complete state of knowledge, i.e. the *maximum entropy* distribution (MED).

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.[4] Because $\mathcal{K}$ includes beliefs its choice could be very subjective, particularly at the level of an individual analyst. Ideally, a manager's, regulator's or other policy maker's state of knowledge would provide the benchmark MED against which to assess model risk on a firm-wide or industry-wide basis.

We simply characterise a 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. Quantile model risk arises because $\{\hat{F}, \hat{\mathcal{K}}\} \neq \{F, \mathcal{K}\}$ for two reasons. 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 $\hat{\mathcal{K}}$;[5] Secondly $\hat{F}$ is not, typically, the MED even based on $\hat{\mathcal{K}}$, e.g. the execution of firm-wide models for a large institution may present such a formidable time challenge that senior managers may require that $\hat{F}$ be based on a simplified data generation processes. Model risk may also spill 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.

In practice, 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. However, in the presence of model risk the quantile of the MED

---

[4]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.

[5]As is the case under current banking regulations for the use of VaR to estimate risk capital reserves - see Basel Committee on Banking Supervision (1996).

corresponding to the quantile estimated by the model is not the α quantile. That is, $q_\alpha^{\hat{F}} \neq q_\alpha^F$, and the model's α quantile $q_\alpha^{\hat{F}}$ is at a different quantile of F. We use the notation $\hat{\alpha}$ for this quantile, i.e. $q_\alpha^{\hat{F}} = q_{\hat{\alpha}}^F$, or equivalently,

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

If $\hat{F}$ were identical to the MED there would be no model risk, and $\hat{\alpha} = \alpha$ for every α. In general 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

$$e(\alpha|F, \hat{F}) = \hat{\alpha} - \alpha. \tag{3}$$

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. For instance, in example 2 below we report the standard deviation of the errors, the mean absolute error (MAE), the root mean squared error (RMSE) and the range.[6]

3 Selecting the Maximum Entropy Distribution

Shannon (1948) defined the *entropy* of a probability density function $g(x)$, $x \in \mathcal{R}$ as

$$H(g) = \mathrm{E}_g[\log g(x)] = -\int_{\mathcal{R}} g(x) \log g(x) dx. \tag{4}$$

The Shannon entropy (henceforth called simply entropy) is a measure of the uncertainty in a probability distribution and its negative is a measure of information.[7] The *maximum*

---

[6]If the model has a significant bias the MAE, RMSE and range should be applied with caution because they include the bias. Anyway, when the bias is large it is better that the model be inefficient: the worst possible case is an efficient but biased model.

[7]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 (1999) for further insight.

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 that is available to the analyst. The criterion here is to be as vague as possible (i.e. to maximize uncertainty) given the constraints imposed by the testable information. This way, the maximum entropy distribution (MED) represents no more (and no less) than this information. If the testable information consists only of an 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 conditions imposed by the analyst's beliefs about the shape of the MED. Often the MED is assumed to belong to some parametric family.

An empirical MED based on historical data alone has the advantage of making no parametric assumption, but it limits beliefs about the future to what has been experienced in the past. Instead we advocate using the new class of generalized beta generated (GBG) distributions introduced by Alexander and Sarabia (2010), and in particular the generalized beta normal (GBN) distribution, for three reasons: (a) they are MEDs under three fairly general shape constraints; (b) their shape parameters have an intuitive interpretation as controlling the peakness, skew and tail weights of the distribution; and (c) they are flexible enough to fit many distributional shapes. As such they may be fit to sample data, as in example 1 below, or their parameters may be set (or modified from sample estimates) by formulating beliefs about the shape of the distribution.

The simplest model in this class is the beta normal distribution, introduced by Eugene, Lee and Famoye (2002). It is generated by replacing the uniform $U[0,1]$ distribution in the normal probability integral transform by a beta distribution $\mathcal{B}(p,q)$, with density

$$f_{\mathcal{B}}(u;p,q) = B(p,q)^{-1}[u^{p-1}(1-u)^{q-1}], \quad 0 < u < 1, \ p,q \geq 0 \tag{5}$$

where $B(p,q)$ is the beta function. Denote the standard normal distribution and density functions by $\Phi(.)$ and $\phi(.)$ respectively, and recall the probability integral transform: if $X = \Phi^{-1}(U)$ with $U \sim U[0,1]$ then X has density function ϕ. The beta normal distribution generalizes this, so that we sample from any beta distribution. That is, if $X = \Phi^{-1}(U)$ with $U \sim \mathcal{B}(p,q)$ then X has a beta normal distribution with density function:

$$g_{\Phi}(x) = B(p,q)^{-1}\phi(x)[\Phi(x)]^{p-1}[1-\Phi(x)]^{q-1}, -\infty < x < \infty, \tag{6}$$

5

Replacing $\Phi(x)$ and $\phi(x)$ in (6) by any other 'parent' distribution $F(x)$ and density $f(x) = F'(x)$, yields the general class of beta generated distributions introduction by Jones (2004).

However, the use of a classical beta distribution in place of the uniform still limits the behaviour of the generated random variable. For instance, the skewness and kurtosis of beta normal random variables (tabulated by Eugene, Lee and Famoye, 2002) lie in a very narrow range. Changing the parent distribution so that it is not normal but skewed and leptokurtic loses the tractability of a normal parent, as well as the connection between p and q and the skewness and tail weight of the generated distribution. For the specification of relevant information, which is captured by the entropy constraints, these properties may seem important to retain. So instead of changing the parent Alexander and Sarabia (2010) change the generator, to a generalized beta distribution (of the first kind, introduced by McDonald, 1984) which has density function

$$f_{\mathcal{GB}}(u; a, p, q) = B(p, q)^{-1}[au^{ap-1}(1 - u^a)^{q-1}], \quad 0 < u < 1, \ a, p, q \geq 0. \tag{7}$$

In general, given any continuous parent distribution $F(x)$, $x \in \mathcal{R}$ with density $f(x) = F'(x)$, X has a GBG distribution when the probability transformed variable $U = F(X)$ has density (7). Then the distribution of X may be characterised by its density function:

$$f_{\mathcal{GBG}}(x; a, p, q) = B(p, q)^{-1}f(x)[aF(x)^{ap-1}(1 - F(x)^a)^{q-1}], \ x \in \mathcal{R}. \tag{8}$$

Two special cases are: the classical beta generated distributions ($a = 1$) and the Kumaraswamy generated distributions ($p = 1$). Alexander and Sarabia also show that a GBG is the MED under the following three contraints:

$$-\mathrm{E}_{\mathcal{GBG}}[\log F(X)^a] = \zeta(p, q), \tag{9}$$

$$-\mathrm{E}_{\mathcal{GBG}}[1 - \log F(X)^a] = \zeta(q, p), \tag{10}$$

$$\mathrm{E}_{\mathcal{GBG}}[\log f(X)] - \mathrm{E}_U[\log f(F^{-1}(U^{1/a}))] = a^{-1}(a - 1)^2\zeta(p, q) \tag{11}$$

with $U \sim \mathcal{B}(p, q)$ and $\zeta(p, q) = \psi(p+q) - \psi(p)$ where ψ denotes the digamma function. Furthermore, the GBG has the maximum entropy of all distributions satisfying the information constraints (9) – (11) and its entropy is:

$$
\begin{aligned}
-\mathrm{E}_{\mathcal{GBG}}[\log f_{\mathcal{GBG}}(x)] = {} & \log B(p, q) - \log a + a^{-1}(a - 1)\zeta(p, q) \\
& + (p - 1)\zeta(p, q) + (q - 1)\zeta(q, p) - \mathrm{E}_U[\log f(F^{-1}(U^{1/a}))]. \tag{12}
\end{aligned}
$$

It follows that the GBG entropy is the sum of the entropy of the generalized beta generator, which is independent of the parent, and another term $-\mathrm{E}_U[\log f(F^{-1}(U^{1/a}))]$ that is related to the entropy of the parent. Furthermore, the constraints (9) and (10) reflect information only about the generalized beta generator: (9) is related to the information in the left tail and (10) is related to information in the right tail. The tail weights are symmetric if F is symmetric and $p = q$, and a controls the peakness of the distribution: as it increases so does the weight in the centre. Note that (12) takes its maximum value of zero when $a = p = q = 1$; otherwise, the structure in the generator adds valuable information to the shape of the GBG distribution.

The case of the normal generator $N(\mu, \sigma^2)$ where $X \sim \mathcal{GBN}(a, p, q; \mu, \sigma^2)$ is of particular interest. Since $(X - \mu)/\sigma \sim \mathcal{GBN}(a, p, q; 0, 1)$ the generalized beta distribution controls the shape parameters but does not affect the location and scale parameters of X, which are inherited from the $N(\mu, \sigma^2)$ parent distribution. We now consider how one might specify beliefs about the weight in the centre and the tails of the MED by choosing suitable values for the generalized beta parameters a, p and q given a normal parent. The following example fits the GBN parameters a, p, q, μ, σ^2 to two quite different samples on the S&P 500 index returns and interprets the entropy components of each sample. It suggests that a pragmatic approach to specifying the MED would be to fit a GBN to historical data and then adjust parameters to reflect subjective beliefs; e.g. increase $\hat{q}$ and/or decrease $\hat{p}$ to increase the negative skew, and increase (decrease) $\hat{a}$ to increase (decrease) the peakness of the density.

EXAMPLE 1: **GBN maximum entropy distributions**
Consider two samples of daily returns on the daily returns on the SP 500 index: Sample 1 (January 2004 - December 2006) and Sample 2 (January 2007 to December 2009). The samples are chosen to represent two very different market regimes: in the first sample markets were relatively stable with a low volatility; the second sample encompasses the credit crunch and the banking crisis, when market risk factors were extremely volatile.

Table 1 reports standard sample statistics. Then $\hat{\mu}$ and $\hat{\sigma}$ are set equal to the sample mean and standard deviation and the parameters a, p and q are estimated using the method of moments.[8] The last section of the table reports the lower and upper tail components of the generalized beta entropy, i.e. $\zeta(\hat{p}, \hat{q})$ and $\zeta(\hat{q}, \hat{p})$ in (9) and 10), and the term $-\log \hat{a} + \hat{a}^{-1}(\hat{a}-1)\zeta(\hat{p}, \hat{q})$ in (12) which is zero for the classical beta generated distribution but which

---

[8]We fit the three parameters of (7) to the standardized return $Z = (X - \mu)/\sigma$ and afterwards translate the standard $\mathcal{GBN}(a, p, q; 0, 1)$ distribution to a general $\mathcal{GBN}(a, p, q; \mu, \sigma^2)$ model using $X = Z\sigma + \mu$.

Table 1: S&P 500 daily returns: descriptive statistics,
GBN parameter estimates and entropy decomposition.

	Sample 1 Jan 2004 - Dec 2006	Sample 2 Jan 2007 - Dec 2009
Mean	0.00034	-0.00014
Std. Dev.	0.00659	0.01885
Skewness	0.00259	0.05921
XS Kurtosis	0.25954	6.35811
$\hat{a}$	3.02177	2.32358
$\hat{p}$	0.33261	0.44417
$\hat{q}$	1.00898	1.05067
$\zeta(\hat{p}, \hat{q})$	3.01637	2.30010
$\zeta(\hat{q}, \hat{p})$	0.43951	0.52847
$-\log \hat{a} + \hat{a}^{-1}(\hat{a}-1)\zeta(\hat{p}, \hat{q})$	0.91232	0.46710

more generally controls the information in the centre of a GBG distribution.

Considering first the estimates of the parameter a, which increases with the weight in the centre of the GBN distribution, it is greater for sample 1 than it is for sample 2. By the same token, the last row of the table indicates that there is more information in the centre of sample 1 than there is in sample 2. Since both samples have positive skewness, $\hat{p} < \hat{q}$ in both cases. Finally, $\zeta(\hat{p}, \hat{q})$ is greater in sample 1 and $\zeta(\hat{q}, \hat{p})$ is greater in sample 2, thus sample 1 has less uncertainty surrounding negative returns but more uncertainty about positive returns, compared with sample 2.

Following the theoretical work of Shannon (1948), Zellner (1977), Jaynes (1983) and others, an analyst may be tempted to assume the testable information is given by a set of moment functions derived from a sample, in addition to the normalization condition. If only the first two moments are deemed relevant, the MED is a normal distribution (Shannon, 1948). More generally, when the testable information contains the first n sample moments, $f(x)$ takes an exponential form. This is found by maximizing (4) subject to the conditions:

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

where $\mu_0 = 1$ and μ_n $(n = 1, ..., N)$ are the moments of the distribution. The solution is:

$$f(x) = \exp\left(-\sum_{n=0}^{n=N} \lambda_n x^n\right) \tag{13}$$

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

$$\mu_n = \int x^n \exp\left(-\sum_{n=0}^{n=N} \lambda_n x^n\right) dx, \quad n = 0, \ldots, N. \tag{14}$$

Rockinger and Jondeau (2002), Wu (2003) several others have applied the four-moment MED to econometric and risk management problems. Park and Bera (2009) and Chan (2009a) apply the four-moment MED to the generalized autoregressive conditional heteroscedasticity (GARCH) model of Bollerslev (1987). Chan (2009b) extends this to the computation of VaR. But, perhaps surprizingly, none of these papers consider the tail weight that is implicit in the use of a four-moment MED.

Our next empirical example quantifies the model risk of two popular VaR models relative to a four-moment MED.[9] We have selected these VaR models because they are commonly adopted, having been popularized by the 'RiskMetrics' methodology introduced by JP Morgan in the mid-1990's – see RiskMetrics (1997).

EXAMPLE 2: **VaR model risk relative to a four-moment MED**
We consider a portfolio tracking the S&P 500 index and a risk manager with the (ostensibly sensible) belief that all information relevant to the VaR assessment over a 1-week horizon is contained in the first four moments of the weekly portfolio returns over the past year. We shall quantify the model risk of a VaR model, relative to this four-moment MED, by measuring the error (3) over a long period of time. To this end we base all parameter estimates on a one-year rolling window of weekly returns: starting with January - December 2001, the sample is rolled over weekly until we reach the last sample from January - December 2009. In total there are $520 - 52 + 1 = 469$ different rolling windows of 52 observations where the window at time t contains the weekly returns $x_{t-1}, \ldots, x_{t-52}$.

The two VaR models are based on an assumed zero-mean normal distribution for the portfolio returns, so the $\alpha\%$ VaR estimate at time t is $\Phi^{-1}(\alpha)\hat{\sigma}_t$ where $\hat{\sigma}_t$ is the standard deviation of the portfolio returns and Φ is the standard normal distribution function. In

---

[9]The $\alpha\%$ VaR of a portfolio is the loss that would be equalled or exceeded with probability α, assuming the portfolio is left unmanaged over a predetermined risk horizon. When VaR is expressed as a percentage of the current portfolio value, it is -1 times the α quantile of the distribution of the portfolio return over the risk horizon.

VaR model 1, the 'Regulatory' model,[10] $\hat{\sigma}_t$ is the square root of the arithmetic average of $x_{t-1}^2, \ldots, x_{t-52}^2$. We denote the normal distributions thus generated by $\hat{F}_{1t}$. In VaR model 2 the variance is assumed to follow an exponentially weighted moving average (EWMA) process given by the recursion:

$$\hat{\sigma}_t^2 = (\eta - 1)x_{t-1}^2 + \eta\hat{\sigma}_{t-1}^2, \tag{15}$$

for some $0 < \eta < 1$. Following RiskMetrics (1997) we set $\eta = 0.97$ and the normal distributions generated by VaR model 2 are denoted $\hat{F}_{2t}$. It is commonly held that model 2 has less model risk than model 1, because its standard deviation is more responsive to current market conditions. For each model we estimate time series of 1% and 5% VaR over a 1-week horizon at each time t, i.e. $-\hat{F}_{1t}^{-1}(0.05)$, $-F_{1t}^{-1}(0.01)$, $-\hat{F}_{2t}^{-1}(0.05)$ and $-\hat{F}_{2t}^{-1}(0.01)$. These time series are depicted in Figure 1.

Now consider the maximum entropy density (13) with $N = 4$. This is fit to each sample by computing the first four moments $\mu_1, \ldots, \mu_4$ of $x_{t-1}, \ldots, x_{t-52}$ and solving (14).[11] The estimates of $\lambda_0, \ldots, \lambda_4$ at time t, denoted $\hat{\lambda}_t$, are displayed in Figure 2.[12] Denoting by $F(x; \hat{\lambda}_t)$ the MED with parameters $\hat{\lambda}_t$, for each model $(i = 1, 2)$ we compute $\hat{\alpha}_t = F(\hat{F}_{it}^{-1}(\alpha); \hat{\lambda}_t)$, the tail probability under the MED associated with the α quantile estimate from model $\hat{F}_{it}$. Thus we obtain four time series of tail probabilities $\hat{\alpha}$ under the fitted four-moment MED corresponding to each of the two VaR models for $\alpha = 1\%$ and 5%.

Table 2 reports some sample statistics on the quantile probability errors (3). Model 2 errors have the smaller standard deviation and range, indicating less uncertainty due to model risk than model 1, yet model 2 has the greater bias. In each case the mean is very significantly different from zero, and negative, indicating a substantial overestimation of the MED VaR at both quantiles in each model. In fact, the Max statistics show that *all* the quantile errors in each sample are negative, so there is not even one time t when the probability associated with either VaR estimate under the four-moment MED is greater than 1% or 5% respectively. This is because the belief that returns are light tailed is implicit in the four-moment MED assumption, yet the samples have heavy-tailed distributions.

---

[10]So called because it employs the minimum one-year data period that is required under the Basel Committee recommendations.

[11]We have used an adapted version of the Matlab code for the Zellner and Highfield (1988) algorithm, given by Mohammad-Djafari (2001).

[12]For brevity we do not report the time series of sample moments. We simply remark that the samples have positive excess kurtosis, indicating empricial distributions that are heavry tailed.

Table 2: Sample statistics on (3) from a four-moment MED

VaR Model	Model 1		Model 2	
Quantile	**5%**	**1%**	**5%**	**1%**
Mean	-3.03%	-0.78%	-3.20%	-0.82%
Std. Dev.	0.99%	0.17%	0.74%	0.15%
MAE	3.03%	0.78%	3.20%	0.82%
RMSE	3.19%	0.80%	3.28%	0.83%
Max	-0.15%	-0.29%	-1.31%	-0.33%
Min	-4.83%	-1.00%	-4.66%	-0.99%
Range	4.68%	0.71%	3.34%	0.66%

The above example has demonstrated how to quantify the model risk of two popular VaR models relative to a four-moment MED and has illustrated the undesirable effects of using moment-based MEDs for model risk assessment of heavy-tailed random variables. In fact:

- 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.

- 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.

4 Endogenizing Model Risk

The previous example generated a time series of of observations on $\hat{\alpha} = F(\hat{F}^{-1}(\alpha))$ where uncertainty about the value of $\hat{\alpha}$ was generated by model risk. We now formalize this by regarding $\hat{\alpha}$ as a random variable with a distribution that is generated by our two sources of model risk, i.e. model choice and parameter estimation error. Because $\hat{\alpha}$ is a probability it has range $[0, 1]$; so we may approximate its distribution by a generalized beta distribution with density (7). Assuming $\hat{\alpha} \sim \mathcal{GB}(a, p, q)$, the α quantile of our model, adjusted for model risk, becomes a random variable:

$$Q(\alpha|F, \hat{F}) = F^{-1}(\hat{\alpha}), \ \ \hat{\alpha} \sim \mathcal{GB}(a, p, q). \tag{16}$$

That is, the model-risk-adjusted quantile $Q(\alpha|F, \hat{F})$ is a random variable whose GBG distribution is generated from the MED F and the parameters of the generalized beta representation of $\hat{\alpha}$. Our quantile $q_\alpha^{\hat{F}}$ $(= q_{\hat{\alpha}}^{F})$ is an observation on this random variable.

11

If G_B denotes the distribution of $\mathcal{GB}(a, p, q)$, the distribution function of $Q(\alpha|F, \hat{F})$ is

$$G_F(v; a, p, q) = \Pr(F^{-1}(\hat{\alpha}) \leq v) = G_B(F(v); a, p, q), \quad v \in \mathcal{R}, \tag{17}$$

and, denoting by f the density of F, the density of $Q(\alpha|F, \hat{F})$ is

$$g_F(v; a, p, q) = B(p, q)^{-1} f(v) F(v)^{ap-1} [1 - F(v)^a]^{q-1}, \quad v \in \mathcal{R}. \tag{18}$$

We are interested in the mean of $Q(\alpha|F, \hat{F})$, given by

$$E[Q(\alpha|F, \hat{F})] = B(p, q)^{-1} \int_{\mathcal{R}} v f(v) F(v)^{ap-1} [1 - F(v)^a]^{q-1} dv, \tag{19}$$

because it quantifies any systematic bias in the quantile estimates: e.g. if the MED has a heavier tails than the model then extreme quantiles $q_\alpha^{\hat{F}}$ will be biased: if α is near zero $E[Q(\alpha|F, \hat{F})] > q_\alpha^F$ and if α is near one $E[Q(\alpha|F, \hat{F})] < q_\alpha^F$. 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.

Unfortunately, analytic solutions for (19) are only available for some particular choices of F and some values of a, p and q. However, it is always possible to obtain an approximate expression for (19). We write, $E[Q(\alpha|F, \hat{F})] = \mathrm{E}_{\mathcal{GB}(a,p,q)}[F^{-1}(\hat{\alpha})]$. Upon expanding $F^{-1}(\hat{\alpha})$ in a Taylor series around the point $\hat{\alpha}_m = E(\hat{\alpha})$, we obtain the following a series expansion for $F^{-1}(\hat{\alpha})$ in terms of its derivatives $F^{-1(i)}$, for $i = 1, \ldots, 4$:

$$\begin{aligned} F^{-1}(\hat{\alpha}) &= F^{-1}(\hat{\alpha}_m) + F^{-1(1)}(\hat{\alpha}_m)(\hat{\alpha} - \hat{\alpha}_m) + \frac{1}{2} F^{-1(2)}(\hat{\alpha}_m)(\hat{\alpha} - \hat{\alpha}_m)^2 \\ &+ \frac{1}{6} F^{-1(3)}(\hat{\alpha}_m)(\hat{\alpha} - \hat{\alpha}_m)^3 + \frac{1}{24} F^{-1(4)}(\hat{\alpha}_m)(\hat{\alpha} - \hat{\alpha}_m)^4. \end{aligned} \tag{20}$$

Taking expectations in (20) yields

$$\begin{aligned} -E[Q(\alpha|F, \hat{F})] &\approx F^{-1}(\hat{\alpha}_m) + \frac{1}{2} F^{-1(2)}(\hat{\alpha}_m) \sigma_{\hat{\alpha}}^2 \\ &+ \frac{1}{6} F^{-1(3)}(\hat{\alpha}_m) \gamma_1(\hat{\alpha}) \sigma_{\hat{\alpha}}^3 + \frac{1}{24} F^{-1(4)}(\hat{\alpha}_m) \gamma_2(\hat{\alpha}) \sigma_{\hat{\alpha}}^4, \end{aligned} \tag{21}$$

where $\sigma_{\hat{\alpha}}^2 = E[(\hat{\alpha} - \hat{\alpha}_m)^2]$, $\gamma_1(\hat{\alpha}) = \sigma_{\hat{\alpha}}^{-3} E[(\hat{\alpha} - \hat{\alpha}_m)^3]$, and $\gamma_2(\hat{\alpha}) = \sigma_{\hat{\alpha}}^{-4} E[(\hat{\alpha} - \hat{\alpha}_m)^4]$.

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 risk estimates are typically constructed to be conservative,

we add an *uncertainty buffer* to the bias-adjusted α quantile in order to be $(1-y)\%$ confident that the model-risk-adjusted α quantile is no less than q_α^F. That is, we add to the bias-adjusted α quantile estimate a quantity equal to the difference between the mean of $Q(\alpha|F, \hat{F})$ and $G_F^{-1}(y)$, the $y\%$ quantile of $Q(\alpha|F, \hat{F})$. Finally, our point estimate for the model-risk-adjusted α quantile becomes:

$$q_\alpha^{\hat{F}} + \overbrace{\{q_\alpha^F - E[Q(\alpha|F, \hat{F})]\}}^{\text{bias adjustment}} + \overbrace{\{E[Q(\alpha|F, \hat{F})] - G_F^{-1}(y)\}}^{\text{uncertainty buffer}} = q_\alpha^{\hat{F}} + q_\alpha^F - G_F^{-1}(y). \quad (22)$$

The model-risk-adjusted quantile will be greater than the unadjusted quantile when $G_F^{-1}(y)$ is less than q_α^F; however, negative adjustments may result from a high value for y, i.e. a low degree of confidence for the model-risk-adjusted quantile to exceed the maximum entropy quantile. Roughly speaking, if $q_\alpha^{\hat{F}}$ overshoots q_α^F more than $(1-y)\%$ of the time, the negative bias adjustment outweighs the uncertainty buffer. Note that the computation of $E[Q(\alpha|F, \hat{F})]$ can be circumvented except when the decomposition into bias and uncertainty components is required. Also, the confidence level $1-y$ here reflects a penalty for model risk which is a matter for subjective choice.

5 Application to Value-atRisk

The portfolio VaR corresponds to an amount that could be lost, with a specified probability, if the portfolio remains unmanaged over a specified time horizon. VaR 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, used since the 1996 amendment to the Basel I Accord,[13] will be extended to include stressed VaR under an amended Basel II and the new Basel III Accords.[14]

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. A popular topic is the introduction of new VaR models,[15] and another

---

[13]See Basel Committee on Banking Supervision (1996).

[14]See Basel Committee on Banking Supervision (2009).

[15]Historical simulation (Hendricks, 1996) is the most popular approach amongst banks (Perignon and Smith, 2010) but data-intensive and prone to pitfalls (Pritsker, 2006). Filtering the data makes it more risk

very prolific strand of literature focuses on testing their accuracy.[16] However, the stark failure of many banks to set aside sufficient capital reserves during the banking crisis of 2008 sparked an intense debate on the risk of using of VaR models for this purpose. Turner (2009) is critical of the manner in which VaR models have been applied and Taleb (2007) even questions the very idea of using statistical models for risk assessment. Despite the warnings of Turner, Taleb and other critics of VaR models,[17] most financial institutions continue to employ them as their primary tool for market risk assessment with the proviso, following Basel II recommendations, that VaR model risk be formally considered.[18]

When choosing a VaR model, returns on financial assets and interest rates are commonly assumed to have a simple parametric distribution. For instance, normal VaR models are consistent with the assumption of geometric Brownian motion for pricing derivative products. Another common choice is to base VaR calculations on historical simulation, thus making the simplistic assumption that history will repeat itself with certainty. In general, the model risk associated with model choice arises because risk analysts are obliged, by senior management and/or regulators, to make simpifying but pragmatic assumptions about the behaviour of portfolio returns, and because the firm-wide use of more realistic but complex VaR models requires computational resources that are not available. When VaR model parameters are estimated from historical data, parameter estimation error is a result of constraints on these data: the sampling error stemming from size limitations imposed by internal management, regulators, or the limited data availability of new or illiquid products. Moreover, the regime-switching nature of financial assets implies that much historical data are irrelevant for the forward-looking risk horizon. Given the regulatory requirement that banks address these common sources of VaR model risk, banks and their regulators now require a framework that can answer the following questions:

1. Given a VaR model, are some portfolios more sensitive to its model risk than others?

---

sensitive (Barone-Adesi et al. (1998), Hull and White (1998a)) and this impoves its accuracy (Alexander and Sheedy, 2008). Other popular VaR models assume normal risk factor returns with the RiskMetrics EWMA or Regulatory covariance matrix estimates (RiskMetrics, 1997). More complex VaR models are proposed by Hull and White (1998b), Mittnik and Paollela (2000), Christoffersen et al. (2001), Berkowitz and O'Brien (2002), Ventner and de Jongh (2002), Angelidis et al. (2004), Hartz et al. (2006), Kuan et al. (2009) and many others.

[16]This is frequently based on the coverage tests introduced by Kupiec (1998) and refined by Christoffersen (1998), Christoffersen, Hahn and Inoue (2001) and Christoffersen and Pelletier (2004). Berkowitz and O'Brien (2002) and Berkowitz et al. (2010) focus on the accuracy of VaR models used by major banks.

[17]Even the use of a quantile has been subject to criticism, as it is not necessarily sub-additive, and related metrics such as conditional VaR may be preferred. See Beder (1995) and Artzner et al. (1999).

[18]See Basel Committee on Banking Supervision, 2006.

2. Given a portfolio, which of the available VaR models has least model risk?

3. How can VaR model risk be included in risk capital requirements?

Relative to the vast literature on VaR model construction and applications there has been little academic research on VaR model risk. Derman (1996), Simons (1997), Crouhy et al. (1998), Green and Figlewski (1999), Kato and Yoshiba (2000) and Rebonato (2001) examine the general causes of model risk in finance but in these papers a formal definition of model risk that allows its assessment is elusive. The quantification of model risk in the risk-neutral measure has been addressed by Hull and Suo (2002), who define model risk relative to the price distribution implied by market prices of vanilla options, and Cont (2006) who quantifies the model risk of a contingent claim by the range of prices obtained under all possible valuation models. Few papers deal explicitly with VaR model risk, notable exceptions being Jorion (1996) and Talay and Zheng (2002), who examine parameter error, and Brooks and Persand (2002) who empirically assess the VaR estimates obtained from different VaR models. Yet the only paper that attempts to address the questions posed above is Kerkhof et al. (2010), who derive a model risk 'add on' for capital reserves computed as the adjustment necessary for two simple econometric models to pass regulatory backtests. We remark that this approach has virtually nothing in common with ours.

We now illustrate how the concepts introduced in Section 4 may be used to provide answers to the questions posed above, by means of an experiment in which a portfolio's returns are simulated based on a known data generation process. The advantage of this experiment is that we can control the degree of VaR model risk relative to the MED, but our method for estimating model-risk-adjusted risk capital estimates would apply equally well in practice, when MEDs are derived from informational constraints. Our purpose is to illustrate and empirically validate the bias and inefficiency adjustments to VaR for computing risk capital that includes both portfolio and model risk components.

It is widely accepted that of all the parsimonious discrete-time variance processes the asymmetric GARCH class provide the best fit to asset returns. Thus we shall assume that our 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. First the return x_t from time t to $t + 1$ and its variance σ_t^2 are simulated using:

$$\sigma_t^2 = \omega + \alpha(x_{t-1} - \lambda)^2 + \beta\sigma_{t-1}^2, \quad x_t|\mathcal{I}_t \sim \mathcal{N}(0, \sigma_t^2), \qquad (23)$$

where $\omega > 0, \alpha, \beta \geq 0, \alpha + \beta \leq 1$ and $\mathcal{I}_t = (x_{t-1}, x_{t-2}, \ldots)$.[19] For the simulated returns the parameters of (23) are assumed to be:

$$\omega = 1.5 \times 10^{-6}, \alpha = 0.04, \lambda = 0.005, \beta = 0.95, \tag{24}$$

and so the steady-state annualized volatility of the portfolio return is 25%.[20] 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 $\mathcal{I}_t$ and the model $X_t|\mathcal{I}_t \sim \mathcal{N}(0, \hat{\sigma}_t^2)$. We now consider three analysts employing difference processes for $\hat{\sigma}_t^2$. The first analyst has the correct choice of model but uses incorrect parameter values: instead of (24) he employs the fitted model:

$$\hat{\sigma}_t^2 = \hat{\omega} + \hat{\alpha}(x_{t-1} - \hat{\lambda})^2 + \hat{\beta}\hat{\sigma}_{t-1}^2, \tag{25}$$

with

$$\hat{\omega} = 2 \times 10^{-6}, \hat{\alpha} = 0.0515, \hat{\lambda} = 0.01, \hat{\beta} = 0.92. \tag{26}$$

His steady-state volatility estimate is therefore correct, but since $\hat{\alpha} > \alpha$ and $\hat{\beta} < \beta$ his volatility process is more 'jumpy' than the simulated variance 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 second analyst uses a simplified version of (23) with:

$$\hat{\omega} = \hat{\lambda} = 0, \hat{\alpha} = 0.06, \hat{\beta} = 0.94. \tag{27}$$

That is, the analyst employs the RiskMetrics EWMA estimator of example 2 with $\eta = 0.94$ in (15), under which a steady-state volatility is not defined. The third analyst uses the RiskMetrics 'Regulatory' estimator, also described in example 2, therefore taking:

---

[19]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.

[20]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.

$$\hat{\alpha} = \hat{\lambda} = \hat{\beta} = 0, \hat{\omega} = \sqrt{\frac{1}{250} \sum_{i=1}^{250} x_{t-i}^2}. \tag{28}$$

A time series of 10,000 returns $\{x_t\}_{t=1}^{10,000}$ is simulated from the 'true' model (23) with parameters (24). Then, for each of the three models defined above we use this time series to (a) estimate the daily VaR using $\Phi^{-1}(\alpha)\hat{\sigma}_t$, 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

$$\hat{\alpha}_t = \Phi\left[\Phi^{-1}(\alpha)\frac{\hat{\sigma}_t}{\sigma_t}\right]. \tag{29}$$

Taking $\alpha = 1\%$ for illustration, the empirical densities of the three VaR estimates are depicted in Figure 4. The AGARCH VaR is less variable than the EWMA VaR, and the Regulatory VaR has a multi-modal distribution, a feature that results from the lack of risk-sensitivity of this model.[21] Now, for each VaR model and for each α we shall use (7) to obtain the density of $Q(\alpha|F, \hat{F})$. But to apply this we first need a density to represent the empirical distribution of $\hat{\alpha}$, given F. Thus for each VaR model we estimate $\hat{\alpha}$ at every time point in the simulations and then fit a generalized beta distribution to these values.

For $\alpha = 0.1\%, 1\%$ and 5%, Table 3, reports the mean and standard deviation of $\hat{\alpha}$, and we quantify model risk using 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. And 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 smallest RMSE. These observations are supported by Figure 5, which depicts the empirical distribution of $\hat{\alpha}$ for $\alpha = 1\%$.

For simplicity, we shall set the generalized beta parameter $a = 1$ and fit $\hat{\alpha}$ with a classical beta distribution. Table 4 reports the method of moment estimates $(\hat{p}, \hat{q})$ that are computed using the means and standard deviations given in Table 3, and Figure 3 depicts the beta distributions corresponding to these values of $(\hat{p}, \hat{q})$. Since $\hat{p} >> \hat{q}$ all beta distributions have a large positive skew. Both $\hat{p}$ and $\hat{q}$ are inversely proportional to α, and to the degree of model risk. Thus, at one extreme we have the highly-peaked beta density

---

[21]Specifically, a single extreme negative return makes the VaR jump to a level which is sustained for exactly 250 days, and then the VaR jumps down again when this return drops out of the in-sample period.

Table 3: Sample statistics for quantile probabilities

		AGARCH	EWMA	Regulatory
0.10%	Mean	0.11%	0.16%	0.23%
	Stdev	0.0007	0.0013	0.0035
	RMSE	0.07%	0.14%	0.37%
1%	Mean	1.03%	1.25%	1.34%
	Stdev	0.0042	0.0059	0.0119
	RMSE	0.42%	0.64%	1.22%
5%	Mean	4.97%	5.44%	5.27%
	Stdev	0.0103	0.0123	0.0268
	RMSE	1.03%	1.31%	2.66%

of the AGARCH model when $\alpha = 5\%$, and at the other we have the 'U'-shaped density of the Regulatory model when $\alpha = 0.1\%$.

Table 4: Values of $(\hat{p}, \hat{q})$ for the distributions given in Table 3.

α		AGARCH	EWMA	Regulatory
0.10%	$\hat{p}$	2.35	1.50	0.43
	$\hat{q}$	2145.39	959.21	188.35
1%	$\hat{p}$	5.92	4.43	1.24
	$\hat{q}$	570.98	350.90	91.04
5%	$\hat{p}$	22.10	18.29	3.61
	$\hat{q}$	422.04	318.15	64.78

We now apply (22) to compute a point estimate for model-risk-adjusted VaR (RaVaR, for short). The maximum entropy VaR (MeVaR, for short) which is based on the simulated process and the three VaR estimates are time-varying, so the RaVaR will depend on the time it is measured. For illustration, we select a time when the simulated volatility is at its steady-state value of 25% – so the MeVaR is is 4.886%, 3.678% and 2.601% at the 0.1%, 1% and 5% levels, respectively – but similar conclusions would be drawn at other volatility levels. Drawing at random from the points when the simulated volatility was 25%, we obtain AGARCH, EWMA and Regulatory volatility forecasts of 27.00%, 23.94% and 28.19% respectively.[22] These volatilities determine the VaR estimates that we shall

---

[22]So the AGARCH and Regulatory models overestimated VaR at this point and the EWMA model underestimated VaR; of course, this was not the case every time the simulated volatility was 25%.

now adjust for model risk.

Again taking $\alpha = 1\%$ for illustration, the simulated densities of $Q(\alpha|F, \hat{F})$ are depicted in Figure 6. The empirical mean is $E[\widehat{Q(\alpha|F, \hat{F})}] = -10^{-4} \sum_{i=1}^{10,000} F^{-1}(\hat{\alpha}_i)$. Alternatively, we could use the beta parameter estimates in Table 4 to derive an analytic beta normal density function. In this case the mean is computed using numerical integration of (19), or via the analytic approximation (21).[23]

Table 5: Comparison of three methods for estimating the mean of $Q(\alpha|F, \hat{F})$

α	Mean	**AGARCH**	**EWMA**	**Regulatory**
0.10%	(Emp.)	4.919	4.793	4.961
	(19)	4.953	4.832	-
	(21)	4.988	5.017	1.292
1%	(Emp.)	3.703	3.608	3.735
	(19)	3.722	3.603	3.729
	(21)	3.710	3.614	4.067
5%	(Emp.)	2.618	2.551	2.641
	(19)	2.626	2.227	2.643
	(21)	2.618	2.552	2.654

All three means are subject to error, albeit in slightly different ways, but the errors are small relative to the VaR model risk. In Table 5 the three estimates are close, except in the Regulatory model at extreme quantiles (1% and 0.1%). The approximation (21) is good for the AGARCH model, but the fit deteriorates as the bias increases: the means for the Regulatory model are similar at the 5% quantile, but becomes more diverse at 1% and at the 0.1% quantile, where the Regulatory model has a value of $\hat{p}$ less than 1, it becomes impossible to integrate (19) using standard numerical methods.

Table 6 summarizes the bias, calculated as the difference between the MeVaR and the empirical mean of $Q(\alpha|F, \hat{F})$.[24] 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 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

---

[23]The derivatives of $\Phi^{-1}(z)$ in (21) are not complicated: $d\phi(z)/dz = -z\phi(z)$, $\Phi^{-1(1)}(z) = \phi(\Phi^{-1}(z))^{-1}$, $\Phi^{-1(2)}(z) = \Phi^{-1}(z)[\phi(\Phi^{-1}(z))]^{-2}$, $\Phi^{-1(3)}(z) = [1 + 2\{\Phi^{-1}(z)\}^2][\phi(\Phi^{-1}(z))]^{-3}$, and $\Phi^{-1(4)}(z) = \Phi^{-1}(z)[7 + 6\{\Phi^{-1}(z)\}^2][\phi(\Phi^{-1}(z))]^{-4}$.

[24]The very similar results based on other means in Table 5 are not reported for brevity.

Table 6: Bias estimates and 5% quantiles for RaVaR computation

α		AGARCH	EWMA	Regulatory
0.1%	Bias	-0.067	0.054	-0.075
	Quantile	0.291	0.332	0.891
1%	Bias	-0.025	0.070	-0.056
	Quantile	0.237	0.270	0.663
5%	Bias	-0.017	0.050	-0.040
	Quantile	0.157	0.174	0.454

α decreases it is not significant for any model.[25] Beneath the bias we report the 5% quantile of the simulated distribution for $Q(\alpha|F, \hat{F})$, since we shall first compute the RaVaR so that it is no less than the MeVaR with 95% confidence.

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

$$\text{RaVaR}(y) = \text{VaR} + \overbrace{(\text{MeVaR} - E[Q(\alpha|F, \hat{F})])}^{\text{bias adjustment}} + \overbrace{(E[Q(\alpha|F, \hat{F})] - G_F^{-1}(y))}^{\text{uncertainty buffer}}.$$

Table 7 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 α is 0.1%, 1% and 5% respectively. The (small) bias is corrected by adding the bias from Table 6 to each VaR estimate. The main source of model risk here concerns the potential for a large (positive or negative) errors in the quantile probabilties, i.e. the dispersion of the densities in Figure 3. To adjust for this we add to the bias-adjusted VaR an uncertainty buffer equal to the difference between the MeVaR and the 5% quantile given in Table 6. 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:

$$\% \text{ risk capital increase} = \frac{\text{MeVaR} - G_F^{-1}(y)}{\text{VaR}} . \tag{30}$$

---

[25]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 3.

Table 7: Computation of 95% RaVaR

α		AGARCH	EWMA	Regulatory
	Volatility	27.00%	23.94%	28.19%
0.10%	VaR	5.277%	4.678%	5.509%
	Bias Adj. VaR	5.244%	4.772%	5.434%
	RaVaR	5.716%	5.387%	6.483%
1%	VaR	3.972%	3.522%	4.147%
	Bias Adj. VaR	3.948%	3.592%	4.091%
	RaVaR	4.303%	4.055%	4.880%
5%	VaR	2.809%	2.490%	2.932%
	Bias Adj. VaR	2.791%	2.540%	2.892%
	RaVaR	3.043%	2.867%	3.451%

Table 8: Percentage increase in risk capital from model risk adjustment of VaR

$1 - y$	AGARCH	EWMA	Regulatory
95%	8.33%	15.15%	17.68%
90%	6.05%	12.62%	14.16%
85%	4.47%	10.99%	11.37%
80%	3.19%	9.57%	8.81%
75%	2.25%	8.38%	6.68%

The penalty (30) 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 8 reports the percentage increase in risk capital due to model risk when RaVaR is no less than the MeVaR with $(1 - y)\%$ confidence. This is directly proportional to the degree of model risk in the VaR model. At 95% confidence, the first row of the table shows that risk capital based on the AGARCH model would be increased by about 8%, by about 15% in the EWMA model and about or 17.5% in the Regulatory model. The other rows in Table 8 shows how much extra risk captial would be required for other reasonable confidence levels. Interestingly, since the EWMA VaR is exceptionally low at the time of the adjustment, the EWMA VaR-based risk capital would require a greater adjustment than the Regulatory VaR-based risk capital if we required only 80% or lower confidence that the RaVaR will be less than the MeVaR.

6 Summary and conclusions

We assert that it is meaningless to postulate the existence of a unique, measurable 'true' distribution as a benchmark for model risk because it is beyond our realm of knowledge, except in simulation experiments. So one can either confine the concept to contolled environments or define model risk with reference to a distribution has meaning in practical situations. We argue that there is no better choice of model risk benchmark than a maximum entropy distribution (MED) as this embodies the entirety of information and beliefs, no more and no less. To reach an industry or firm-wide consensus for model risk assessment requires the adoption of common MEDs for major risk factors. To this end, we provide advice on the choice of MED. We demonstrate that moment-based MEDs make implicit light-tailed assumptions, and we advocate using the flexible class of GBG distributions, as these are maximum entropy under simple and intuitive shape conditions.

Model risk arises when the analyst uses a simplification in place of his MED because he is constrained by management directives and/or the complexity of the system and the available data. In the presence of model risk, a quantile estimate that purports to be at some quantile of the MED is in fact at a different quantile, which has an associated tail probability under the MED that is stochastic. Using a generalized beta distribution to model this tail probability we obtain a model-risk-adjusted quantile that is a GBG random variable whose distribution endogenizes the bias and uncertainty due to model risk. From this distribution we derive a point model-risk-adjusted quantile that is no less than the MED quantile with some pre-specified degree of confidence. This confidence level may be firm-specific, or policy makers and regulators may consider setting industry-wide directives.

We have applied our framework, using real data and in a simulation environment, to the problem of VaR assessment in financial risk analysis. For VaR models with controlled degrees of model risk we have quantified the increase in risk capital when the model-risk-adjusted VaR (RaVaR) is used in place of the standard VaR. This depends on the confidence required for the RaVaR to be no less than the VaR based on the maximum entropy distribution. If this confidence level is high enough that disallowed VaR models have model-risk-adjusted VaRs that pass the regulatory backtests, the use of our framework is likely to be considerably less expensive than investment in a new VaR system.

The use of maximum entropy as a benchmark for estimating model risk presents a new direction for research. This paper has focussed on developing a statistical framework for quantile risk assessments, and future research may seek to extend this to other quantile-based metrics such as conditional VaR. Other interesting topics include the development

of standardized MEDs for major risk factors and the backtesting of model-risk-adjusted estimates for commonly-used models. Quantifying model risk presents a significant challenge for numerous problems in insurance, finance, hydrology and other areas. There is considerable potential for research based on model risk assessment via maximum entropy in all these areas, rather than on the elusive concept of a 'true' model.

References

Alexander, C., and Sheedy, E. (2008). Developing a stress testing framework based on market risk models. *Journal of Banking and Finance.* 32(10), 2220-2236.

Alexander, C. and Sarabia, J.M. (2010). Generalized beta generated distributions. *ICMA Centre Discussion Papers in Finance.*

Angelidis, T., Benos, A., and Degiannakis, S. (2004). The use of GARCH models in VaR estimation. *Statistical Methodology.* 1(2), 105-128.

Anghel, C. (2001). Statistical process control methods from the viewpoint of industrial application. *Economic Quality Control.* 16(1), 49 - 63.

Artzner, P., Delbaen F., Eber, J.M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance.* 9, 203-228.

Beder, Tanya Styblo. (1995). VAR: Seductive but dangerous. *Financial Analysts Journal.* 51(5), 12-24.

Barone-Adesi, G., Bourgoin, F. and Giannopoulos, K. (1998). Don't look back. *Risk.* 11(8), 100-103.

Basel Committee on Banking Supervision (1996). Amendment to the capital accord to incorporate market risks. *Bank for International Settlements*, Basel.

Basel Committee on Banking Supervision (2006). International convergence of capital measurement and capital standards: A revised framework. *Bank for International Settlements*, Basel.

Basel Committee on Banking Supervision (2009). Revisions to the Basel II market risk framework. *Bank for International Settlements*, Basel.

Berkowitz, J., and O'Brien, J. (2002). How accurate are Value-at-Risk models at commercial banks? *Journal of Finance.* 55, 1093-1111.

Berkowitz, J., Christoffersen, P. and Pelletier, D. (2010). Evaluating VaR models with desk-level data. *Management Science.* Forthcoming

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics.* 69(3), 542-547.

Branger, N. and Schlag, C. (2004). Model risk: A conceptual framework for risk measurement and hedging. Working Paper, *EFMA Basel Meetings.* Available from *SSRN.*

Brooks, C., and Persand, G. (2002). Model choice and Value-at-Risk performance. *Financial Analysts Journal.* 58(5), 87-98.

Cairns, A. J. G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics.* 27, 313-330.

Chan, F. (2009a). Modelling time-varying higher moments with maximum entropy density. *Mathematics and Computers in Simulation.* 79(9), 2767-2778.

Chan, F. (2009b). Forecasting Value-at-Risk using Maximum Entropy. *18th World IMACS / MODSIM Congress*, Cairns, Australia.

Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review.* 39, 841-862.

Christoffersen, P., Hahn, J and Inoue A. (2001). Testing and comparing Value-at-Risk measures. *Journal of Empirical Finance.* 8, 325-342.

Christoffersen, P. and Pelletier, D. (2004). Backtesting Value-at-Risk: a duration-based approach. *Journal of Financial Econometrics.* 2(1), 84-108.

Christoffersen, P. (2009). Value-at-Risk models. In *Handbook of Financial Time Series.* Edited by Mikosch, T., Kreiss, J.P., Davis R.A., and T.G. Andersen. Springer, Berlin.

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

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivatiave instruments. Mathematical Finance, 16(3), 519-547.

Derman, E. (1996). Model risk. *Risk*, 9:5 139-145.

Ebrahimi, N., E. Maasoumi and Soo, E.S. (1999). Ordering univariate distributions by entropy and variance. *Journal of Econometrics.* 90, 317-336.

Eugene, N., Lee, C., and Famoye, F. (2002). The beta normal distribution and its applications. *Communications in Statistics, Theory and Methods.* 31, 497-512.

Green, T. C., and Figlewski, S. (1999). Market risk and model risk for a financial institution writing options. *Journal of Finance.* 54, 1465-1499.

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

Hendricks, D. (1996). Evaluation of Value-at-Risk models using historical data. *FRBNY Economic Policy Review.* April. 39-69.

Hull, J. and White, A. (1998a). Incorporating volatility updating into the historical simulation method for Value-at-Risk. *Journal of Risk.* 1, 5-19.

Hull, J. and White, A. (1998b). Value at risk when daily changes in market variables are not normally distributed. *Journal of Derivatives.* 5(3), 9-19.

Hull, J. and Suo, W. (2002). A methodology for assessing model risk and its application to the implied volatility function model. *Journal of Financial and Quantitative Analysis.* 37(2), 297-318.

Jaynes, E.T. (1983) *Papers on probability, statistics and statistical physics.* Edited by R.D. Rosenkrantz, Reidel Publishing Company, Dordrecht.

Jones, M.C. (2004). Families of distributions arising from distributions of order statistics. *Test.* 13(1), 1-43.

Jones, M.C. (2009). Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology.* 6(1), 70-81.

Jorion, P. (1996). Measuring the risk in value at risk. *Financial Analysts Journal* 52, 47-56.

Kato, T. and Yoshiba, T. (2000). Model risk and its control. *Monetary and Economic Studies.* December. 129-157.

Katz, R.W., Parlange M. B. and Naveau P. (2002). Statistics of extremes in hydrology. *Advances in Water Resources.* 25, 1287-1304.

Kerkhof, J., Melenberg, B. and Schumacher, H. (2010). Model risk and capital reserves. *Journal of Banking and Finance.* 34, 267-279.

Kuan, C-M., Yeh, J-H. and Hsu, Y-C. (2009). Assessing Value-at-Risk with CARE: the conditional autoregressive expectile models. *Journal of Econometrics.* 150(2), 261-270.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives.* 3(2), 73-84.

Kumaraswamy, P. (1980). A generalized probability density function for double-bounded random processes. *Journal of Hydrology.* 46, 79-88.

Matthys, G., Delafosse, E., Guillou A. and Beirlant, J. (2004). *Insurance: Mathematics and Economics.* 34, 517-537.

McDonald, J.B. (1984). Some generalized functions for the size distribution of income. *Econometrica.* 52(3), 647-663.

Mittnik, S. and Paolella, M. (2000). Conditional density and Value-at-Risk: prediction of Asian currency exchange rates. *Journal of Forecasting.* 19, 313-333.

Mkhandi, S.H., Kachroo, R.K. and Guo S.L. (1996). Uncertainty analysis of flood quantile estimates with reference to Tanzania. *Journal of Hydrology.* 185, 317-333.

Mohammad-Djafari, A. (2001). A Matlab program to calculate maximum entropy distributions. arXiv:physics/0111126v1 [physics.data-an]

Park, S. and Bera, A. (2009). Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics.* 150(2), 219-230.

Peng, L. and Qi, Y. (2006). Confidence regions for high quantiles of a heavy tailed distribution. *Annals of Statististics.* 34(4), 1964-1986.

Perignon, C. and Smith, D. (2010). The level and quality of Value-at-Risk disclosure by commercial banks. *Journal of Banking and Finance.* 34(2), 362-377.

Pritsker, M. (2006). The hidden dangers of historical simulation. *Journal of Banking and Finance.* 30, 561-582.

RiskMetrics (1997). *Technical Document.* Available from the RiskMetrics website.

Rebonato, R. (2001). Managing Model Risk. In Volume 2 of *Mastering Risk*, C. Alexander (ed.), Pearson UK.

Reiss, R.D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.* Birkhauser Verlag, Basel.

Rockinger, M. and Jondeau, E. (2002). Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics.* 106, 119-142.

Shannon, C.E. (1948). 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, Urbana, IL 3-91.

Simons, K. (1997). Model error - evaluation of various finance models. *New England Economic Review.* Nov-Dec.

Talay, D. and Zheng, Z. (2002). Worst case model risk management. *Finance and Stochastics.* 6, 517-537.

Taleb, N. (2007). *The black swan: The impact of the highly improbable.* Penguin, London.

Turner, L. (2009). The Turner review: A regulatory response to the global banking crisis. *Financial Services Authority*, London.

Unnikrishnan Nair, N. and Sankaran, P.G. (2009). Quantile-based reliability analysis. *Communications in Statistics - Theory and Methods.* 38(2), 222-232.

Venter, J. and de Jongh, P. (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk.* 4, 1-23.

Wheelocka, D.C. and Wilson, P.W. (2008). Non-parametric, unconditional quantile estimation for efficiency analysis with an application to Federal Reserve check processing operations. *Journal of Econometrics.* 145, 209-225.

Wu, X. (2003). Calculation of maximum entropy densities with application to income distribution. *Journal of Econometrics.* 115(2), 347-354

Zellner, A. (1977). Maximal data information prior distributions. In: A. Aykac and C. Brumat, eds., *New methods in the applications of Bayesian methods.* North-Holland, Amsterdam.

Zellner, J. and Highfield, R. (1988). Calculation of maximum entropy distributions and approximation of marginal MRA distributions. *Journal of Econometrics.* 37, 195-209
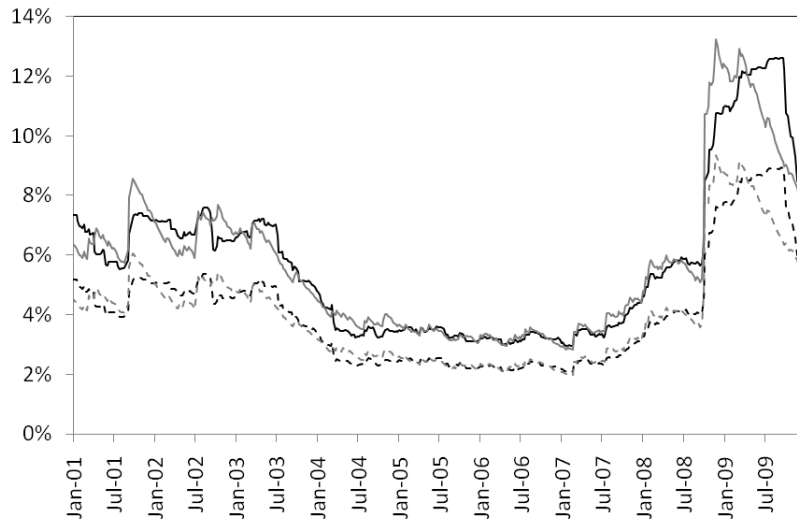
Figure 1: Weekly normal VaR estimates as a percentage of the S&P 500 index at the 1% quantile (solid lines) and 5% quantile (dotted lines) where standard deviation is as described in Model 1 (black lines) and Model 2 (grey lines).
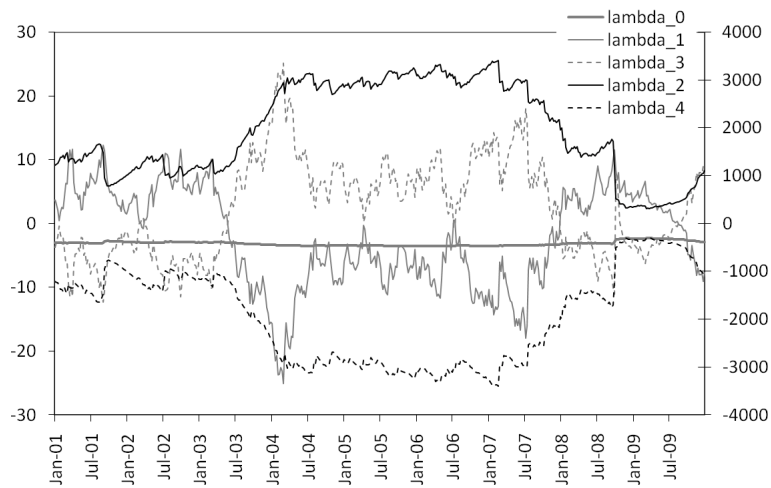

Figure 2: Parameters of the four-moment MED estimated on a rolling window of 250 weekly returns on the S&P 500 index: λ_0, λ_1 and λ_3 on left-hand scale; λ_2 and λ_4 on right-hand scale.
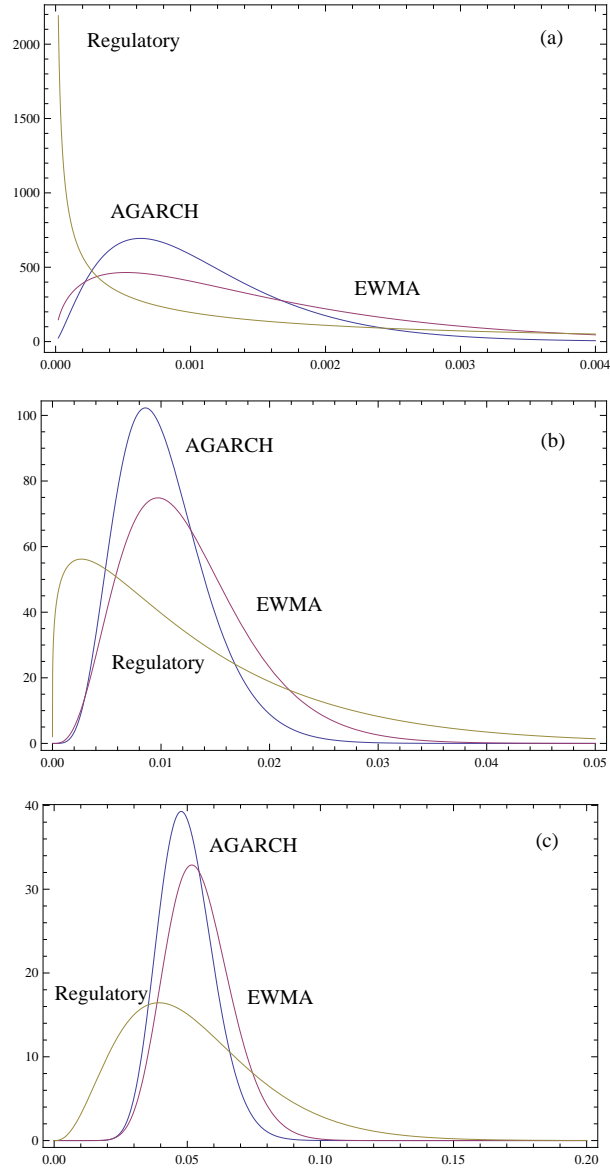
Figure 3: Plots of the beta density for the models AGARCH, EWMA and Regulatory (Table 4) for (a) 0.10%, (b) 1% and (c) 5%.
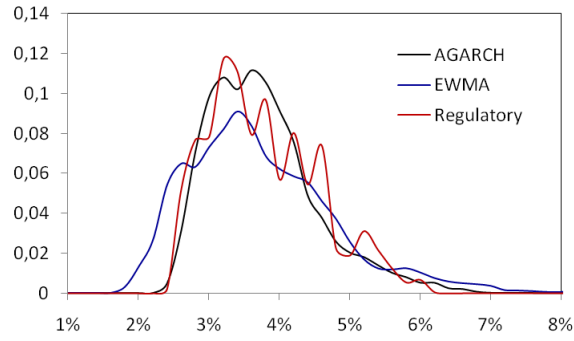
29

Figure 4: Distribution of daily 1% VaR estimates using parameters AGARCH (20), EWMA (21) and Regulatory (22).
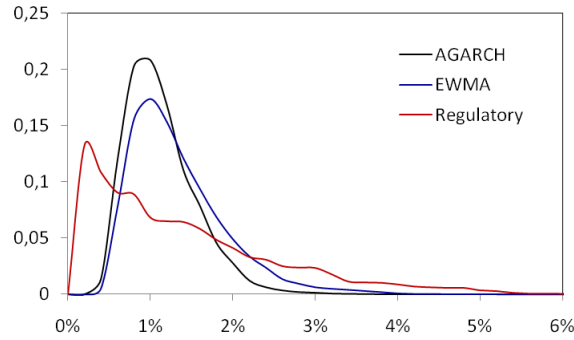

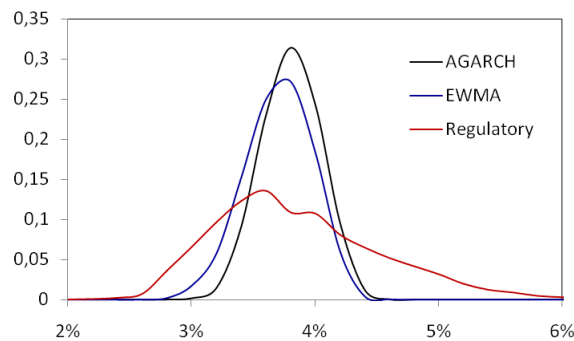Figure 5: Distribution of quantile probabilities for $\alpha = 1\%$ in Figure 4.

Figure 6: Distributions of $Q(\widehat{\alpha | F}, \hat{F})$ under the three VaR models at $\alpha = 1\%$.

30