Bayesian Value-at-Risk and the Capital Charge Puzzle

Matthew Pollard*

13th November 2007

Abstract

This paper presents a general Bayesian estimator for Value-at-Risk and applies it in analysing bank VaR time-series. The estimator optimally incorporates estistimation risk into VaR by integrating over the posterior of each unknown variables. It is shown that Bayesian VaR estimates are uniformly larger in magnitude (more conservative) than usual "plug-in" estimates, which ignore parameter uncertainty. An unusual finding of empirical VaR analysis is that commercial banks' appear to consistently overstate their VaR level ("the Capital Charge puzzle"). Using a sample of 5 international banks' daily VaR and trading revenue, I test whether parameter uncertainty reconciles the apparent overstatement using Bayesian VaR.

1 Introduction

A major concern for financial institutions and their regulators is extreme market events and the adequacy of capital to meet such events. An important tool in measuring the market risk is Value-at-Risk (VaR). Following the Market Risk Amendment to the Basel accord (Basel Committee, 1996b), VaR has become the the adopted industry standard for measuring market risk. The amendment requires commercial banks with large trading portfolios to disclose their daily VaR level, and banks' required capital reserves, or "charges", are set in proportion to this level. Banks are free to specify their own model for VaR estimation and the accuracy of these estimates is important to regulators and banks themselves.

Few commercial banks publicly disclose their VaR and consequently there are very few empirical studies of banks' VaR time-series. Berkowitz and O'Brien (2002) analysed the VaR six anonymous commercial banks with demeaned and standardized revenues. Berkowitz, Christoffersen and Pelletier (2006) use anonymous data from one bank on seven different trading desks. Pérignon, Deng and Wang and Pérignon and Smith (2006), extract data from annual report plot of six Canadian commercial banks and five international banks respectively.

A consistent finding is that banks appear to substantially overstate their VaR-level when measured against ex-post trading revenue. A survey conducted by the Basel Committee (1999) examined forty banks in nine countries and reports that half the banks in the sample did not experience a single day where losses exceeding their VaR level. Berkowitz and O'Brien (2002) found that reported 1% VaR were, on average, 1.3 to 1.6 times larger than the actual percentile of the revenue distribution for four out of the six major US commercial banks in their sample. Pérignon, Deng and Wang (2006) show that the reported VaR at the six largest

* *Email:* matthewcpollard@berkeley.edu,
I would like to thank Daniel Smith and Tom Smith for their invaluable advice, criticism and cuban cigars, as well APRA for the Brian Gray Scholarship and as seminar participants at the UNSW National Honours Colloquium, 2007.

Canadian commercial banks result in capital charges that are 1.25 to 5.0 times higher than those justified by the revenue distribution. An economic consequence is that banks set-aside too much market-risk capital and suffer an investment opportunity cost. This finding has been termed the "Capital Charge Puzzle" (Bakshi and Panayotov, 2007).

At present, there is no agreed-upon explanation for this finding. Berkowitz and O'Brien (2002) argue that that risk estimates appear conservative since the trading revenues used inappropriately contain non-trading income from market-making and net interest income, although later studies use pure trading income and find the effect. Jorion (2004) suggests that over-reporting may only be a temporary phenomenon, due to the excess of economic capital over regulatory capital that banks have been holding in the recent years. Ewerhart (2002) provides an adverse selection argument: since banks cannot credibly communicate their risk exposure, the more prudent banks signal their quality by reporting conservative VaRs. Perignon and Smith (2007) argue that diversification across different trading activities in a bank is not taken into account in calculating aggregate VaR numbers.

This paper proposes another explanation: banks' are conservative due *estimation risk* arising from having to estimate many unknown parameters. The trading portfolios of large banks are complex, with positions that change on daily or intra-daily bases. These portfolios also typically include options and credit derivatives. Aggregating the riskiness of each position into a single VaR level is extremely difficult (Gourieroux & Jasiak, 2001) and involve many unknown variables. For example, consider the simple VaR model where trading revenue for each position is drawn from a multivariate-normal distribution with constant mean and variance. For N positions, there are $N^2 + N$ parameters to estimate (N^2 variance-covariances and N means). It seems reasonable that that risk managers face considerable uncertainty in the distribution of revenue. Given an assumed model for the revenue distributions (multivariate normal, multivariate-GARCH), the uncertainty arises from parameters being unknown.

This paper quantifies the effect of parameter uncertainty using Bayesian inference. Bakshi and Panayotov (2007) also propose that parameter uncertainty explains the Capital Charge Puzzle. Their investigation *models* parameter uncertainty by considering processes with fat tails and jumps and fitting these to returns by maximum likelihood. The method used here directly *measures* parameter uncertainty using posterior distributions.

We derive the optimal parameter-uncertainty VaR estimate in a Bayesian framework. The estimator, called the "Bayesian VaR", is applicable to any parametric model for the distribution of trading revenue. The properties of this estimator are studied and we prove that the size of Bayesian VaR estimates are always larger than the VaR estimated by ignoring uncertainty, i.e. by "plugging-in" estimator values. This is first proven in the case of i.i.d. normal returns, and then generalized to all models where bank revenue is conditionally normal. This result has has important consequences for the current approach to "back-testing" VaR models, which ignores the effects of parameter uncertainty.

The estimator requires calculation of the Bayesian posterior distribution of each unknown parameter. Except for the simple model of i.i.d normal revenues, the posteriors are very hard (or impossible) to analytically calculate. Markov Chain Monte Carlo (MCMC) is instead used to perform this calculation. MCMC generates samples from the posterior distribution by simulating a Markov chain. The chain is constructed so that it has an equilibrium distribution equal to the desired posterior. We presents a method to calculate sequential Bayesian VaR as new information arrives, such as new revenue observations.

To test the parameter-uncertainty hypothesis, we have obtained dataset of daily trading revenue and VaR for five large commercial banks, courtesy of Pérignon and Smith (2006).

Whether banks' overstate their VaR is tested by fitting four "naive" models: historical simulation, filtered historical simulation, GARCH(1,1) and IGARCH(1). These models are termed "naive" since the information set for estimation consists only of historical trading revenue. Previous empirical studies, such as Brian & Berkowitz (2002) and Perignon & Smith (2006), directly compare time series of banks' reported VaR to the estimates from a "naive" model (both use a GARCH(1,1) model).

We argue direct comparison of the VaR series is misguided. The banks' VaR is estimated conditionally on their information set, such as composition of the trading portfolio. Without knowledge of this set, it is impossible to reject a particular VaR at time t as "too conservative" as there always exists a conditioning information set that justifies a given VaR level. Instead, we adopt an *economic* test of overstatement: are the the banks' capital charges lower under the alternative models. Banks incur an investment opportunity cost by over allocating capital reserves. The 1996 Basel Amendment specifies an explicit formula for capital charges as a function of daily VaR. We use this formula to calculate the daily capital changes for each bank using their reported VaR and the alternative model VaRs.

The capital charges for each bank are first calculated ignoring parameter uncertainty. The model parameters fit out-of-sample and the capital charges are calculated under each model. Using the Bayesian VaR estimator, the parameter-uncertainty adjusted VaR is then calculated for each bank using the integrated-GARCH(1) model and an window of 200 trading day. Each VaR estimate is calculated by running Metropolis-Hastings MCMC algorithm over the new window.

The chapter is structured as follows. Section 4.1 formally defines VaR and discusses the regulatory framework placed by the Basel Accords. Section 4.2 discusses the capital charge puzzle and section 4.3 presents the dataset used. Section 4.4 considers four popular VaR models and 4.5 discusses tests for VaR miss-specification. Section 4.6 analyses the banks' reported VaR against four alternative "naive" models and calculates the Basel amendment capital charges for each model. Section 4.7 presents our work on Bayesian VaR. Section 4.7.1 derives the Bayesian VaR estimator, section 4.7.2 shows how to estimate it using MCMC, 4.7.3 prove three theorems regarding the Bayesian VaR and 4.7.4 estimates and compares Bayesian and plug-in VaR in the case of *i.i.d.* normal returns. Section 4.8 considers misspecification tests on Bayesian VaR and proves that, due to estimation risk, Bayesian VaR have coverage probability less than α. Section 4.9 presents our analysis method of the bank data using Bayesian VaR under an IGARCH(1) model and section 4.10 presents the findings.

2 VaR and Regulatory Framework

Definition 1 *Let r_{t+h} denote the dollar revenue realized at time $t+h$, α be a significance level, and $p(r_{t+h} < x|\mathcal{I}_t)$ denote the probability of loss exceeding a level x, conditioned on information $\mathcal{I}_t$. The Value-at-Risk at level α is*

$$VaR_{t+1|t}(\alpha) = x \quad \text{such that } p(r_{t+1} < x|\mathcal{I}_t) = \alpha.$$

Value-at-Risk is a negative number measured in dollars, the time period h usually is 1 or 10 days and the level α usually equals 1% or 5%. If the VaR at 1% is one million dollars, this amounts to the statement "there is a 1% chance over the next period of realizing a loss in *excess* of one million dollars." An overstated VaR refers to an estimate that has greater magnitude than the true value.

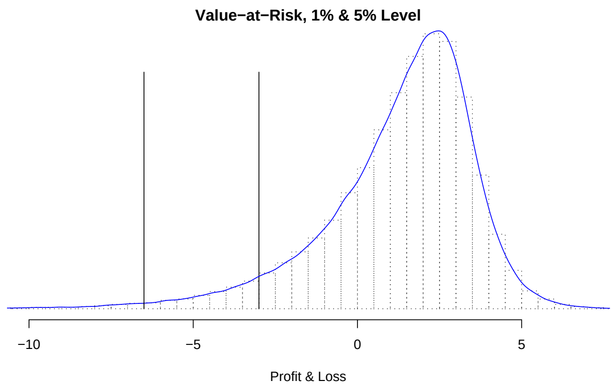


Figure 1: Value-at-Risk at the α =1% (*left)* and α =5% (*right*) level. Value-at-Risk is the worst possible loss in events occurring with probability greater than α. VaR is not informative for rarer events, or losses occurring with probability less than α.

The current supervision framework was created in the *Amendment to Capital Accord to Incorporate Market Risks* (Basel Committee, 1996b) and uses Value-at-Risk the principle measure of risk. The amendment requires banks with substantial trading activity to set aside capital as insurance for extreme portfolio losses. The size of this reserve, or capital charge, is set in proportion to the VaR of the portfolio. The amendment grants banks freedom to use internal models in measure its exposure to market risks and requires this to be summarized as a 1% Value-at-Risk over a 10 day horizon.

The amendment directly specifies capital charges as a function of 1% Value-at-Risk. The capital charge is is proportional to both the level of estimated 1% VaR and the quality of previous past estimates, measured through a "back-test." The back-test compares the expected number of days where losses exceed the VaR ("exceedences") to the realized number. The capital charge is set as the the larger of either is a function of three values: the current Value-at-Risk, $VaR_t(0.01)$, the last 60 day average value at risk and a multiplier term.

$$CC_t := \max\left(VaR_t(0.01), M_t \times \frac{1}{60}\sum_{t=0}^{59} VaR_{t-i}(0.01)\right)$$

At the 1% level, a properly specified VaR model should experience losses in excess of VaR on 2.5 trading days per every 250 trading days (effectively a year). The back-test quality test punishes banks with $N > 4$ exceptions per year by increasing the multiplier M_t. There are three distinct categories for performance:

$$M_t = \begin{cases} 3.0 & \text{if} \quad N \leq 4 \quad \text{green} \\ 3 + 0.2(N-4) & \text{if} \quad 5 \leq N \leq 9 \quad \text{yellow} \\ 4.0 & \text{if} \quad 10 < N \quad \text{red} \end{cases}$$

In the event that more than 10 exceptions at 1% are recorded the span of 250 days, the VaR model is deemed inaccurate and immediate steps are required to improve the risk management system.

3 Value-at-Risk Models

Models used in VaR estimation consist of two statistical statements. At each time t, the model specifies:

1. The distribution of r_t, conditioned on a vector parameters Θ and state variables X_t : $p(r_t|\Theta, X_t)$.

2. A process describing how X_t evolves in time, or transition density $f(X_{t+1}, X_t)$.

The conditioning information $\mathcal{I}_t$ in definition 1 is summarized the conditional moments of $p(r_t|\Theta, X_t)$. State-variable X_t are defined as any parameter that affects $p(r_t)$ that and changes in time. The usual candidate is conditional volatility, although studies have considered higher moments such as conditional skewness and kurtosis (Bali, Mo & Tang, 2006).

Given $p(r_t|\Theta, X_t)$, and full-knowledge in Θ and X_{t+1}, the Value-at-Risk at time t is the α quantile over the lower tail of $p(r_{t+1}|\Theta, X_{t+1})$:

$$VaR_{t+1}(\alpha|\Theta, X_{t+1}) = \{x : \int_{-\infty}^{x} p(r_{t+1}|\Theta, X_{t+1}) = \alpha\}$$

where X_{t+1} has a distribution specified by $f(X_t, X_{t+1})$. The only difference between VaR models is choice of $p(r_t|\Theta, X_t)$ and $f(X_t, X_{t+1})$.

Different conditional distributions for r_t change the different tail probabilities. Distributions with excess kurtosis or with negative skew have thinner, longer lower tails and estimate larger VaR sizes. For example, the 1% tail quantile of the normal distribution and the Student-t distribution with 4 degrees of freedom with unit variance and zero mean are:

$$\begin{aligned} \text{Normal} &= -2.33 \\ \text{Student-}t &= -2.65 \end{aligned}$$

The VaR under the student-t model is 14% higher than under the normal model.

The distribution $p(r_t)$ of bank revenue is unlikely to be stable in time. Two important reasons for this are:

1. The composition a banks' trading portfolio changes, often rapidly;

2. The distribution of market returns is not stable.

If the portfolio weights on risky assets varies in time and if each asset has, different revenue distribution, then the aggregate revenue distribution varies. In particular, trading portfolios that contain options (which have highly asymmetric payoffs), can dramatically alter the aggregate distribution shape after purchase or sale (O'Donnell, 2003). The distribution of market returns is can rapidly change. In particular, conditional volatility of returns is not constant and can rapidly increases in times of market distress (Jacquier, Polson and Rossi, 2002). It is also documented that other higher order moments, such as conditional skewness and kurtosis, change over time (Campbell and Siddique, 1999; Smith, 2006). Both (1) and (2) cause the the conditional distribution of a banks' aggregate revenue r_t, to be extremely complicated and time-varying.

Four popular models for VaR estimation that capture time-variation in $p(r_t)$ are: historical-simulation , filtered historical-simulation, GARCH(1,1) and IGARCH(1) ("Risk-Metrics"). Historical simulation is a non-parametric model using the empirical quantiles over a sliding window.

Filtered historical-simulation is a semi-parametric model where returns are recalled by conditional volatility and historical simulation is applied. The GARCH and IGARCH are parametric models where the conditional volatility follows an autoregressive process.

Historical Simulation. This models the distribution r_{t+1} as the empirical distribution of historical revenue over a window $[t-K, t]$. The state X_{t+1} is a vector of empirical quantiles, and the model estimates $p(r_{t+1})$ as

$$\hat{p}(r_{t+1} < x) = \hat{\Phi}_{[t-K,t]}(x) = \frac{\#\{r_\tau \leq x\}_{\tau=t-K+1}^{t}}{K}$$

The VaR estimator is,

$$V\hat{a}R^{HS}_{t+1}(\alpha) \quad = \quad \{x: \;\; \hat{\Phi}_{[t-K,t]}(x) = \alpha\} = \hat{\Phi}^{-1}_{[t-K,t]}(\alpha)$$

HS is a non-parametric estimator: given a sufficient observations drawn from stable distribution, it can estimate the true VaR with arbitrary precision. Unfortunately, the estimator is very inefficient compared to well specified parametric alternatives. Empirical distributions are poor in estimating tail probabilities since there are few observations from these regions.

The estimator allows for the distribution $p(r_t)$ to change shape by using a sliding window over observations. HS simulation is very slow to to adjust for new shapes $p(r_t)$ and is biased when moments shift rapidly (Pritsker, 2001). These faults withstanding, historical simulation is currently the most popular VaR estimator with banks and regulatory authorities (Campbell, 2005).

Filtered Historical Simulation This is a semi-parametric model proposed by Borone-Adesi, Giannopoulos and Vosper (1999), where historical simulation is applied returns scaled by estimated volatility, $r_t/\hat{h}_t^{1/2}$, and then the time-series is multiplied by $\hat{h}_t^{1/2}$. Specifically,

$$V\hat{a}R^{FHS}_{t+1}(\alpha) = \hat{\Phi}^{-1}_{[t-K,t]}\left(\frac{r_\tau}{\hat{h}_t}, \alpha\right) \times \hat{h}_t.$$

Filtered HS remedies the "stickiness" of historical simulation in reacting to volatility changes, while keeps the flexible, non-parametric feature. $\hat{h}_t$ is usually estimated by a GARCH(1,1) or IGARCH(1). Performance of the method hinges on having low standard error in $\hat{h}_t$. An IGARCH(1) model for implementation of this model.

GARCH(1,1) A popular discrete-time stochastic volatility model proposed by Engel (1982) and Bollerslev (1986). The model specifies that returns conditioned on the mean and instantaneous volatility are normal:

$$r_t|\mu, h_t \sim \mathcal{N}(\mu, h_t^2)$$

where the state variable h_t^2 follows the autoregressive moving-average process

$$h_t^2 = \alpha_0 + \alpha h_{t-1}^2 + \beta(r_{t-1} - \mu)^2$$

The next period volatility prediction is $\hat{h}_{t+1}^2 = \hat{\alpha}_0 + \hat{\alpha}\hat{h}_t^2 + \hat{\beta}(r_t - \hat{\mu})^2$ where $(\alpha_0, \alpha, \beta, \mu)$ are estimated from the data. The VaR estimator is

$$V\hat{a}R^{G}_{t+1}(\alpha) = \hat{\mu} + \hat{h}_{t+1} Z_\alpha.$$

where Z_α is the standard-normal α quantile, $Z_{0.01} = -2.33$. The model suffers from unit-root degeneration ("blows up") if $(\alpha+\beta) > 1$. Models fitted over stock returns are very close to unit root behaviour.

IGARCH(1) (also"Risk-Metrics"): an integrated-GARCH model proposed by Engel and Bollerslev (1986) and popularized for VaR applications by J.P. Morgan in RiskMetrics software. The coefficients on past variance states h_{t-1}^2 and shocks $(r_{t-1}-\mu)^2$ sum to one:

$$h_t^2 = (1-\beta)h_{t-1}^2 + \beta(r_{t-1}-\mu)^2, \;\; V\hat{a}R_{t+1|t} = \hat{\mu} + \hat{h}_{t+1}Z_\alpha$$

Conditional volatility (or variance) is highly persistent in an IGARCH(1) model due to unit-root behaviour in the above equation. However, there is typically very little difference between the IGARCH(1) and GARCH(1,1), since GARCH(1,1) estimated parameters sum very closely to one (Engle & Bollerslev, 1986).

4 Bayesian Method for Estimating VaR

This section presents a general Bayesian method for estimating Value-at-Risk under parameter uncertainty. This estimate is called the "Bayesian VaR", and the method is applicable for any parametric model for trading revenue.

The idea behind Bayesian VaR is as follows. Under parameter uncertainty and data Y_t, the predictive distribution for the subsequent revenue, $p(r_{t+1}|Y_t)$, is wider than corresponding predictive distribution when the parameter and state values are known with certainty, or $p(r_{t+1}|\Theta, X_{t+1})$. In the Bayesian framework, the optimal predictive distribution for r_{t+1} integrates over uncertainty in X_t and Θ, or

$$p(r_{t+1}|Y_t) = \int_{\Theta, X_{t+1}} p(r_{t+1}|\Theta, X_{t+1})p(\Theta, X_{t+1}|Y_t)d\Theta dX_{t+1}$$

The Bayesian VaR estimate is defined as the α quantile of this predictive distribution, or

$$V\tilde{a}R_{t+1}(\alpha|Y_t) = \{x : \int_{-\infty}^{x} p(r_{t+1}|Y_t)dr_{t+1} = \alpha\}.$$

The usual approach to estimating VaR is to first estimate the unknown parameter values in the model and then substitute them into a conditional formulae for VaR,

$$V\hat{a}R_{t+1}(\alpha) = \{x : \int p(r_{t+1}|\hat{\Theta}, \hat{X}_{t+1})dr_{t+1} = \alpha\},$$

This estimator ignores uncertainty in Θ and X_{t+1}. For example, under the GARCH(1,1) model, the distribution of returns conditional on mean μ and volatility h_{t+1}, is $r_{t+1} \sim \mathcal{N}(\mu, h_{t+1}^2)$. The true VaR given μ and h_{t+1} is

$$VaR_{t+1}(\alpha|h_{t+1}, r_{t+1}) = \mu + h_{t+1}Z_\alpha.$$

When μ and h_{t+1} are unknown, a popular estimator proposed Engel (2001) is obtained by plugging-in estimate $E(\mu|Y_t) = \mu$ and forecast $E(h_{t+1}|Y_t) = \hat{h}_t$ into the above formula, or

$$V\hat{a}R_{t+1}(\alpha) = \hat{\mu} + \hat{h}_t Z_\alpha.$$

This estimator is equivalent to the approximating the predictive distribution for r_{t+1} as

$$p(r_{t+1}|Y_t) \approx p(r_{t+1}|\hat{\mu}, \hat{h}_t).$$

These plug-in estimators under-estimate the total risk facing decision makers: market risk *and* estimation risk. Bayesian VaR incorporate estimation risk into the overall VaR. One consequence is that Bayesian VaR are uniformly lower than the equivalent plug-in estimators, or

$$VaR_{t+1|t}(\alpha, Y_t) \leq VaR_{t+1}(\alpha, \hat{\Theta}).$$

This inequality is proven in section 4.3, first under the simplifying case of *i.i.d* normal returns, and then generally for state-space models where, conditioned on state-variables and parameters, revenue is conditionally normal. The inequality, of course, does not *quantify* the the difference between Bayesian VaR and plug-in VaR. This depends on the posterior $p(\Theta|Y)$ and hence priors $p(\Theta)$; typically, this posterior has no closed form expression. MCMC simulation allows the difference to be quantified and section 4.2 presents Monte Carlo estimator for Bayesian VaR. An important exception is where trading revenues are he i.i.d normal, which has a closed form expression for the posterior distribution of revenue. This is investigated in section 4.3. A consequence of the inequality is that parameter uncertainty adjusted VaR estimates have lower ex-post probability of exceedences (α). Section 4.4 shows that existing back-tests based on exceedences probability will, if sufficiently powerful, always reject VaR models that are properly specified and incorporate parameter uncertainty.

4.1 General Estimator of Bayesian VaR

Let $\mathcal{M} = \{\Theta, X_t, (r_t|\Theta, X_t), (X_{t+1}|\Theta, X_{t+1})\}$ denote a Value-at-Risk model where parameters Θ and state variables X_t are unknown. The model fully specifies the likelihood $p(r_{t+1}|\Theta, X_{t+1})$ and tradition distribution $p(X_{t+1}|\Theta, X_{t+1})$. Let the dataset be denoted $Y_t = \{r_\tau\}_{\tau=1}^{t}$. The VaR at $t+1$ is the α quantile over the predictive distribution of $p(r_{t+1}|Y_t)$,

$$VaR_{t+1|t} = \{x : \int_{-\infty}^{x} p(r_{t+1}|Y_t) dr_{t+1} = \alpha\}.$$

Characterizing $p(r_{t+1}|Y_t)$ is the problem at hand. The Bayesian VaR estimator integrates over parameter uncertainty summarized in $p(\Theta, X_{t+1}|Y_t)$. By application of Bayes theorem, we have

$$p(r_{t+1}|Y_t) = \int_{\Theta, X_{t+1}} p(r_{t+1}|\Theta, X_{t+1}) p(\Theta, X_{t+1}|Y_t) d\Theta dX_{t+1} \tag{1}$$

The first term is the likelihood for r_{t+1}. The second term $p(\Theta, X_{t+1}|Y_t)$ is not specified by the model and is complex. This contains the predictive density for the unobserved state variable and the posterior for Θ. By Bayes theorem, it equals

$$p(\Theta, X_{t+1}|Y_t) = \int_{X_t} p(X_{t+1}|\Theta, X_t) p(X_t|\Theta, Y_t) p(\Theta|Y_t) dX_t.$$

The first term $p(X_{t+1}|\Theta, X_t)$ the transition density for the state variable and is given by the model. The second is the filtering distribution for the state variable $p(X_t|\Theta, Y_t)$. The third is the smoothed posterior forΘ. Both posteriors do not have analytical expressions. Fortunately, MCMC allows samples to be drawn from them and used to construct a Monte Carlo estimator of the Bayesian VaR.

4.2 MCMC Estimator of Bayesian VaR

A difficulty with Bayesian VaR is that the the predictive distribution $p(r_{t+1}|Y_t)$ is analytically intractable for all but the simplest VaR models. We present here a Monte Carlo method that can use used to samples from $p(r_{t+1}|Y_t)$. The method uses Markov Chain Monte Carlo to draw a random sample from $p(X_t, \Theta|Y_t)$. Once this is performed, Monte Carlo simulation is used to obtain the distribution of $p(X_{t+1}, \Theta|Y_t)$. The method is as follows:

1. Use MCMC to draw N samples $(X_t^{(n)}, \Theta^{(n)})_{n=1}^N$ from $p(X_t, \Theta|Y_t)$,
2. For n from 1 to N
   (a) draw $X_{t+1}^{(n)} \sim p(X_{t+1}|\Theta^{(n)}, X_t^{(n)})$
   (b) draw $r_{t+1}^{(n)} \sim p(r_{t+1}|\Theta^{(n)}, X_{t+1}^{(n)})$
3. Calculate the empirical distribution function, $\hat{\Phi}(x) = \frac{\#(\{r_{t+1}^{(n)}\}_{n=1}^N < x)}{N}$.
4. Set $V\tilde{a}R_{t+1}(\alpha|Y_t) = \hat{\Phi}^{-1}(\alpha)$.

The distributions $p(X_{t+1}|\Theta, X_t)$ and $p(r_{t+1}|\Theta^{(n)}, X_{t+1})$ are specified by the model and in most cases (all diffusion processes, GARCH), both are normal.

4.3 Three Inequality Theorems

It can be proven that for a given VaR model, the Bayesian VaR estimate is uniformly lower than than VaRs obtained by plugging-in point-estimates $\hat{\Theta}$ and $\hat{X}_t$,

$$VaR_{t+1}(\alpha|Y_t) \leq VaR_{t+1}(\alpha, \hat{\Theta}, \hat{X}_{t+1}).$$

We first prove this in the case of $i.i.d$ normal returns. We then generalize the proof for VaR models where r_t conditioned on Θ and X_{t+1} is normal. All diffusion and GARCH-type models for r_t satisfy this.

Theorem 1 (i.i.d normal) Let $r_t \sim$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$ for all t, with unknown μ and σ^2, and let $Y_t = \{r_\tau\}_{\tau=1}^t$. The Value-at-Risk obtained by point estimates $\hat{\mu} = E(\mu|Y_t)$, $\hat{\sigma}^2 = E(\sigma^2|Y_t)$ underestimates the Bayesian Value-at-Risk.

Proof Let $p(r_t \leq x|\mu, \sigma^2) := \Phi(x, \mu, \sigma^2)$, the normal cumulative distribution function. The point estimate VaR at level α is found by solving $\Phi(x, \hat{\mu}, \hat{\sigma}^2) = \alpha$. The true VaR depends on the predictive distribution $p(r_{t+1}|Y_t)$. By Bayes theorem,

$$\begin{aligned} p(r_{t+1}|Y_t) &= \int_{\sigma^2,\mu} p(r_{t+1}, \mu, \sigma^2|Y_t) d\mu d\sigma^2 \\ &= \int_{\sigma^2,\mu} p(r_{t+1}|\mu, \sigma^2, Y_t) p(\mu, \sigma^2|Y_t) d\mu d\sigma^2 \\ &= \int_{\sigma^2,\mu} p(r_{t+1}|\mu, \sigma^2) p(\mu|Y_t) p(\sigma^2|Y_t) d\mu d\sigma^2, \end{aligned}$$

where the last line uses the fact that $r_{t+1}|\mu, \sigma^2$ is conditionally independent of Y_t and assume that $\mu|Y_t$ is independent of $\sigma^2|Y_t$. The posterior distribution for r_{t+1} is a mixture of the normal likelihood $p(r_{t+1}|\mu, \sigma^2)$ and the posteriors for μ and σ^2.

Now consider the distribution function for $(r_{t+1}|Y_t)$, $p(r_{t+1} \leq x|Y_t) := \Phi(x|Y_t)$. This is given by

$$\begin{aligned}
\Phi(x|Y_t) &= \int_{-\infty}^{x} \int_{\sigma^2,\mu} p(r_{t+1}|\mu,\sigma^2)p(\mu|Y_t)p(\sigma^2|Y_t)d\mu d\sigma^2 dr_{t+1} \\
&= \int_{\sigma^2,\mu} \Phi(x,\mu,\sigma^2)p(\mu|Y_t)p(\sigma^2|Y_t)d\mu d\sigma^2.
\end{aligned}$$

$\Phi(x,\mu,\sigma^2)$ is convex in both μ and σ^2 for all $x \leq \mu$. Proof of convexity is omitted; it may be confirmed through lengthy differentiation. Jensen's inequality states $E[g(X)] \geq g(E[X])$ for convex g. The bivariate version states $E[g(X,Y)] \geq g(E[X],E[Y])$. The value x corresponds to a lower tail value and hence $x < \mu$. Invoking Jensen's inequality yields

$$\begin{aligned}
\Phi(x|Y_t) &\geq \Phi\left(x, \int \mu p(\mu|Y_t)d\mu, \int \sigma^2 p(\sigma^2|Y)d\sigma^2\right) \\
&= \Phi(x,\hat{\mu},\hat{\sigma}^2).
\end{aligned}$$

$\Phi(x,\mu,\sigma^2)$ increases monotonically in x. It follows from the above inequality that

$$\{x: \ \Phi(x|Y_t) = \alpha\} \leq \{x: \ \Phi(x,\hat{\mu},\hat{\sigma}^2) = \alpha\},$$

for all $\alpha < \frac{1}{2}$. Equivalently, $VaR_{t+1}(\alpha, Y_t) \leq VaR_{t+1}(\alpha,\hat{\mu},\hat{\sigma}^2)$.

$\square$

Theorem 2 (Generalized Inequality) The Bayesian Value-at-Risk estimate for model $(\mathcal{M},\Theta,X_t)$ where Θ and X_t are unknown is always lower than the estimate obtained by plugging-in estimates $\hat{\Theta} = E(\Theta|Y_t)$ and $\hat{X} = E(X_t|Y_t)$, or

$$VaR_{t+1}(\alpha|Y_t) \leq VaR(\alpha,\hat{\Theta},\hat{X}_t).$$

Proof See section A.4 of the appendix.

Theorem 3 (Certainty Inequality) The Bayesian VaR when facing parameter uncertainty in Θ and X_t is lower than then the Value-at-Risk when (μ,σ) are known.

Proof Under certainty, $E(\mu|Y_t) = \mu$, $E(\sigma|Y_t) = \sigma$ and using the previous theorem, we have

$$\Phi^{-1}(x|Y_t) \geq \Phi^{-1}(x,\mu,\sigma), \ \Leftrightarrow \ VaR_{t+1}(\alpha,Y_t) \leq VaR_{t+1}(\alpha,\mu,\sigma).$$

$\square$

Theorem 3 has important consequences for misspecification tests of Value-at-Risk models. Bayesian VaR have lower coverage probabilities than α since they hedge an additional risk factor, estimation risk, which is ignored by back-testing methods.

4.4 Bayesian VaR under i.i.d. Normal returns

Theorems 1 and 2 do not measure how *much* the plug-in VaR understate the Bayesian VaR. The extent depends on the posterior distribution of $\Theta|Y_T$, and for nearly all models, this posterior has no closed form expression and require MCMC simulation. An important exception is where returns are drawn independently and identically from the normal distribution, or $r_t \sim i.i.d\ \mathcal{N}(\mu, \sigma^2)$. The posterior $p(\mu, \sigma^2|Y_T)$ is known in closed form when the Jeffreys prior for μ and σ^2 are used:

$$p(\mu, \log\sigma) \propto 1.$$

When μ is unknown and σ known, the posterior distribution for the mean is normal:

$$p(\mu|\sigma, Y_T) = \mathcal{N}\left(\hat{\mu}, \frac{\sigma^2}{T}\right)$$

where $\hat{\mu}$ is the sample mean, $\overline{r}$. This implies the following posterior distribution for r_{T+1} :

$$p(r_{T+1}|\sigma, Y_T) = \int_\mu p(r_{T+1}|\mu, \sigma)p(\mu|\sigma, Y_T) = \mathcal{N}\left(\hat{\mu}, \sigma^2 + \frac{\sigma^2}{T}\right).$$

If the mean is known and the variance is unknown, the posterior for σ has an inverse gamma distribution proportional to:

$$p(\sigma|\mu, Y_T) \propto \frac{1}{\sigma^{T+1}}\exp\left\{-\frac{(N-1)\hat{\sigma}^2}{2\sigma^2}\right\},$$

where $\hat{\sigma}^2$ is the unbiased sample variance estimator. This implies the following posterior return distribution:

$$p(r_{T+1}|\mu, Y_T) = \int p(r_{T+1}|\mu, \sigma)p(\sigma|\mu, Y_T)d\sigma = \mathcal{T}(\mu, \hat{\sigma}, T-1).$$

where $\mathcal{T}(m,s^2,v)$ denotes the density of the non-central Student-t distribution with mean m, variance s^2, and v degrees of freedom. This is defined by

$$\mathcal{T}(m, s^2, v) = \mu + s \times \sqrt{\frac{v-2}{v}}\mathcal{T}(v)$$

where $\mathcal{T}(v)$ is a central Student-t with v degrees of freedom.

If both the mean and variance are unknown, the posterior $p(\mu, \sigma|Y_T)$ is proportional to:

$$p(\mu, \sigma|Y_T) \propto \frac{1}{\sigma^{N+1}}\exp\left\{\frac{-N(\mu-\hat{\mu})^2}{2\sigma^2} - \frac{(N-1)\hat{\sigma}^2}{2\sigma^2}\right\},$$

and implies the posterior return distribution:

$$p(r_{T+1}|Y_T) = \int_\sigma \int_\mu p(r_{T+1}|\mu, \sigma)p(\mu, \sigma|Y_T)d\mu d\sigma = \mathcal{T}\left(\hat{\mu}, \hat{\sigma}^2 + \frac{\hat{\sigma}^2}{T}, T-1\right).$$

Comparing these to the plug-in predictive distribution distribution, $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$, nicely illustrates the distributional effect of parameter uncertainty. When μ is uncertain, the variance of returns increases from σ^2 to $\sigma^2 + \frac{\sigma^2}{T}$. When σ is uncertain, the return variance is unchanged and the kurtosis of the distribution increases from 3 to $3 + \frac{6}{T-5}$. When both the mean and variance is unknown, both the variance and kurtosis increase.

The Value-at-Risk is also known in closed form. The plug-in estimator using $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$ sets

VaR as:

$$VaR(\alpha|\hat{\mu}, \hat{\sigma}) = \hat{\mu} + \hat{\sigma} Z_\alpha,$$

where Z_α is the α^{th} quantile of the standard normal distribution; at the 1% quantile this equals -2.33. For the unknown μ and known σ case, the Bayesian VaR is:

$$VaR(\alpha|\sigma, Y_T) = \hat{\mu} + \left(\hat{\sigma}^2 + \frac{\hat{\sigma}^2}{T}\right)^{1/2} Z_\alpha.$$

For the known μ and unknown σ case, the Bayesian VaR is:

$$VaR(\alpha|\mu, Y_T) = \mu + \sigma t_\alpha(T-1),$$

where $t_\alpha(T-1)$ is the α^{th} quantile of the central Student-t distribution; at 1% with 49 degrees of freedom this equals -2.40. In the last case where both μ and σ are unknown, the Bayesian VaR is:

$$VaR(\alpha|\mu, Y_T) = \hat{\mu} + \left(\hat{\sigma}^2 + \frac{\hat{\sigma}^2}{T}\right)^{1/2} t_\alpha(T-1).$$

To illustrate the difference, suppose $T = 10$ observations are used estimate the mean and variance,$\alpha = 0.01$ and $\hat{\mu} \approx \mu = 0$, $\hat{\sigma} \approx \sigma = 1$. The corresponding VaR estimates are:

$$\begin{aligned} \text{Plug-in} &: \quad -2.326 \\ \text{Bayesian, } \mu \text{ unknown} &: \quad -2.439 \\ \text{Bayesian, } \sigma \text{ unknown} &: \quad -2.821 \\ \text{Bayesian, } \mu, \sigma \quad \text{unknown} &: \quad -2.959 \end{aligned}$$

The plug-in point-estimates VaR underestimates the Bayesian VaR by 4.85% with unknown mean, 21.2% when the variance is unknown and 27.2% when both are unknown. The assumption of $T = 10$ observations is not unreasonable: when both the mean and variance are unobserved and stochastic state-variables, the effective number of data points used in estimate a state (μ_t, σ_t^2) is only a fraction of the total sample size, and depends on the persistence of each variable.

5 Misspecification Tests under Parameter Uncertainty

Misspecification tests of Value-at-Risk time-series are useful tools assessing the quality of estimates. Many tests have been proposed, such as the Kupiec (1995) unconditional coverage test, the Christoffersen (1998) Markov chain test and the Engle and Mangelli (1999) CAViaR test. Regardless of construction and test statistic, the each compares realized or *ex-post* trading revenue to *ex-ante* VaR estimates. We argue this introduces a hindsight bias effect when the decision maker faces parameter uncertainty, which is always the case.

Theorem 3 states that the Bayesian VaR with uncertainty is less than the VaR with certainty when using the same model. The Bayesian VaR is optimal optimal respect to the posterior revenue distribution but suboptimal with respect to the true distribution, when parameters are known with certainty. A consequence is that well specified Bayesian VaR models (and more generally, any parameter-uncertainty adjusted VaR), have lower coverage probability than α: $p[r_t < VaR_t(\alpha)] < \alpha$. Misspecification tests based on the property should always reject these well-specified models. A proof for this result is given in section 5.1, and we briefly overview

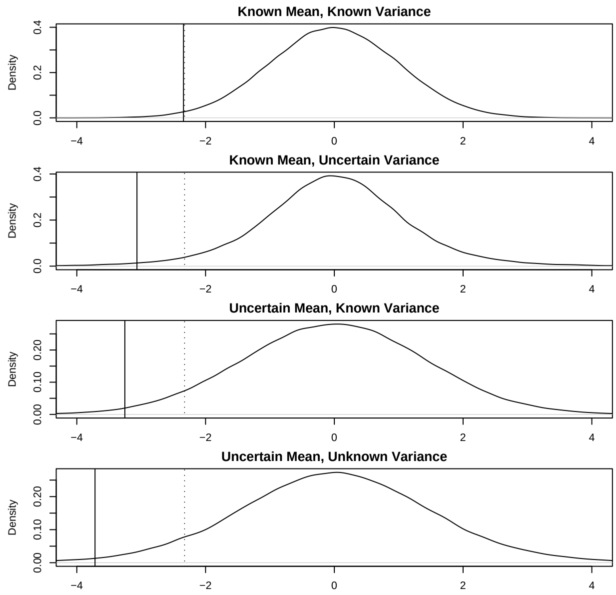


Figure 2: Effect of mean and variance uncertainty on Value-at-Risk. The solid line is the 1% VaR incorporating uncertainty in μ and σ^2 and broken line is the 1% VaR using obtained by plugging-in point-estimates $\hat{\mu}$ and σ^2. Returns are $i.i.d \sim \mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are either known ($\mu = 0,\ \sigma^2 = 1$) or unknown. Distributions calculated assuming of $T = 10$ effective observations and diffuse conjugate priors.

the definition for well-specified VaR estimates, and the Kupiec (1995) unconditional coverage test.

A VaR model is well specified if and only if probability of exceeding the VaR equals the level for each period of time, conditional on all available information. The model may be misspecified either because: (1) the average probability of exceedences does not equal the VaR level, or (2), the average probability equals the level, but the instantaneous probability significantly deviates from the level, or (3) both. Formally, a model is well specified if:

Definition 2 *A model $V\hat{a}R^{\alpha}_{t+1|t}$ is well specified if and only if*

$$E[X_{t+1,\alpha}|\mathcal{I}_t] = \alpha \ \ almost \ \ surely, \ \ for \ \ all \ \ t \tag{2}$$

where $X_{t+1} = I(r_{t+1} < V\hat{a}R^{\alpha}_{t+1|t})$, the indicator of exceedences times.

Direct tests of (2) are impossible. The information set $\mathcal{I}_t$ pertaining to the bank is unobservable to regulatory authorities and econometricians. Instead, existing back-testing procedures are based on testing implications of (2) rather than the the condition itself.

The two testable consequences are, (1) the unconditional probability of exceeding the VaR at any given moment is $p(r_{t+1} < V\hat{a}R_{t+1}(\alpha)) = \alpha$, which follows from the law of iterated expectations; and (2) the sequence of hits $\{X_1, X_2, ...X_t\}$ must be independent from each other; or equivalently the sequence must not convey any information about of X_{t+1}.

The unconditional test of Kupiec (1995) tests whether (1) holds,

$$H_0: \ \ E(X_{t+1}) = \alpha \ \text{ for all } t.$$

or, whether the reported VaR is violated more (or less) than $\alpha \times 100\%$ of the time. Kupiec proposed a likelihood ratio test for H_0 with test statistic

$$\begin{aligned} T_\alpha &= 2\log\left(\left(\frac{1-\hat{\alpha}}{1-\alpha}\right)^{T-I(\alpha)}\left(\frac{\hat{\alpha}}{\alpha}\right)^{I(\alpha)}\right) \\ I(\alpha) &= \sum_{t=1}^{T} I(r_t < VaR_t(\alpha)),\ \ \hat{\alpha} = \frac{1}{T}I(\alpha) \end{aligned}$$

Under H_0, $T_\alpha \sim \chi^2_1$. This test is implicitly used in the Basel Amendment capital charge formula to determine the multiplier term.

5.1 Coverage Probability of the Bayesian VaR

Theorem 3 proved that the the optimal Bayesian VaR for a decision maker facing parameter uncertainty is lower than the equivalent Value-at-Risk with complete certainty:

$$VaR_{t+1}(\alpha, Y_t) \le VaR_{t+1}(\alpha|\Theta).$$

A consequence is that the ex-post probability of exceeding $VaR_{t+1}(\alpha, Y_t)$ is less than α.

Theorem 4 (Coverage Probability) Suppose returns r_t follow a model $(\mathcal{M}, \Theta, X_t)$ but Θ and X_t are unknown. The Bayesian $VaR_{t+1}(\alpha|Y_t)$ that has a coverage probability less than α, or

$$p(r_{t+1} < VaR_{t+1}(\alpha|Y_t)) \le \alpha.$$

Proof Suppose Θ, X_{t+1} are known with certainty. Then the *true* VaR given $(\mathcal{M}, \Theta, X_t)$ is

$$VaR_{t+1}(\alpha|\Theta, X_{t+1}) = \{x : \int_x^\infty p(r_{t+1}|\Theta, X_{t+1})dr_{t+1} = \alpha\}$$

By construction, $p(r_{t+1} < VaR_{t+1}(\alpha|\Theta, X_{t+1})) = \alpha$.From theorem 3, the optimal VaR under uncertainty is less than the VaR with certainty,

$$VaR_{t+1}(\alpha|Y_t) \le VaR_{t+1}(\alpha|\Theta, X_{t+1}).$$

and since $p(r_{t+1} < x)$ is monotonically increasing in x,

$$p(r_t \le VaR_{t+1}(\alpha|Y_t)) \le p(r_t \le VaR_{t+1}(\alpha|\Theta)) = \alpha$$

$\square$

This poses a problem for unconditional misspecification tests. It is optimal for decision makers to estimate VaR conservatively under uncertainty, and this level corresponds to a significance level α^* less than the actual (or mandated) level α. A sufficiently powerful coverage test will always reject these well-specified estimates.

6 Trading Revenue and VaR Data

A large data set has of daily trading revenue and 1% VaR for five major commercial banks has been obtained, courtesy of Pérignon and Smith (2006). The series span 2001 to 2004 for four banks, and consist of approximately 1000 observations; the fifth spans 2002 to 2005 with 775 observations. The banks selected are among the largest commercial banks in the US (Bank of America), Switzerland (Credit Suisse First Boston), Germany (Deutsche Bank), Canada (Royal Bank of Canada) and France (Société Générale). Figure (3) show the data series.

In sampling banks, the procedure started with the largest bank in each country. If the bank did not disclose revenue and VaR, the second and third largest banks were considered. Under this procedure, a sample of five banks was obtained. For the US, Germany and Canada, the largest commercial banks were selected: Bank of America (BoA), Deutsche Bank (DEU) and Royal Bank of Canada (RBC) respectively. For Switzerland and France, the 2nd and 3rd largest banks were sampled: Credit Suisse First Boston (CSFB) and Société Générale (SG) respectively.

The definition of trading revenues varies slightly across the sample. Royal Bank of Canada and Deutsche Bank report hypothetical revenue based on the previous day portfolio allocation. The Bank of America, Credit Suisse First Boston and Société Générale report actual revenue affected by intra-day changes to the portfolio allocation. The banks do not disclose whether the trading revenues include trading fees and commissions.

The numerical values of VaR and trading revenue for these banks is not publicly disclosed. Instead, the series were obtained by extracting data from VaR and revenue graphs included in annual reports. See the appendix of Pérignon and Smith (2006) for details of the extraction procedure. The nature of the data-source and the extraction procedure both introduce measurement error into the data-set. Pérignon and Smith perform a controlled experiment to measure the extraction error using simulated trading revenue and VaR. The extraction error ratio, defined as mean absolute error divided by mean absolute return, is found to be small at 0.010% for bar and line plus marker graphs.

7 Analysis of the Banks' Reported VaR

We investigate whether the each bank over-reports their VaR level when ignoring parameter uncertainty. Table (1) contains the number of days where the 1% VaR was exceeded. Each bank has fewer exceedences than the expected 10 days over the sample – DEU, RBC and SG have no exceedances days. This suggests, prima facie, that each bank over-reported its Value-at-Risk.

To test the hypothesis of over-statement, each banks' VaR were calculated using four "naive" models (historical simulation, filtered HS, GARCH and IGARCH). These models are naive in the sense that VaR is estimated only form past revenue data. The banks' VaR is estimated conditional on their information set. This set includes the composition of trading portfolio and the riskiness of each component. Without knowledge of this set, it is impossible to reject a particular VaR at time t as "too conservative." There always exists a possible conditioning information set that justifies a given VaR level. For instance, the bank might have issued out-of-the-money puts that act as catastrophe insurance. Losses on these contracts will be realized very infrequently, but their addition will increase the VaR over the sample period. Over very a time-series of bank revenue, there may be enough realized losses to justify the bank's *average* VaR. However, the few empirical studies of banks' VaR have all been very short, consisting of no more than a thousand observations for each bank.

The problem is avoided in this paper by not directly comparing VaR levels. Instead, we compare the *economic* loss that a bank faces from over-reporting their VaR: the capital charge. This is the mandated level of capital reserves a bank must set aside It is in the interest of

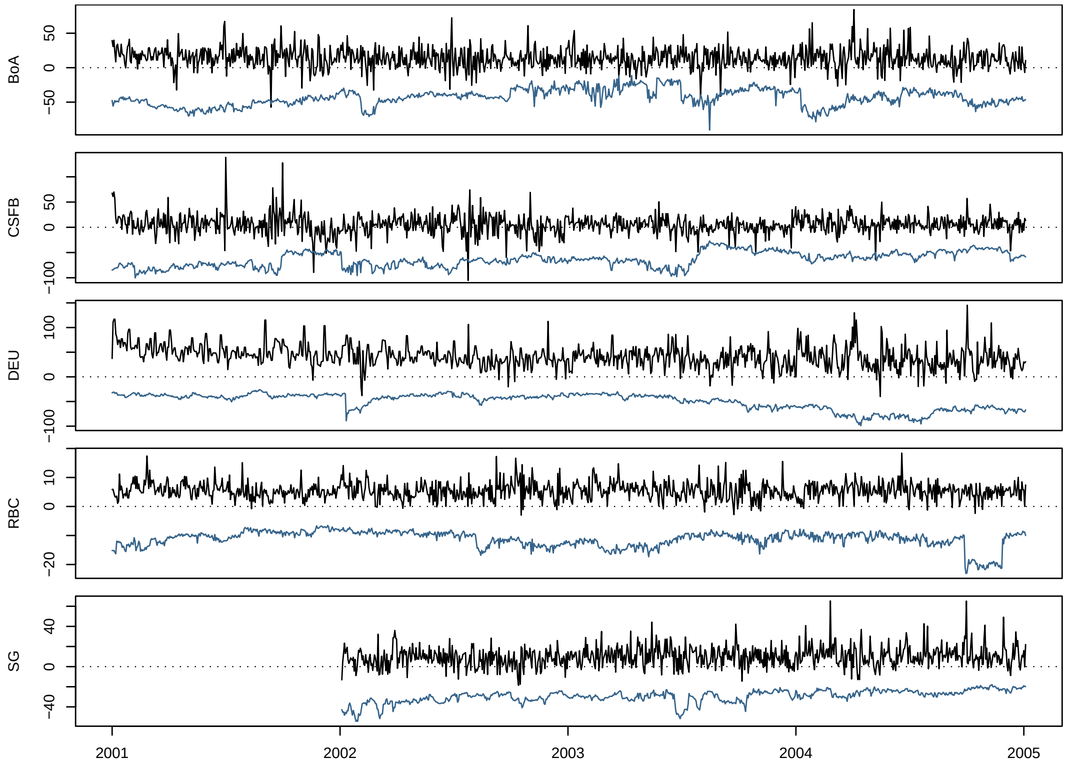


Figure 3: The five banks daily trading revenue and reported 1% Value-at-Risk from 2001 to 2004. The banks are Bank of America (BoA), Credit Suisse First Boston (CSFB), Deutsche Bank (DEU), Royal Bank of Canada (RBC) and Société Générale (SG). Only BoA and CSFB exceed the Value-at-Risk (4 and 6 instances). The expected number at 1% is 10 exceedances. Data were kindly provided by Smith and Pérignon (2006).

banks to minimize this amount since excess set-aside capital incurs an investment opportunity cost. A $\$X$ reduction in capital charges has an value to the bank of at least $\$r_f X$ per annum, where r_f is the annual risk-free interest rate.

The framework for bank capital requirements was specified by the 1996 Basel Amendment to Market Risk and is adhered to by each country in our sample (United States, Switzerland, Germany, France and Canada). Under this framework, the formula for setting capital charges is:

$$
\begin{aligned}
CC_t &= \max\left(VaR_t(0.01), M_t \times \frac{1}{60}\sum_{t=0}^{59} VaR_{t-i}(0.01)\right), \qquad (3)\\
M_t &= \begin{cases} 3.0 & \text{if} \quad N \leq 4 \quad \text{green} \\ 3 + 0.2(N-4) & \text{if} \quad 5 \leq N \leq 9 \quad \text{yellow.} \\ 4.0 & \text{if} \quad 10 < N \quad \text{red} \end{cases}
\end{aligned}
$$

where N is the number of exceedences, $r_{t+1} < VaR_{t+1}(0.01)$, over the last 250 days. The charge formula is incentive compatible with truthful VaR reporting (Cuoco and Liu, 2006). Banks that overstate their VaR will have fewer than expected exceedences and a$M_t = 3$, but have a high average VaR. Banks that understate their VaR will have too many exceedance days and be penalized by a multiplier $M_t > 3$. The equation (3) is used to calculate the hypothetical capital charges for each bank, using the reported VaR time-series, and the estimated VaR time-series for each naive model.

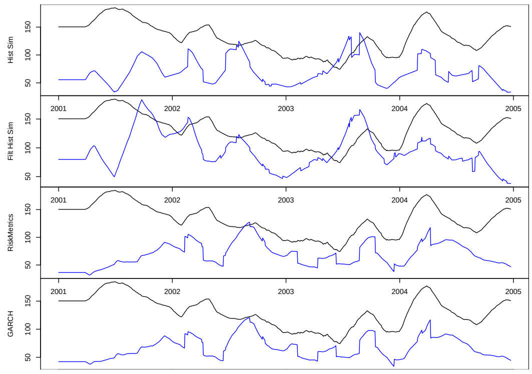
The four models considered are historical simulation (HS), filtered HS, GARCH(1,1) and IGARCH(1). The filtered HS estimator uses the IGARCH model to estimate the volatility path. Point-estimates for the parameters in the IGARCH and GARCH models are not estimated using the data set, and are instead fitted using daily S&P 500 index returns, 1986-2006. This is done to yield a real out-of-sample capital charge test; the charges are lower when the models are fitted in sample. The model fits from S&P 500 returns are:

$$
\begin{aligned}
\text{GARCH:} \quad h_t^2 &= 8.81\text{e-}7 + 6.14\text{e-}02(r_t - \overline{r}_t)^2 + 0.931 h_{t-1}^2 \\
\text{IGARCH}: \quad r_t &= 0.06(r_t - \overline{r}_t)^2 + 0.94 h_{t-1}^2.
\end{aligned}
$$

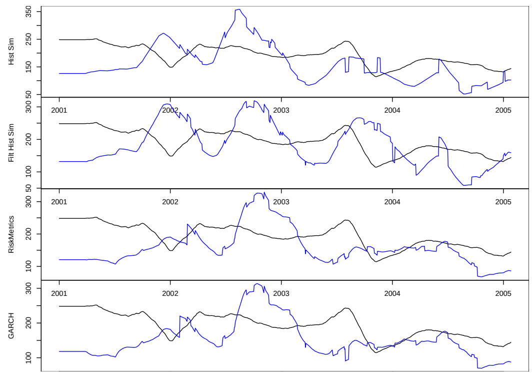
There is little difference between the two models: the GARCH autoregressive parameter on h_{t-1}^2 is almost equal to that of the IGARCH. Average $\overline{r}_t$ was set as a sliding window over the last 100 past daily revenues. The historical simulation and filtered historical simulation use a window of 50 days.

7.1 Results

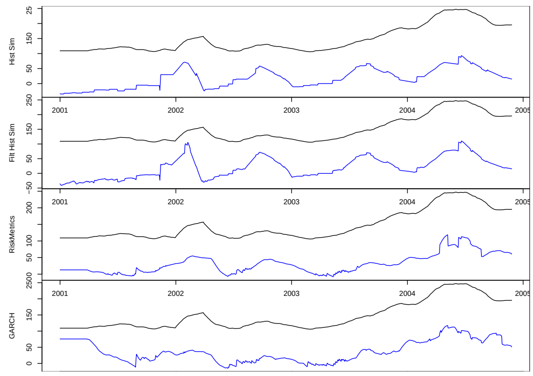
Figure (4) shows the time-series of instantaneous capital charges and table (2) contains the average capital charges. For all four naive models and all five banks, in all 20 cases, the average capital charges under the naive models are *lower* than those from the banks' VaR. This unanimous result is surprising. Each naive model in each case produces VaR that cut capital charges by a large margin. Conversely, judged by these naive models, each bank over-allocates capital according to the Basel formula by a large margin. In particular, DEU, RBC and SG over-allocate capital by factors 4.4, 12 and 2.2 with their reported VaR than necessary. In conclusion, each bank over-reports its VaR when measured by capital charges paid and when parameter uncertainty is ignored.



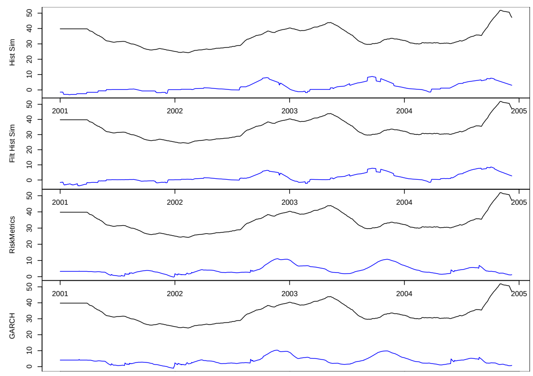
Bank of America



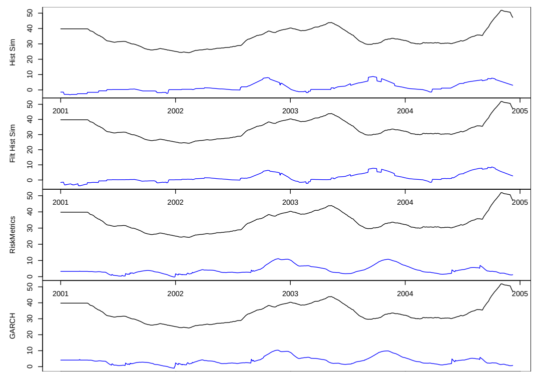
Credit Suisse First Boston



Deutsche Bank



Royal Bank of Canada



Société Générale

Figure 4: The capital charges of each bank and capital charges under each alternative VaR model (historical simulation, filtered historical simulation , IGARCH(1)/Risk-metrics and GARCH(1,1)). Each model is estimated out-of-sample (IGARCH and GARCH use S&P 500 index returns). *Black:* Actual capital charge from reported VaR, *Blue:* Capital charge from alternative model.

	BoA	CSFB	DEU	RBC	SG	Expected X_t
Full Sample	4	6	0	0	0	10
2001	1	2	0	0	NA	2.5
2002	0	1	0	0	0	2.5
2003	3	2	0	0	0	2.5
2004	0	1	0	0	0	2.5

Table 1: Number of exceedence days, $r_t < VaR_t$, for each bank. The expected number is given by $\alpha \times N_{days}$. Each bank has considerably lower exceedences than expected, particularly DEU, RBC and SG which have none over the four year sample.

Models	BoA	CSFB	DEU	RBC	SG
Bank's Internal	131	193	146	33.9	90.8
Hist. Sim	69.5	157	17.8	1.79	40.3
Filtered HS	94.3	182	19.6	1.64	37.6
IGARCH(1)	68.7	158	31.4	4.12	48.0
GARCH(1,1)	66.2	152	37.7	3.71	46.1
Overcharge %	75.2	19.6	440	1104	112

Table 2: Averaged Capital Charges for the banks internal VaR model and the four alternative models. Each bank incurs higher capital charges using their internal model than each alternative model. Alternative models are all fit out-of-sample using the sliding prediction window method. Overcharge % is calculated by dividing the banks internal charge with the average charge across the alternative models. IGARCH and GARCH are fitted using the S&P500 index returns, 1926 to 2007.

8 Bayesian VaR and the Capital Charge Puzzle

This section investigates whether the Capital Charge puzzle can be explained banks incorporating estimation risk into their VaR estimates. Section 7.1 found that each banks' reported VaR seems overstated by a very large margin when parameter uncertainty is ignored. Using the estimator derived in section 4, the Bayesian VaR each bank is estimated. Estimation is performed out-of-sample using a sliding window of $T = 200$ observations of trading revenue. A sequential estimation method is employed to estimate the day-ahead VaR.

8.1 Sequential Estimation Method

Estimating a series of Bayesian VaRs for $t = 1,...,T$ requires sequential estimation using data $Y_1, Y_2..., Y_T$, where $Y_t = \{r_\tau\}_{\tau=1}^t$. Suppose we have at time t obtained $VaR_{t+1}(\alpha|Y_t)$ from calculating the necessary posteriors $p(X_t|\Theta, Y_t)$ and $p(\Theta|Y_t)$. At time $t+1$, new data is observed $(Y_t \to Y_{t+1})$. To calculate the $t+1$ VaR estimate, it is necessary to calculate the posteriors $p(X_{t+1}|\Theta, Y_{t+1})$ and $p(\Theta|Y_{t+1})$. Ideally, we would like to update the posteriors $p(X_t|\Theta, Y_t) \to p(X_{t+1}|\Theta, Y_{t+1})$ and $p(\Theta|Y_t) \to p(\Theta|Y_{t+1})$ without re-performing an entire new MCMC run on Y_{t+1}, which is inefficient. Methods to do this exist, but are complicated to implement. Gordon, Salmon and Smith (1993) proposed a "particle filter" method that yields an approximation to $p(X_{t+1}|\Theta, Y_{t+1})$ without re-running MCMC. The method simulates of X_t through the transition density and uses unequal sampling to approximate the new posterior. Polson, Stroud and Muller (2002) haave developed a "practical filtering" method that also approximates the update step by re-running MCMC only on a block of most recent states $(X_{t-k}, ..., X_t)$.

We use the brute force method to perform updating. After each new observation, the

MCMC algorithm is re-run using the new data-set. This method is slow – the algorithm needs to be run N times for N observation, but is computationally feasible a modest laptop using WinBUGS.

A sliding observation window is used. Instead of using the entire series $Y_t = \{r_1, ..., r_t\}$, a window $Y_{t,w} = \{r_{t-w+1}, ..., r_t\}$ of size w is data used to calculate posteriors. This is done to avoid a Bayesian learning effect, where the posterior distributions for parameters thin as t increases. Instead, we wish to quantify the parameter uncertainty effect with a given length of data. This also has the important benefit of reducing the time to run each MCMC algorithm step. A window of trading days 200 days is used . This window slightly under to the 1-year (≈250 days) minimum window the Basel framework requires banks to back-test their portfolios.

8.2 Model and Priors

We considered the IGARCH(1) model. This model was picked since it is the simplest of the parametric models considered, has fewest parameters (2) of any stochastic volatility model, and is fastest to estimate by MCMC. The model is formally specified by

$$\begin{aligned} r_{t+1}|\mu, h_{t+1} &\sim \mathcal{N}(\mu, h_{t+1}) \\ h_{t+1} &= \lambda h_t + (1-\lambda)(r_t - \mu)^2 + \varepsilon_t \\ \varepsilon_t &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

A small Gaussian error term is included to smooth the transition density for h_{t+1}, which is necessary to sample from $p(r_{t+1}|r_{t+1}, \lambda)$.

Choice of priors significantly affects small sample Bayesian inference. The goal is to see if parameter uncertainty can explain the capital charge puzzle without having to assume extreme ignorance about the parameters. Consequently, we try to pick as realistic priors as possible, which are intentionally informative. Each bank,k, is assumed know their average return μ_k with precision at least $1/\sigma^2$, where σ^2 is the return variance, and the prior for μ_k is set as $\mu_k \sim \mathcal{N}(\overline{r}_k, S_k^2)$, where S_k^2 is the sample variance. It is also assumed that the banks have previously observed that volatility is persistent and priorly believe that the volatility autoregressive term, λ, is close to 1. The prior used is $\lambda \sim Beta(20, 1.5)$. The error variance has a thin prior inverse gamma distribution, $\sigma^2 \sim \mathcal{IG}(20, 20)$, with shape and location parameter of 20.

8.3 Results

The Bayesian VaR estimates for the 200 day inference window are plotted in figure (5). The average VaR for each bank is given in table (3) and the average capital charges are given in table (4). Figure (6) shows the instantaneous capital charges.

The Bayesian VaRs seen in figure (5) are consistently lower than the plug-in VaR estimates. They are also higher, on average, than the banks' reported VaR. The banks still seem to overstate the magnitude of their VaR after taking parameter uncertainty into account. For Bank of America and Credit Suisse First Boston, the difference is reasonably small: their reported VaR is 39% and 23% higher than the Bayesian VaR, and this difference may not be significant. Société Générale overstates by 52%, Deutsche Bank by 88% and Royal Bank of Canada by 198%. These three banks continue to overstate by a considerable margin. The Bayesian VaR respond downward more aggressively the point-estimate VaR after large negative

Models	BoA	CSFB	DEU	RBC	SG
Bank's Internal	-43.4	-63.5	-49.8	-11.3	-29.6
Plug-in IGARCH(1)	-19.9	-37.7	-9.60	-1.28	-12.8
Bayesian IGARCH(1)	-31.1	-51.5	-26.4	-3.78	-19.5
VaR Overstate %, plug-in	118	68.4	419	783	131
Var Overstate %, Bayes	39.5	23.3	88.6	198	52.0

Table 3: Average Value-at-Risk for the banks internal VaR model and the IGARCH(1) model using plug-in estimates (ignoring parameter uncertainty) and the Bayes IGARCH(1) using a 200 days training window. VaR units are millions, local currency.

Models	BoA	CSFB	DEU	RBC	SG
Bank's Internal	131	193	146	33.9	90.8
Plug-in IGARCH(1)	67.5	158	31.4	4.12	48.0
Bayes IGARCH(1)	94.2	163	78	7.12	58.9
Overcharge %, Point Est.	75.2	19.6	464	722	189
Overcharge %, 200 day	41.7	17.0	325	376	154

Table 4: Average capital charges for the banks internal VaR model and the IGARCH(1) model using point-estimates (ignoring parameter uncertainty) and the Bayes IGARCH(1) using 200 days training window. The Bayes IGARCH incur higher capital charges than the point-estimate IGARCH. Even after adjusting parameter uncertainty, each bank incurs higher capital charges under their internal models.

returns. Also, Bayesian VaR are noisier than both the point-estimate VaR and banks' VaR. This is likely due to Monte Carlo simulation error involved in estimation.

The capital charges for each bank under the Bayesian VaR were calculated. Figure (6) show the instantaneous capital charges under the Basel 1996 formula. The capital charges for the 200-day Bayesian VaR remain lower than the capital charges calculated by the banks' VaR; all but CSFB have considerably lower capital charges under Bayesian VaR than internal VaR. Incorporating for parameter uncertainty reduces the percentage amount overpaid under the internal VaR by approximately a factor of two for RBC and BoA, and a factor of 1.5 for DEU and SG. Curiously, the CSFB charges do not change by much after incorporating uncertainty.

Overall, after adjusting for parameter uncertainty reduces the average VaR in the IGARCH model. Over a relatively short 200 day window for inference, the estimated Bayesian VaR are approximately twice as large (in absolute size) than the plug-in VaR, which ignores parameter uncertainty. However, when compared judged against the Bayesian VaR, three of the five banks continue to overstate their VaR and over-allocate capital. Given the choice for model used and training window, parameter uncertainty can only partially explain the capital charge puzzle.

This does not necessarily reject parameter uncertainty as an explaination, but it does indicate that more extreme ignorance is required to justify the banks level. Unfortunately, assumptions for the level of parameter ignorance within banks is somewhat subjective, although there is a limit reasonable assumptions. The IGARCH model has only two parameters and is one of the simplest possible specifications for stochastic volatility; therefore, the size of the parameter uncertainty effect in the model is fairly modest and the results should be interpreted as a lower bound for the effect of parameter uncertainty in stochastic volatility models.

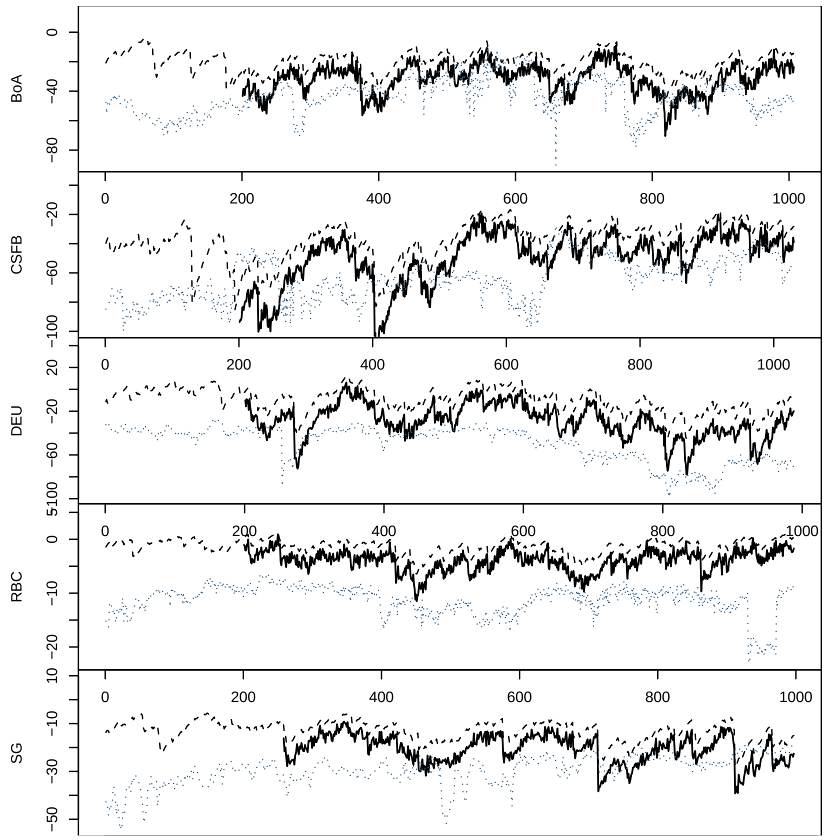


Figure 5: Bayesian VaR with 200 day window under the IGARCH(1) model (black, solid), the non-adjusted VaR (black, dashed line) and the banks' reported VaR (red, dashed line). The adjusted IGARCH(1) VaR are uniformly lower than the non-adjusted VaR. The banks VaR still appear conservative.

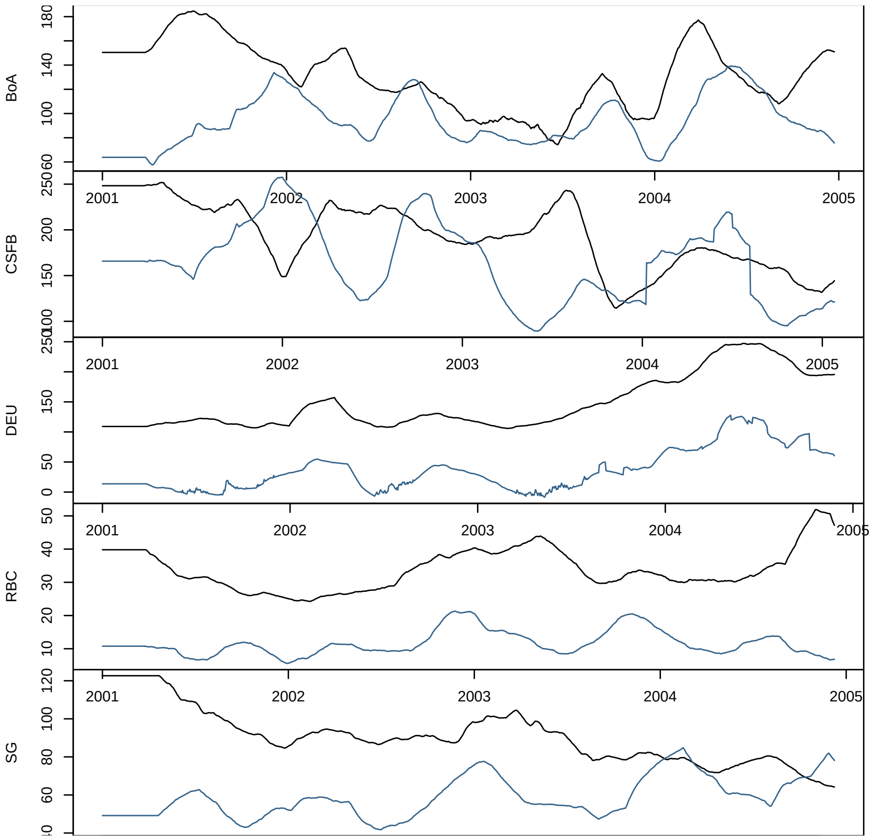


Figure 6: Capital charges under the 200 day window Bayesian VaR estimator in an IGARCH(1) model and banks' reported VaR. The BoA and CSFB average capital charges under the two models are almost equal. The DEU, RBC and SG capital charges are considerably lower on the adjusted-IGARCH(1) model than the reported VaR.

9 Conclusion

This paper presents a general Bayesian estimator for Value-at-Risk and uses it to analyse bank VaR time-series. Bayesian VaR optimally incorporates parameter uncertainty into estimates by integrating over the posterior of each unknown variables. We show that Bayesian VaR estimates are uniformly larger in magnitude (more conservative) than usual "plug-in" estimates, which ignore parameter uncertainty.

The Bayesian VaR estimator is then applied to testing whether parameter uncertainty can explain the capital charge puzzle, or the apparent overstatement VaR by commercial banks. A sample of 5 commercial banks' daily VaR and trading revenue is analysed, and using very simple alternative models, we show that each bank overstates their VaR by comparing capital charges under each model. Using Markov Chain Monte Carlo and a sequential estimation procedure, the Bayesian VaR is estimated for each bank using an IGARCH model. Parameters in the model are fitted using a sliding inference window of 200 trading days. Given the priors, model and window size, we find that the Bayesian VaR from the IGARCH model is approximately twice the absolute size than when using the plug-in estimator. However, even adjusting for parameter uncertainty, three out of five banks in the sample continue to overstate their VaR.

References

[1] Bakshi, G., Panayotov, G., "The Capital Adeqacy Puzzle," *working paper, Smith Business School, University of Maryland*, 2006.

[2] Barone-Adesi, G., Giannopoulos, K., Vosper, L., "Filtering Historical Simulation: Backtest Analysis," working paper, *University of Westminster*, 2000.

[3] Basel Committee, "Overview of the amendment to the capital accord to incorporate market risks," working paper, Basel Committee on Banking Supervision, 1996a.

[4] Basel Committee, "Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements," working paper, Basel Committee on Banking Supervision, 1996b.

[5] Berkowitz, J. & O'Brien, J., "How accurate are value-at-risk models at commercial banks?" *Journal of Finance*, 57, 1093-1111, 2002.

[6] Berkowitz, J., Christoffersen, P., & Pelletier, D., "Evaluating Value-at-Risk models with desk-level data," *working paper*, 2006.

[7] Bollerslev, T., "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307-327, 1986.

[8] Campbell, H., & Siddique, A., "Autoregressive conditional skewness." *Journal of Financial and Quantitative Analysis,* 34, 465-487, 1999.

[9] Christoffersen, P., "Evaluating interval forecasts," *International Economic Review*, 39, 841-862, 1998.

[10] Cuoco, D., & Liu, H., "An analysis of VaR-based capital requirements," *Journal of Financial Intermediation*,15(3), 362-394, 2006.

[11] Engle, R., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987-1008, 1982.

[12] Engle, R., & Bollerslev, T., "Modeling the Persistence of Conditional Variances,"*Econometric Reviews* 5, 1-50, 1986.

[13] Ewerhart, C., "Banks, internal models and the problem of adverse selection," working paper, *University of Bonn,* 2002.

[14] Gourieroux, C., & Jasiak, J. "Value at Risk," *working paper,* to appear in *Handbook of Financial Econometrics*, 2001.

[15] Jacquier, E., Polson, N., & Rossi, P., "Bayesian analysis of stochastic volatility models (with discussion)," Journal of Business and Economic Statistics 12, 371- 417, 1994.

[16] Jorion, P., "Bank trading risk and systemic risk," working paper, *University of California,* 2004.

[17] Kupiec, P., "Techniques for Verifying the Accuracy of Risk Measurement Models," *Journal of Derivatives*, 3, 73-84, 1995.

[18] Pérignon, C., Smith, D., "The quality and level of Value-at-Risk disclosure by commercial banks," working paper, *Simon Fraser University,* 2006.

[19] Pérignon, C., & Smith, D., "Diversification and Value-at-Risk," working paper, *Simon Fraser University*, 2007.

[20] Polson, N., Stroud, J., and Muller, P., "Practical Filtering for Stochastic Volatility Models," working paper, 2002.

[21] Pritsker, M., "The Hidden Dangers of Historical Simulation," *Federal Reserve Finance and Economics Discussion Series*, 27, 2001.

[22] Smith, D., "Conditional Coskewness and Asset Pricing," working paper, *Simon Fraser University,* 2006.

A Proof of Generalized Inequality

Proposition For all models $(\mathcal{M}, \Theta, X_t)$ where $r_t|\Theta, X_t$ has normal distribution, the Value-at-Risk estimator incorporating parameter uncertainty in parameters Θ and state variables X_{t+1} is always lower than the VaR obtained by plugging-in point estimates $\hat{\Theta} = E(\Theta|Y_t)$, $\hat{X}_t = E(X_t|Y_t)$.

Proof By Bayes theorem,

$$p(r_{t+1}|Y_t) = \int_{\Theta, X_{t+1}} p(r_{t+1}|\Theta, X_{t+1}, Y_t) p(\Theta, X_{t+1}|Y) d\Theta dX_{t+1},$$

and conditional on Θ and X_{t+1},$p(r_{t+1}|\Theta, X_{t+1}, Y_t) = p(r_{t+1}|\Theta, X_{t+1})$. Let $\Phi_{t+1}(x|Y_t) := p(r_{t+1} \leq x|Y_t)$. Then,

$$\begin{aligned}
\Phi(x|Y_t) &= \int_x \int_{\Theta, X_{t+1}} p(r_{t+1}|\Theta, X_{t+1}) p(\Theta, X_{t+1}|Y) d\Theta dX_t dr_{t+1} \\
&= \int_{\Theta, X_{t+1}} \Phi(x, \Theta, X_{t+1}) p(\Theta, X_{t+1}|Y) d\Theta dX_t
\end{aligned}$$

The distribution of $r_{t+1}|\Theta, X_t$ is normal with mean $\mu = f(\Theta)$ and variance $\sigma_t^2 = g(\Theta, X_t)$. It is always possible to find representations of Θ and $X_t$so that f and g are linear functions. For x corresponding to $\Phi(x, \mu, \sigma_t^2) = \alpha < \frac{1}{2}$, $\Phi(x, \mu, \sigma_t^2)$ is convex both μ and σ_t^2. This can be shown by lengthy differentiation using the product rule. Since f and g are linear, $\Phi$is also convex in Θ and X_{t+1}. Jensen's inequality states $E[h(X)] \geq h(E[X])$ for convex h. The bivariate version is similarly $E(h(X, Y)) \geq h(E[X], E[Y])$. Invoking this yields the inequality,

$$\begin{aligned}
\Phi(x|Y_t) &\geq \Phi\left(x, \int_\Theta \Theta p(\Theta|Y_t) d\Theta, \int_{X_{t+1}} X_{t+1} p(X_{t+1}|Y_t) dX_{t+1}\right) \\
&= \Phi(x, \hat{\Theta}, \hat{X}_{t+1}).
\end{aligned}$$

It immediately follows that

$$VaR_{t+1}(\alpha, Y_t) \leq VaR_{t+1}(\alpha, \hat{\Theta}, \hat{X}_{t+1})$$

□