Gibbs posterior inference on value-at-risk

Nicholas Syring,* Liang Hong,† and Ryan Martin‡

September 4, 2018

Abstract

Accurate estimation of value-at-risk (VaR) and assessment of associated uncertainty is crucial for both insurers and regulators, particularly in Europe. Existing approaches link data and VaR indirectly by first linking data to the parameter of a probability model, and then expressing VaR as a function of that parameter. This indirect approach exposes the insurer to model misspecification bias or estimation inefficiency, depending on whether the parameter is finite- or infinite-dimensional. In this paper, we link data and VaR directly via what we call a discrepancy function, and this leads naturally to a Gibbs posterior distribution for VaR that does not suffer from the aforementioned biases and inefficiencies. Asymptotic consistency and root-n concentration rate of the Gibbs posterior are established, and simulations highlight its superior finite-sample performance compared to other approaches.

Keywords and phrases: Direct posterior; discrepancy function; M-estimation; model misspecification; risk capital; robust estimation.

1 Introduction

European insurance regulation Solvency II (2009, e.g., Article 101) stipulates that insurers set their risk capital to the 99.5% Value-at-Risk (VaR) of the potential loss, so inference on VaR has become an important problem for both researchers and practitioners in the field of insurance. Technically, VaR is a quantile of the loss distribution (Artzner et al. 1999) and, therefore, risk capital calculation is estimation of an extreme quantile of the loss distribution. To this end, many classical methods are available (e.g., Embrechts et al. 1999, Dowd 2001 and Gourieroux et al. 2000). Since the accurate estimation of risk capital is directly tied to the insurer's solvency, the assessment of associated uncertainty is a key concern for both insurers and regulators. Gerrard and Tsanakas (2012), Fröhlich and Weng (2015) and Bignozzi and Tsanakas (2016) suggest several different approaches to accounting for estimation uncertainty. The downside is that these approaches require the insurer to pin down a parametric model for the future loss, which subjects the insurer to potential model misspecification bias (e.g., Hong and Martin 2018). To avoid potential model misspecification, an attractive alternative is a nonparametric approach, where the (infinite-dimensional) parameter is the loss distribution itself. For example, Hong and Martin (2017) suggest a mixture of log-normal distributions and

*Department of Mathematics, Washington University, nasyring@wustl.edu.
†Department of Mathematics, Robert Morris University, hong@rmu.edu,
‡Department of Statistics, North Carolina State University, rgmarti3@ncsu.edu.

take a Bayesian approach with an appropriate Dirichlet process prior (e.g., Ferguson 1973; Müller and Quintana 2004; Ghoshal 2010) on the mixing distribution. This yields a posterior for the loss distribution, from which the corresponding posterior for VaR can be obtained in a conceptually straightforward way. However, data are only weakly informative about extreme quantiles, and it is not clear how a nonparametric prior for the full loss distribution would affect the corresponding posterior for VaR. Therefore, based on the currently available tools, we face the following dilemma: either use a parametric model and risk misspecification bias, or use a more flexible nonparametric model and risk inefficiency and/or unexpected bias coming from the prior.

Towards a new approach, first note that the classical Bayesian approach links the data and model parameters through a likelihood function. But a likelihood function requires the insurer to give a complete specification of a statistical model, whether it be parametric or nonparametric, finite- or infinite-dimensional. So the only way to avoid the aforementioned two extremes is to link data and parameters in a different way, not via a likelihood. In this paper we propose to link data and parameters via a *discrepancy function*, of which the log-likelihood is a special case. Specifically, we will introduce a suitable discrepancy function for VaR, and then combine an empirical version of this discrepancy with an informative prior distribution for VaR to obtain what is often called a *Gibbs posterior* (Zhang 2006ab and Bissiri et al. 2016). Roughly, the advantages of this approach are two-fold: first, by working with discrepancy instead of likelihood, we avoid specification of a statistical model, which reduces our susceptibility to model misspecification biases; second, we can express the connection between data and VaR directly, rather than indirectly as a function of other model parameters or as a functional of the loss distribution, which allows us to directly incorporate available prior information about VaR into our Gibbs posterior, improving efficiency.

The remainder of the paper is organized as follows. In Section 2, we introduce the Gibbs posterior. Then, in Section 3, we derive the Gibbs posterior for VaR, establish its asymptotic consistency, and discuss scaling of the Gibbs posterior for the purpose of uncertainty quantification. Section 4 gives two numerical examples to illustrate the accuracy of the Gibbs posterior in estimating VaR. Finally, we give a few concluding remarks in Section 5.

2 Review of Gibbs posteriors

The Gibbs posterior is well-developed in statistics literature. There are at least two derivations of the Gibbs posterior, from the points of view of coherence in inference (Bissiri et al., 2016) and efficiency in prediction (Zhang, 2006b). There are also many applications of Gibbs posteriors in several fields: econometric models in Chernozhukov and Hong (2003), variable selection in Jiang and Tanner (2008), clinical trials in Syring and Martin (2017a), and image analysis in Syring and Martin (2017c). For the specific problem of estimating VaR, we argue the merits of the Gibbs posterior approach by establishing its good asymptotic properties in Section 3.2 and Appendix A, and by demonstrating its accuracy in inference in numerical examples in Section 4. Here we introduce necessary notations, define the Gibbs posterior, and discuss its unique scale parameter.

To describe the Gibbs posterior, we let X be a random variable with distribution function P. A data scientist observes a random sample $X^n = (X_1, \ldots, X_n)$ generated from P, and is interested in a feature/parameter $\theta = \theta(P)$ of P, where θ takes value in a parameter set Θ. Suppose there is a *linking function* $\ell_\theta(x)$ that links data to the parameter such that the true parameter $\theta^\star$ satisfies $\theta^\star = \arg\min_{\theta\in\Theta} E_P\{\ell_\theta(X)\}$. Regarded as a function of θ, $E_P\{\ell_\theta(X)\}$ is called the *discrepancy function* and denoted as $D(\theta)$. Its empirical version $n^{-1}\sum_{i=1}^n \ell_\theta(X_i)$ is denoted by $D_n(\theta)$. For a chosen prior distribution Π for θ, the Gibbs posterior Π_n for a Π-measurable set B is calculated as

$$\Pi_n(B) = \frac{\int_B e^{-\omega n D_n(\theta)}\,\Pi(d\theta)}{\int_\Theta e^{-\omega n D_n(\theta)}\,\Pi(d\theta)}, \tag{1}$$

where ω is a *scale parameter* to be determined; see Section 3.3. The Gibbs posterior combines the strengths from both M-estimation (e.g., van der Vaart 1998, Chapter 5) and Bayesian analysis. As in M-estimation, here data and the parameter are linked together via the discrepancy function instead of a likelihood function, thus model misspecification risk is completely avoided. On the other hand, the Gibbs posterior gives a distribution estimator rather than a point estimator so that uncertainty quantification, in the form of credible intervals, say, can be readily obtained. In contrast to the Bayesian approach in which nuisance parameters are often introduced through the likelihood, we target the parameter of interest directly, so no marginalization steps are needed to obtain inference on the interest parameter. Moreover, since the only parameter appearing in the Gibbs posterior is the interest parameter, informative prior specification is straightforward.

The scale parameter ω in (1), also called the *learning rate*, essentially controls the spread of the Gibbs posterior distribution. It is chosen at the discretion of the data scientist. Regarding the choice of ω, several different proposals have been made by various authors, such as the SafeBayes method in Grünwald and van Ommen (2014), the coherence method in Bissiri et al (2016), and the unit information method in Holmes and Walker (2017), among others. Our approach to setting the value of the scale parameter is to ensure that the posterior credible intervals are *calibrated* in the sense that their frequentist coverage probability is approximately equal to the nominal credibility level. More detailed discussion of our implementation of this choice of ω will be given in Section 3.3.

3 Gibbs posterior for VaR

3.1 Construction

To set the stage, let $X_1,\ldots,X_n$ be independent and identically distributed (iid) insurance claims generated from an unknown loss distribution P. For a given $q\in(0,1)$, let $\theta^\star = \theta_q^\star(P)$ denote the corresponding VaR, the $100q^{\text{th}}$ quantile of P:

$$\theta^\star = \inf\{\theta : P((-\infty,\theta]) \geq q\}.$$

Throughout, we assume X_1 is integrable and that the VaR $\theta^\star$ is identifiable in the sense that, if F is the distribution function of P, then $F(\theta) > F(\theta^\star)$ for all $\theta > \theta^\star$. We are particularly interested in the case $q = 0.995$. Since losses are nonnegative, here Θ is taken to be $\mathbb{R}_+ = [0,\infty)$. Next, define the discrepancy function

$$D(\theta) = \frac{1}{2}\int_{\mathbb{R}_+} \bigl(|\theta - x| - x\bigr)\, P(dx) + \frac{(1-2q)\theta}{2},$$

the expectation of the linking function $\ell_\theta(x) = \frac{1}{2}(|\theta - x| - x) + (\frac{1}{2} - q)\theta$. Koltchinskii (1997) shows that $\theta^\star$ is the minimizer of this discrepancy function; if F is smooth in a neighborhood of $\theta^\star$ this can be verified by setting the derivative of D equal to zero and solving for θ. The point is that expressing VaR as the minimizer of a discrepancy function creates an opportunity to construct a Gibbs posterior as described in Section 2. Given an iid sample $X_1, \ldots, X_n$ from P, define the empirical discrepancy function

$$D_n(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \bigl(|\theta - X_i| - X_i\bigr) + \frac{(1-2q)\theta}{2}.$$

Combining the empirical discrepancy function $D_n(\theta)$ with a prior distribution Π on Θ, we can then construct a Gibbs posterior for VaR according to (1). Note that, for the kinds of applications we have in mind, where $q \gg 0.5$, the discrepancy function $D_n(\theta)$ increases linearly for large θ. Then, the negative sign in the exponent implies that the integrand in (1) is indeed integrable for any reasonable (possibly improper) prior Π. Therefore, the Gibbs posterior Π_n is a well-defined distribution on $\mathbb{R}_+$. And since it is only one-dimensional, computation of any relevant feature of the Gibbs posterior is straightforward; see Section 4. For example, if Π has a density $\pi(\theta)$, the Gibbs posterior has density

$$\pi_n(\theta) \propto \exp\{-n\omega D_n(\theta)\}\, \pi(\theta). \tag{2}$$

For our theoretical investigations in Section 3.2 below, we take ω to be a fixed constant; but some care is needed in setting its value, and we discuss this in Section 3.3. As for the prior distribution, since VaR is a practically meaningful quantity, the user may very well have genuine prior information available (e.g., based on historical data) from which an informative prior distribution can be constructed. As a general recommendation, we suggest a gamma prior distribution with shape and scale parameters chosen so that the mean reflects some genuine prior information and the standard deviation is some (potentially large) fraction of that mean; see Section 4 for more details on this prior specification.

3.2 Properties

A desirable property of any statistical procedure is consistency, i.e., when the sample size is large, the estimator, etc., is close to the true value with high probability. Consistency results such as these are standard for the classical VaR estimators based on quantiles of the empirical distribution (e.g. van der Vaart 1998, Chapter 21). In the present context, we say that our Gibbs posterior Π_n for $\theta^\star = \theta_q^\star(P)$, with $q \in (0,1)$ fixed, is *consistent* if

$$\Pi_n(\{\theta : |\theta - \theta^\star(P)| > \varepsilon\}) \to 0 \quad \text{in } P\text{-probability, for all } \varepsilon > 0, \text{ as } n \to \infty.$$

The next theorem shows that, under mild conditions on the prior, the Gibbs posterior in (2) is consistent over a wide range of distributions P.

Theorem 3.1. *Let P be the true loss distribution and $\theta^\star = \theta_q^\star(P)$ denote the true VaR, with $q \in (0,1)$ fixed. If the prior, Π, is continuous and bounded away from zero on any neighborhood of $\theta^\star$, then the Gibbs posterior* (2) *is consistent.*

Proof. Take any $\epsilon > 0$ and define an ϵ neighborhood around $\theta^\star$, that is, $A_\epsilon = \{\theta : |\theta - \theta^\star| \le \epsilon\}$. Set $\delta = \epsilon/2$ and split the sample space into the two disjoint regions:

$$\mathbb{X}_n = \{(x_1, \ldots, x_n) : |\hat\theta_n - \theta^\star| \le \delta\} \quad \text{and} \quad \mathbb{X}_n^c = \{(x_1, \ldots, x_n) : |\hat\theta_n - \theta^\star| > \delta\},$$

where $\hat{\theta}_n$ denotes a minimizer of $D_n(\theta)$, which is known to be a consistent estimator for $\theta^\star$; see Koltchinskii (1997). Next, write the Gibbs posterior probability of A_ϵ^c as

$$\Pi_n(A_\epsilon^c) = \Pi_n(A_\epsilon^c)\, 1_{\mathbb{X}_n}(X^n) + \Pi_n(A_\epsilon^c)\, 1_{\mathbb{X}_n^c}(X^n),$$

where $1_B(x)$ takes value 1 if $x \in B$ and 0 otherwise. Consistency of $\hat{\theta}_n$ implies the second term vanishes in P-probability as $n \to \infty$. Therefore, we only need to analyze the Gibbs posterior probability for "nice" data sets X^n that reside in $\mathbb{X}_n$. Rewrite this Gibbs posterior probability as a ratio $\Pi_n(A_\epsilon^c) = N_n/I_n$, where

$$N_n = \int_{A_\epsilon^c} e^{-\omega n\{D_n(\theta)-D_n(\theta^\star)\}}\, \Pi(d\theta) \quad \text{and} \quad I_n = \int_{\mathbb{R}_+} e^{-\omega n\{D_n(\theta)-D_n(\theta^\star)\}}\, \Pi(d\theta).$$

Our strategy is to bound the numerator from above and the denominator from below in such a way that the posterior probability vanishes in the limit as $n \to \infty$ in P-probability.

The denominator, I_n, can be bounded by $e^{-\omega na}$ for any $a > 0$ in a manner such as that in Lemma 4.4.1 of Ghosh and Ramamoorthi (2003); the crux of their argument, phrased in our context, is that $D_n(\theta) \to D(\theta)$ in P-probability for all fixed θ, which follows from the law of large numbers. The only condition needed to prove this lower bound is that the prior assign positive mass to all "discrepancy neighborhoods" of $\theta^\star$, i.e., to sets of the form

$$\{\theta : D(\theta) - D(\theta^\star) \leq r\}, \quad \text{for all } r > 0.$$

But the discrepancy function difference may be written as

$$D(\theta) - D(\theta^\star) = \theta\{F(\theta) - q\} - \theta^\star\{F(\theta^\star) - q\} - \int_{\theta^\star}^{\theta} x\, P(dx),$$

when $\theta > \theta^\star$, where F is the distribution function corresponding to P. Then, bounding the last term by $\int_{\theta^\star}^{\theta} x\, P(dx) \geq \theta^\star\{F(\theta) - F(\theta^\star)\}$, we have

$$D(\theta) - D(\theta^\star) \leq (\theta - \theta^\star)\{F(\theta) - q\}.$$

For $\theta^\star > \theta$, the bounding is similar: $D(\theta) - D(\theta^\star) \leq (\theta^\star - \theta)\{q - F(\theta)\}$. Hence we have

$$\{\theta : D(\theta) - D(\theta^\star) \leq r\} \supset \{\theta : |\theta - \theta^\star| < r\},$$

and our assumption on Π implies that the discrepancy neighborhood has positive prior probability. This verifies the only requirement of the denominator bound given by Lemma 4.4.1 of Ghosh and Ramamoorthi (2003).

Now consider the numerator, N_n. In the integrand, add and subtract $D_n(\theta^\star \pm \epsilon)$ in the exponent, where "$\pm$" denotes whichever of $\theta^\star + \epsilon$ or $\theta^\star - \epsilon$ has smaller value at $D_n(\cdot)$. Then, by convexity, and the fact that $\hat{\theta}_n \notin A_\epsilon^c$,

$$\begin{aligned} D_n(\theta) - D_n(\theta^\star) &= D_n(\theta) - D_n(\theta^\star \pm \epsilon) + D_n(\theta^\star \pm \epsilon) - D_n(\theta^\star) \\ &\geq D_n(\theta^\star \pm \epsilon) - D_n(\theta^\star). \end{aligned}$$

Therefore, on $\mathbb{X}_n$,

$$N_n = \int_{A_\epsilon^c} e^{-\omega n\{D_n(\theta)-D_n(\theta^\star)\}} \Pi(d\theta) \leq e^{-\omega n\{D_n(\theta^\star \pm \epsilon)-D_n(\theta^\star)\}}.$$

By the law of large numbers, $D_n(\theta^\star \pm \epsilon) - D_n(\theta^\star) \to D(\theta^\star \pm \epsilon) - D(\theta^\star)$ in P-probability which, by the uniqueness of $\theta^\star$, is some positive value, say, $t > 0$. Therefore, the upper bound above will be less than $e^{-\omega n t/2}$ with P-probability converging to 1.

Putting together the bounds on the numerator and denominator, we have that, with P-probability converging to 1,

$$\Pi_n(A_\epsilon^c) \leq \frac{N_n}{I_n} \leq e^{-\omega n(t/2 - a)}.$$

But this holds for any $a > 0$, so taking $a < t/2$ shows that the left-hand side above must converge in P-probability to 0 as $n \to \infty$, which proves the consistency claim. □

We can also identify the minimal size of the neighborhoods needed such that the Gibbs posterior remains consistent. The size of the neighborhoods is then referred to as the "concentration rate" of the Gibbs posterior. Specifically, we have that

$$\Pi_n(\{\theta : |\theta - \theta^\star| > M_n n^{-1/2}\}) \to 0 \quad \text{in } P\text{-probability} \tag{3}$$

for any $M_n \to \infty$. Therefore, the Gibbs posterior has the same root-n rate as the M-estimator of the quantile (Koltchinskii 1997). For details, see Theorem A.1 in the appendix.

3.3 Tuning the scale

In Section 3.1, we showed the Gibbs posterior for the VaR is formulated using a discrepancy function exactly equal to the loss minimized in M-estimation. However, unlike M-estimation, the Gibbs posterior includes the additional scale parameter ω. Multiplying the discrepancy function by a positive constant has no effect on M-estimation because it does not change the location of its minimizer, but it does affect the Gibbs posterior. In particular, the scale parameter tends to tighten/flatten the Gibbs posterior by placing more/less weight on the discrepancy function versus the prior probability distribution. This has the effect of changing the length and the coverage probability of posterior credible intervals. Hence, by carefully selecting the scale parameter we can ensure posterior credible intervals for the VaR actually attain their nominal coverage probability. Algorithm 1 provides a brief description of the method used to calibrate Gibbs posterior credible intervals for VaR by selecting an appropriate value of the scale parameter ω. The basic idea is to get a bootstrap-based approximation of the coverage probability of the Gibbs posterior credible intervals, and then use stochastic approximation to solve for the ω value that puts this coverage probability close to the target confidence level. The algorithm is applied to both real and simulated data in Section 4. A general version of this algorithm has been implemented successfully for several other Gibbs and Bayesian posteriors; see Syring and Martin (2017b). Additionally, R codes for calibrating the Gibbs posterior for the VaR, including a fast implementation using C++, are available at `https://github.com/nasyring/GPC`.

4 Examples

4.1 1990 Norwegian fire claims data

Our goal here is to produce a Gibbs posterior distribution for the VaR, with $q = 0.995$, based on the 1990 Norwegian fire claims data which has been analyzed by Brazauska and

Algorithm 1 — Gibbs Posterior Calibration for VaR.

Fix a nominal confidence level $\alpha \in (0,1)$, a convergence tolerance $\varepsilon > 0$, and an initial guess $\omega^{(0)}$ of the scale parameter. Resample B data sets of size n from X^n with replacement, denoted $\tilde{X}_1^n, \ldots, \tilde{X}_B^n$. Set $t = 0$ and do:

1. For each $b = 1, \ldots, B$, construct the Gibbs posterior based on data $\tilde{X}_b^n$ and compute the corresponding $100(1-\alpha)\%$ credible interval $\widetilde{C}_{n,\alpha}^b$.
2. Evaluate the empirical coverage probability

$$\hat{c}_\alpha(\omega^{(t)} \mid \mathbb{P}_n) = B^{-1} \sum_{i=1}^{B} 1\{\widetilde{C}_{n,\alpha}^b \ni \hat{\theta}_n\},$$

the proportion of the B intervals containing the M-estimator of the q^{th}-quantile, $\hat{\theta}_n$.

3. If $\left|\hat{c}_\alpha(\omega^{(t)} \mid \mathbb{P}_n) - (1-\alpha)\right| < \varepsilon$, then stop and return $\omega^{(t)}$ as the output; otherwise, update $\omega^{(t)}$ to $\omega^{(t+1)}$ according to $\omega^{(t+1)} = \omega^{(t)} + (t+1)^{-3/4}\{\hat{c}_\alpha(\omega^{(t)} \mid \mathbb{P}_n) - 1 + \alpha\}$, set $t \leftarrow t+1$, and go back to Step 1.

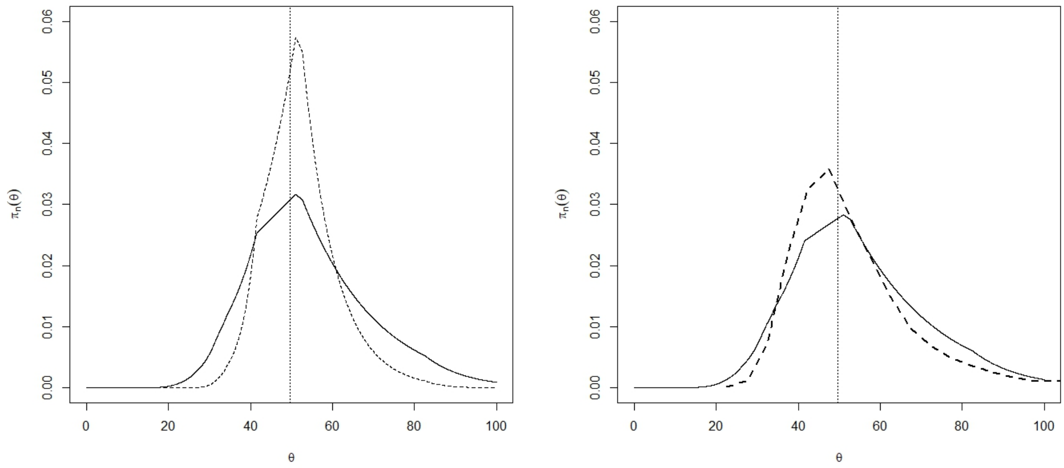
Kleefed (2016) and Mdziniso and Cooray (2018). Here we divide all the claim amounts by 500, so that the range is $[1, 290]$ and $[1, 157]$ for the 1989 and 1990 data sets, respectively. In this case, the naive VaR estimate based on the (scaled) 1989 data is $m = 64.8$. Since it is certainly possible for the 1990 data to differ in significant ways from the 1989 data—in fact, the naive VaR estimate from the 1990 data is 49.8, not so close to that of the 1989 data—we do not want to commit to this value too much, we only want to use it to point our Gibbs posterior in generally the right direction. Therefore, we use a relatively "flat" $\mathsf{Gamma}(a, b)$ prior for VaR, one whose mean is the historical estimate m and whose standard deviation is cm, where $c \in (0, 1]$. This can be achieved by taking the gamma shape and scale parameters as $a = c^{-2}$ and $b = c^2 m$, respectively. Here we take $c = 2^{-1/2}$, so that $a = 2$ and $b = m/2 = 32.4$, but an insurer is free to use a different prior if he/she chooses.

Figure 1 displays Gibbs posteriors with prior distribution as stated above and a nonparametric Bayesian posterior based on that in Hong and Martin (2017) for the Norwegian fire claims 1990 data set. Panel (a) shows that the Gibbs posterior is concentrated near the empirical 99.5% quantile, and that Algorithm 1 decreases the scaling parameter from the default choice of 0.10, resulting in a wider Gibbs posterior but with otherwise the same shape. Panel (b) compares the calibrated Gibbs posterior and the nonparametric Bayesian posterior from Hong and Martin (2017) based on a Dirichlet Process mixture of log-normals. Both the Gibbs posterior and that based on the nonparametric Bayes model have roughly the same center, near the empirical quantile, but simulations in Section 4.2 suggest that the nonparametric Bayesian posterior usually has higher spread.

4.2 Simulated data

The following simulation demonstrates the effectiveness of Algorithm 1 in calibrating Gibbs posterior credible intervals. We simulate $n = 500$ losses from two distributions:

(a) the Pareto (Type II) distribution with parameters $\sigma = 300$ and $\tau = 2$, so that the



(a) Gibbs with two ω values (b) Comparison of Gibbs with DP mixture

Figure 1: Panel (a): Plots of the Gibbs posterior density (2) for the 1990 Norwegian fire claims data based on two settings of the scaling parameter ω: dashed curve corresponds to $\omega = 0.10$ and solid curve corresponds to the calibrated Gibbs posterior with $\omega \approx 0.038$. Vertical dotted line represents the empirical estimate of VaR, i.e., the 99.5% sample quantile. Panel (b): Plots of the calibrated Gibbs posterior and the nonparametric Bayesian posterior from Hong and Martin (2017).

99.5% quantile is approximately $3,943$;

(b) the Weibull distribution with shape and scale parameter values 1.5 and 3000, respectively, so that the 99.5% quantile is approximately $9,118$.

In order to apply Algorithm 1 we resample the data set with replacement $B = 200$ times and for each of the B data sets we sample the corresponding Gibbs posterior $M = 5000$ times, and compute 95%-credible intervals ($\alpha = 0.05$). For comparison, we repeat the simulation using three other approaches, bootstrapping the M-estimate of the VaR, the nonparametric DP mixture posterior in Hong and Martin (2017), and the "oracle" Bayesian posterior based on the true (Pareto or Weibull) likelihood. For the parametric Bayesian posteriors and the Gibbs posterior, a flat prior is used. As discussed in Section 1, the Gibbs posterior has advantages over both the M-estimator and the nonparametric Bayesian approach, which suggest it could produce more efficient inferences about the VaR. Compared to the bootstrap alone, the Gibbs posterior has the advantage of incorporating prior information, while compared to nonparametric Bayesian models the Gibbs posterior is simpler, modeling the VaR directly without needing to estimate the entire loss distribution. For the Gibbs posteriors, the Bayesian oracle posteriors and the bootstrap we repeated the simulation 5000 times, and used 5000 posterior samples/bootstrap resamples to compute interval estimates. However, for the nonparametric Bayesian posteriors we used only 1000 posterior samples and only 1000 repetitions of the simulation due to the high computational time of that approach. The results of the simulations are displayed in Table 1. The "oracle" results we include above represent the best possible inference when the true model is known, but in all realistic problems the true model is not known. We found that using the wrong model, e.g. Weibull for

		Pareto VaR	Weibull VaR
Oracle posterior	coverage	0.96	0.98
	length	4.23	1.82
Gibbs posterior	coverage	0.92	0.97
	length	5.08	2.82
bootstrapped M-est.	coverage	0.88	0.87
	length	6.50	2.46
DP mixture	coverage	0.95	0.90
	length	5.85	3.45

Table 1: Empirical coverage probabilities of 95% confidence/credible intervals and average interval lengths (in thousands) for four methods.

Pareto data and vice versa, could result in enough bias to cause credible intervals to have negligible coverage probability. While these may not be "good" models, the point is that it is not clear how to choose a good model, so it is valuable to avoid modeling altogether. The three robust techniques covered (the Gibbs posterior, the bootstrap, and the nonparametric Bayesian posterior) all attempt to provide reliable inference without the need for the user to specify a finite-dimensional parametric model. And, at least in these two examples, the Gibbs posterior calibrated using Algorithm 1 does the best job.

5 Discussion

Inference on VaR is an important and challenging problem, in part because data are, by definition, not-so-informative about extreme quantiles, but also because it is not possible to directly link data to VaR via a probability model. In this paper, we used special characterization of VaR as the minimizer of a suitable discrepancy function to construct a direct Gibbs posterior for VaR, one that does not have to first pass through a probability model, perhaps with nuisance parameters. This allows the insurer to directly incorporate prior information about VaR, which is likely to be available from historical records, and also to reduce their risk of model misspecification biases and/or inefficiencies that may result from adopting robust-yet-complex model formulations.

The Gibbs posterior has a delicate dependence on a scale parameter ω in the sense that the choice of this parameter can drastically affect the finite-sample performance. Here we have employed a Gibbs posterior calibration strategy from Syring and Martin (2017b) that ensures the Gibbs posterior credible intervals achieve the nominal frequentist coverage probability, thereby leading to valid uncertainty quantification about VaR.

The approach presented in this paper is not specific to VaR. In fact, it is common in econometric and other insurance-related applications to use methods that rely minimally on a statistical model, such as estimating equations. These approaches are useful for producing point estimates, but uncertainty quantification requires asymptotic approximations, etc. The calibrated Gibbs approach employed here provides a nice Bayesian-like alternative to these methods, wherein a posterior distribution is constructed without specifying a statistical model and the readily obtained credible regions attain the nominal coverage probability via the calibration algorithm. We believe that this Gibbs posterior approach is interesting and powerful, and we hope to present on other relevant applications elsewhere.

Acknowledgments

This work is partially supported by the U.S. National Science Foundation, DMS–1811802.

A Gibbs posterior concentration rate

Theorem A.1. *Let P be the true loss distribution and $\theta^\star = \theta_q^\star(P)$ denote the true VaR, with $q \in (0,1)$ fixed. If the prior Π for θ has a density function continuous and bounded away from zero at $\theta^\star$, then the Gibbs posterior Π_n satisfies* (3).

Proof. We prove this by verifying the three conditions of Theorem 4.2.1 in Syring (2017). That is, we will show that there exists constants $C_1, C_2, c > 0$ such that for every n and every sufficiently small $\delta > 0$

$$\sup_{|\theta-\theta^\star|>\delta} \{D(\theta^\star) - D(\theta)\} \leq -C_1\delta^2, \tag{4}$$

$$E \sup_{|\theta-\theta^\star|<\delta} |\mathbb{G}_n(\ell_\theta - \ell_{\theta^\star})| \leq C_2\delta, \tag{5}$$

$$\Pi(\{\theta : D(\theta) - D(\theta^\star) < n^{-\epsilon/6}\}) \gtrsim e^{-5c\epsilon n/6}, \tag{6}$$

where $\epsilon > 0$ and $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ is the empirical process.

By direct calculation, similar to that shown in the proof of Theorem 3.1, Condition (4) can be verified. Next, we take the constant function δ as an envelop for the class $\mathcal{F} = \{\ell_\theta - \ell_{\theta^\star} : |\theta - \theta^\star| < \delta\}$, where $\ell_\theta(x) = \frac{1}{2}(|\theta - x| - x) + (1/2 - q)\theta$. Example 19.7 in van der Vaart (1998) provides a bound on the bracketing number of the set $\mathcal{F}$ from which follows a bound on the empirical process as in Condition (5); see Corollaries 5.53 and 19.35 in van der Vaart (1998). Finally, Condition (6) follows from the fact that

$$\Pi(\{\theta : D(\theta) - D(\theta^\star) < n^{-\epsilon/6}\}) \geq \Pi(\{\theta : |\theta - \theta^\star| < n^{-\epsilon/6}\})$$

and, by the stated assumption on the prior, the right-hand side is lower bounded by a constant times $n^{-\epsilon/6}$, which is greater than anything exponentially small in n. □

References

Artzner, P., Eber, J.M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance* 9(3), 203–228.

Bignozzi, V. and Tsanakas, A. (2016). Parameter uncertainty and residual estimation risk. *Journal of Risk and Insurance* 83 (4), 949–978.

Bissiri, P., Holmes, C. and Walker, S. (2016). A general framework for updating belief distribution. *Journal of Royal Statistical Society, Series B–Statistical Methodology* 78(5), 1103–1130.

Brazauskas, Y. and Kleefeld, A. (2016). Modeling severity and measuring tail risk of Norwegian fire claims. *North American Actuarial Journal* 20(1), 1–16.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics* 115, 293–346.

Dowd, K. (2001). Estimating VaR with order statistics. *Journal of Derivatives* 8 (3) 23–30.

Embrechts, P., Resnick, S.I. and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal* 3 (2), 30–41.

Ferguson, T. S. (1973). Bayesian analysis of some nonparametric problems. *Annals of Statistics* 1, 209–230.

Fröhlich, A. and Weng, A. (2015). Modeling parameter uncertainty for risk capital calculation. *European Actuarial Journal* 5, 79–112.

Gerrard, R. and Tsanakas, A. (2011). Failure probability under parameter uncertainty. *Risk Analysis* 31, 727–744.

Ghosal, S. (2010). The Dirichlet process, related priors and posterior asymptotics. In *Bayesian Nonparametrics*, Cambridge University Press, 35–79.

Ghosh, J. K. and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*, Springer: New York.

Gourierouxa, C., Laurent, J.P. and Scaillet, O. (2000). Sensitivity analysis of Values at Risk. *Journal of Empirical Finance* 7, 225–245.

Grünwald, P. and Van Ommen T. (2014). Inconsistency of Bayesian inference for misspecified linear models, and a proposal for repairing it. *Bayesian Analysis* 12(4), 1069–1103.

Holmes, C. and Walker, S.G. (2017). Assigning a value to a power likelihood in a Bayesian model. *Biometrika* 104, 497–503.

Hong, L. and Martin, R. (2017). A flexible Bayesian nonparametric model for predicting future insurance claims. *North American Actuarial Journal* 21(2), 228–241.

Hong, L. and Martin, R. (2018). Dirichlet process mixture models for insurance loss data. *Scandinavian Actuarial Journal* 6, 545–554.

Jiang, W. and Tanner, M. A. (2008). Gibbs posterior for variable selection in high-dimensional classification and data mining. *Annals of Statistics* 36(5), 2207-2231.

Koltchinskii, V.I. (1997). M-estimation, convexity and quantiles. *Annals of Statistics* 25(2), 435–477.

Mdziniso, N.C. and Cooray, K. (2018). Odd Pareto families of distributions for modeling loss payment data. *Scandinavian Actuarial Journal* 1, 42–63.

Müller, P. and Quintana, F. A. (2004). Nonparametric Bayesian data analysis. *Statistical Science* 19, 95–110.

Solvency II (2009). `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:335:0001:0155:en:PDF`. Accessed on June 26, 2018.

Syring, N. (2017). Gibbs posterior distributions: new theory and applications (doctoral dissertation). University of Illinois at Chicago. Retrieved from `http://hdl.handle.net/10027/22219`.

Syring, N. and Martin, R. (2017a). Gibbs posterior inference on the minimum clinically important difference. *Journal of Statistical Planning and Inference* 187, 67–77.

Syring, N. and Martin, R. (2017b). Calibrating general posterior credible regions, *Biometrika*, to appear. `https://arxiv.org/abs/1509.00922`.

Syring, N. and Martin, R. (2017c). Robust and rate-optimal Gibbs posterior inference on the boundary of a noisy image, *Annals of Statistics*, revision submitted. `https://arxiv.org/abs/1606.08400`.

van der Vaart, A. W. (1998). *Asymptotic Statistics*, New York: Cambridge University Press.

Zhang, T. (2006a). From ϵ-entropy to KL-entropy: analysis of minimum information complexity density estimation. *Annals of Statistics* 34, 2180–2210.

Zhang, T. (2006b). Information theoretical upper and lower bounds for statistical estimation. *IEEE Transaction on Information Theory* 52, 1307–1321.