Preliminary Draft. Please do not cite, circulate or quote without authors permission

Partially censored posterior for accurate left tail density prediction Draft*

Agnieszka Borowska, Lennart F. Hoogerheide, Siem Jan Koopman

July 2017

Abstract

A novel approach to inference for a specific region of the predictive distribution is introduced. An important domain of application is accurate prediction of financial risk measures, where the area of interest is the left tail of the predictive posterior density of (log)returns. It originates from the Bayesian approach to parameter estimation and time series forecasting, however it provides a more accurate estimation of the density in the region of interest in case of misspecification. In the proposed concept of the Partially Censored Posterior the set of parameters is partitioned into two subsets: the first, for which we consider the standard marginal posterior, and the second, for which we consider a censored conditional posterior. The censoring means that observations outside the region of interest are censored: for those observations only the probability of being outside the region of interest matters. In the second subset we choose parameters that are expected to benefit from censoring. This approach yields more precise parameter estimation than a fully censored posterior for all parameters, and has more focus on the region of interest than a standard approach. Finally, novel ways of time-varying censoring is developed, beneficial from the tail prediction perspective. Extensive simulation and empirical studies show the ability of the introduced method to outperform standard approaches.

Keywords: Bayesian inference; censored likelihood; censored posterior; misspecification; density forecasting; mixtures of Student's t; Value at Risk.

*We would like to thank the participants at the 1st EcoSta 2017 conference (Hong Kong, 15–17 June 2017) for their insightful comments. We are also grateful to Paolo Gorgi for for providing useful comments and suggestions.

1 Introduction

The issue of accurate estimation of the left tail of the predictive distribution of returns is crucial from the risk management perspective and is thus commonly investigated by both academics and practitioners. One of the main reasons for its importance is that it serves for obtaining measures of downside risk for investments, such as Value-at-Risk (VaR) and Expected Shortfall (ES), cf. McNeil et al. (2015) and McNeil and Frey (2000). The task of tail prediction can be seen as a subset-focused kind of the broader problem of density forecasting, which has been rapidly growing in econometrics, finance and macroeconomics (cf. Diks et al., 2011). This expansion can be attributed to an increasing understanding of the limited informativeness of point forecasts. In contrast to these, density forecasts provide a full insight into the forecast uncertainty.

A natural framework, therefore, for analysing density forecasts is the Bayesian one, as it treats all unobserved quantities as parameters to be estimated. This includes the predictions for the observation process. Importantly, the Bayesian approach incorporates the parameter uncertainty into analysis and facilitates dealing with model uncertainty, usually via Bayesian Model Averaging. However, the issue of Bayesian model misspecification still seems to be an open question[1]. A formal approach to this problem is provided by Kleijn and van der Vaart (2006), who show (under stringent conditions) that given an incorrectly specified model, the posterior concentrates close the points in the support of the prior that minimise the Kullback-Leibler divergence with respect to the true data generating process (DGP). This result can be seen as the Bayesian counterpart of the MLE being consistent for the pseudo-true values in the frequentist statistics. Nevertheless, differently than the asymptotic distribution of the MLE, the estimated posterior variance is incorrect in case of misspecification (Kleijn and van der Vaart, 2006). Müller (2013) shows that one can rescale the posterior so that credible sets have the correct coverage.

In the context of tail forecasting, the crucial question is: *what if close is not close enough*? From the perspective of accurate tail prediction obtaining estimates being just *close* to their real values is likely to lead to incorrect risk measures and hence to poor managerial decisions. To enhance inference on a particular region of the predictive density, Gatarek et al. (2014) define the Censored Posterior (CP) (and the censored predictive likelihood), who only they use for the estimation and model combination using Bayesian Model Averaging. A concept underlying their approach is the Censored Likelihood scoring function of Diks et al. (2011), which use it for comparing density forecasts in tails. However, as we discuss in the later part of this paper, for densely parametrised models applied in practice the censored posterior approach is likely to lose too much information or to even lead to incorrect inference.

To overcome these shortcomings we propose a novel concept of the Partially Censored Posterior, where the set of parameters is partitioned into two subsets: the first, for which we consider the standard marginal posterior, and the second, for which we consider a censored conditional posterior. The censoring means that observations outside the region of interest are censored: for those observations only the probability of being outside the region of interest matters. In the second subset we choose parameters that are expected to benefit from censoring. This approach yields more precise parameter estimation than a fully censored posterior for all parameters, and has more focus on the region of interest than the standard approach.

The outline of this paper is as follows. In Section 2 we define the risk measure concepts, discuss the censored likelihood based methods for tail inference and present a simple toy example to illustrate potential benefits of censoring. Section 3 introduces our novel concept of the Partially Censored Posterior, which we first motivate by another simple illustration to move next to a more realistic example using a GARCH-type model. Extending the existing literature on censored likelihood based methods, in Section 4 we introduce a time varying threshold for censoring and show its advantages over the time-constant

[1] At the time of writing there is an active, ongoing debate in the Bayesian community about the issue of Bayesian model misspecification. Interestingly, it seems that there is no common ground on it (yet)! Cf. Robert (2017) and Cross Validated (2017).

threshold that has been used in the literature in studies of similar estimation methods. We provide an empirical application in Section 5. Section 6 concludes.

2 Censored likelihood and censored posterior

Let $\{y_t\}_{t\in\mathbb{Z}}$ be a time series of daily logreturns on a financial asset price, with $y_{1:T} := \{y_1, \ldots, y_T\}$ denoting the observed data[2]. We denote $y_{s:r} = \{y_s, y_{s+1}, \ldots, y_{r-1}, y_r\}$ for $s \leq r$. We assume that $\{y_t\}_{t\in\mathbb{Z}}$ is subject to a dynamic stationary process parametrised by θ, on which we put a prior $p(\theta)$. Let $y_h^* := y_{T+h}$ denote the future return h days ahead from the last observation. We are interested in Bayesian estimation of the conditional predictive density of y_h^*, given the observed series $y_{1:T}$. In particular, we are interested in Bayesian estimation of $100\alpha\%$ VaR [3] , i.e. the conditional (in the sense of McNeil and Frey, 2000) $100(1-\alpha)\%$ quantile of the predictive distribution of y_h^*, i.e.

$$100\alpha\%\,\text{VaR} = \inf\big\{x \in \mathbb{R} : p(y_h^*|y_{1:T}) \geq \alpha\big\}.$$

The regular (uncensored) likelihood is given by the standard formula

$$p(y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|\theta, y_{1:t-1}).$$

Now suppose that we are interested in a particular region of the predictive distribution, e.g. the left tail, given by $A = \{A_1, \ldots, A_T\}$, where $A_t = \{y_t | y_t < C_t\}$ with C_t potentially time-varying. In the context of assessing the performance of forecasting methods in tails Diks et al. (2011) introduce the censored likelihood scoring function, which is useful when the main interest lies in comparing the density forecasts' accuracy for a specific region. Building upon Diks et al. (2011), Gatarek et al. (2014) define the censored likelihood (CL) as

$$p^{cs}(y_{1:T}|\theta) = \prod_{t=1}^{T} p^{cs}(y_t|\theta, y_{1:t-1}), \tag{2.1}$$

i.e. the conditional density $p^{cs}(y_t|\theta, y_{1:t-1})$ of the mixed continuous-discrete distribution for the censored variable $\tilde{y}_t$

$$\tilde{y}_t = \begin{cases} y_t, & \text{if } y_t \in A_t, \\ R_t, & \text{if } y_t \in A_t^C. \end{cases} \tag{2.2}$$

Next, Gatarek et al. (2014) use the CL to define the censored posterior (CP) density as

$$p^{cs}(\theta|y_{1:T}) \propto p(\theta) p^{cs}(y_{1:T}|\theta),$$

which they further apply to specify the censored predictive density. They employed the latter in forecasting experiments of Bayesian Model Averaging.

Definition (2.2) means that the censored variable is equal to the original one in the region of interest,

[2] With a slight abuse of notation we will use the same lowercase letter notation for a random variable and its realisation.

[3] In further research we aim to extend the analysis by considering the Expected Shortfall (ES) as an alternative risk measure. This is due to its advantageous properties compared to VaR, mainly sub-additivity (which makes ES a coherent risk measure in the sense of Artzner et al., 1999). We recall that given $100\alpha\%$ VaR, the conditional ES is defined as

$$100\alpha\%\,\text{ES} = \mathbb{E}\big[PL(y_{1:H}^*)|PL(y_{1:H}^*) < 100\alpha\%\,\text{VaR}\big].$$

while everywhere outside it it is equal to a fixed value R_t. In consequence, the distribution of $\tilde{y}_t$ is mixed: continuous (in A_t) and discrete (in A_t^C). The censored likelihood for the uncensored variable y_t can be thus written as

$$\begin{aligned} p^{cs}(y_t|y_{1:t-1}\theta) &= [p(y_t|y_{1:t-1},\theta)]^{I\{y_t\in A_t\}} \times \left[\mathbb{P}(y_t \in A_t^C|y_{1:t-1},\theta)\right]^{I\{y_t\in A_t^C\}} \\ &= [p(y_t|y_{1:t-1},\theta)]^{I\{y_t\in A_t\}} \times \left[\int_{A_t^C} p(x|y_{1:t-1},\theta)dx\right]^{I\{y_t\in A_t^C\}}. \end{aligned} \tag{2.3}$$

Differently than with a likelihood of a *censored dataset* where all $y_t \in A_t^C$ are censored and their exact values are ignored, with the censored likelihood the exact value of $y_t \in A_t^C$ still plays a role in conditioning in subsequent periods. Only in the case of i.i.d. observation when $p(y_t|y_{1:t-1},\theta) = p(y_t|\theta)$ both approaches would be equivalent. Hence, all original observation $y_{1:T}$ are taken into account for the estimation or prediction in the region of interest so that no information is lost. Moreover, conditioning on the true values of y_t and not the censored ones is computationally easier.

2.1 Scoring rules in density forecasting

Gatarek et al. (2014) notice that the density (2.3) is equal to the exponent of the censored likelihood score function of Diks et al. (2011). The latter authors consider Diebold and Mariano (1995) test for comparing accuracy of two sequences of density forecasts $\hat{f}_t$ and $\hat{g}_t$. They notice that taking into account the total probability of the region of interest is crucial from forecasting perspective. Alternative region-focused score functions might simply ignore the observations outside the region of interest, but this results in losing information about the frequency of this region occurring. Below we review the main concepts related to scoring rules for evaluating density forecasts.

Kullback-Leibler information criterion and scoring rules In our empirical study in Section 5 we consider the Kullback-Leibler Information Criterion (KLIC) to perform Diebold-Mariano (DM) tests of equal predictive accuracy using loss functions that depend on the forecasted density[4]. The KLIC is an information theoretic goodness-of-fit measure and formally, for the density forecast $\hat{f}_t$, is defined as

$$\begin{aligned} \mathrm{KLIC}(\hat{f}_t) &= \mathbb{E}_t\left[\log p_t(y_{t+1}) - \log \hat{f}_t(y_{t+1})\right] \\ &= \int p_t(y_{t+1}) \log\left(\frac{p_t(y_{t+1})}{\hat{f}_t(y_{t+1})}\right) dy_{t+1}, \end{aligned}$$

where p_t denotes the true conditional density. Diks et al. (2011) show that for two competing density forecasts $\hat{f}_t$ and $\hat{g}_t$ their relative KLIC values correspond exactly to the difference of their logarithmic scoring rules (SR) (cf. Amisano and Giacomini, 2007), with

$$\begin{aligned} d_t^l &= S^l(\hat{f}_t; y_{t+1}) - S^l(\hat{g}_t; y_{t+1}), \\ S^l(\hat{f}_t; y_{t+1}) &= \log \hat{f}_t(y_{t+1}). \end{aligned}$$

The logarithmic scoring rule assigns a high score to a density forecast if the observation y_{t+1} falls within a region with high predictive density $\hat{f}_t$, and a low score if it falls within a region with low predictive density. Given a set of parameter draws $\{\theta^{(i)}\}_{i=1}^M$ (either from the regular posterior, the CP or the partially censored posterior, formally specified in Section 3.2) and the data $y_{1:T} = \{y_1, \ldots, y_T\}$, the

[4]In the simulation study we know the true quantiles so we can compute the mean square errors (MSEs) of the forecasts.

predictive density can be estimated as

$$\begin{aligned} p(y_{T+1}|y_{1:T}) &= \int p(y_{T+1}|y_{1:T},\theta)p(\theta|y_{1:T})d\theta \\ &= \frac{1}{M}\sum_{i=1}^{M} p(y_{T+1}|y_{1:T},\theta^{(i)}). \end{aligned}$$

The average log score difference of a sequence of H one-step-ahead forecasts

$$\bar{d}_{T,H} = \frac{1}{H}\sum_{h=T}^{T+H} d_h^l$$

can be then used to perform a test on the difference of the predictive accuracy using a Diebold and Mariano (1995) type statistic

$$t_{T,H} = \frac{\bar{d}_{T,H}}{\sqrt{\hat{\sigma}^2_{T,H}/H}},$$

where $\hat{\sigma}^2_{T,H}$ is a heteroskedasticity and autocorrelation-consistent (HAC) variance estimator of $\sigma^2_{T,H} = \mathrm{Var}\left[\sqrt{H}\bar{d}_{T,H}\right]$.

Censored likelihood score function Diks et al. (2011) show that two density forecasts can be compared on a specific region of interest by using SRs based on partial likelihood, e.g. the censored likelihood, without facing the problem that the KLIC-based scoring rule favours fat-tailed density forecasts. Such favouring would result in biased scores towards density forecasts that have more probability mass in the specific region of interest. We can overcome this problem by taking score functions as the logarithm of a likelihood (or partial likelihood) function of the realizations of some statistical experiment, as described inDiks et al. (2011). Note that a necessary condition is that we consider normalized densities that integrate to unity. We measure the distance between a density $\hat{f}_t$ and a true conditional density p_t

Diks et al. (2011) define the censored likelihood (CSL) score function as

$$S^{csl}(\hat{f}_t; y_{t+1}) = \mathbb{I}(y_{t+1} \in A_t)\log \hat{f}_t(y_{t+1}) + \mathbb{I}(y_{t+1} \in A_t^c)\log\left(\int_{A_t^c} \hat{f}_t(s)ds\right) \tag{2.4}$$

This scoring rule does not neglect observations falling outside the region of interest but only as far as their total mass is concerned; the shape of the predictive density over A_t^c is ignored.

2.2 Toy application

Consider the true model for y_t being i.i.d. split normal $\mathcal{SN}(\mu,\sigma_1^2,\sigma_2^2)$ (cf. Geweke, 1989)[5], i.e.

$$p(y_t) = \begin{cases} \phi(y_t;\mu,\sigma_1^2), & y_t > \mu, \\ \phi(y_t;\mu,\sigma_2^2), & y_t \le \mu, \end{cases}$$

where $\phi(x;m,s)$ denotes the Gaussian density with mean m and variance s at x. The mean of a random variable distributed according to $\mathcal{SN}(0,\sigma_1^2,\sigma_2^2)$, i.e. with a split at zero, is equal $-\frac{\sigma_2-\sigma_1}{\sqrt{2\pi}}$, which is nonzero for any asymmetric case. Hence, shifting of the split point accordingly to the chosen variances allows

[5]Cf. also a recent paper by de Roon and Karehnke (2016).

us to consider a zero-mean random variable: $y_t \sim \mathcal{SN}(\mu, \sigma_1^2, \sigma_2^2)$ with $\mu := \frac{\sigma_2-\sigma_1}{\sqrt{2\pi}}$ results in $\mathbb{E}[y_t] = 0$. Such a specification is then equivalent to $y_t = \mu + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{SN}(0, \sigma_1^2, \sigma_2^2)$.

We consider $\sigma_1 = 1$ and $\sigma_2 = 2$ hence to obtain $\mathbb{E}[y_t] = 0$ we set $\mu = \frac{1}{\sqrt{2\pi}}$. We generate $T = 100$, $T = 1,000$ and $T = 10,000$ observations from the true model. We are interested in evaluating 95% and 99% VaR, i.e. in the estimation of 5% and 1% quantile of the distribution of y_t, receptively (the left tail). The true values for these quantities are -2.8908 and -4.2538, receptively.

We consider a misspecified model for estimation, which we take to be i.i.d. normal with unknown mean c and variance σ^2 (common for all the values of y_t). We put flat prior on μ, while for σ^2 we specify $p(\sigma) \propto \frac{1}{\sigma}$, which is equivalent with the flat prior for $\tau = \log(\sigma)$, $p(\tau) \propto 1$.

We first perform uncensored posterior analysis. Then, we consider two specifications for the censored posterior. In each the threshold value is constant over time, $A_t = A = \{x : x \leq c\}$, and we consider two different values for the threshold c: one equal to the 10% quantile of the generated sample and another one equal to zero[6]. All the simulations are carried out with $M = 10,000$ posterior draws after a burn-in of $1,000$. Table 2 and Figure 3 present the results of the simulations (we refer to Appendix C.1 to for the plots for the remaining cases). We can see that the regular posterior provide incorrect estimates from the left tail perspective, which are "averaged out" over the whole domain. In contrast to that, the CP by focusing in the estimation on the relevant region, provides much better parameter estimates. Obviously, the precision of the estimates from the CP depends on the degree of censoring: the more censoring, the less information, the lower the precision.

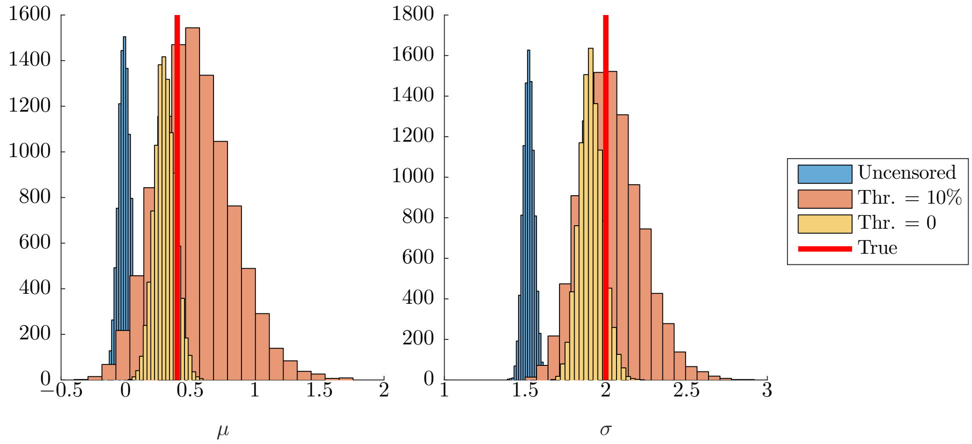


Figure 2.1: Asymmetric (misspecified) i.i.d. mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: histograms of posterior draws without and with censoring (with different thresholds) for $T = 1,000$.

[6] As long as we have grounds to believe that the left tail of the returns distribution is characterised by a higher variance, so that μ needs to be set positive, such a choice is conservative in a sense that the observations from the right tail (right from the split point) will never "contaminate" the censored estimates for the left tail.

Value	True/MC*	Posterior	CP10%	CP0
		$T = 100$		
AR	–	0.8585	0.4671	0.7843
μ	0.0000	0.0926	1.5715	0.1727
		(0.1014)	(1.5261)	(0.1529)
σ	1.0000	1.0226	1.8779	1.1289
		(0.0741)	(0.9051)	(0.1418)
VaR 1%	-2.0976	-2.1245	-2.2322	-2.1519
	[0.5426]	[0.5763]	[0.6812]	[0.6041]
VaR 5%	-1.4804	-1.4899	-1.4668	-1.4951
	[0.2709]	[0.2922]	[0.3123]	[0.2986]
		$T = 1000$		
AR	–	0.8727	0.8143	0.8627
μ	0.0000	0.0071	0.0196	0.0230
		(0.0304)	(0.1473)	(0.0387)
σ	1.0000	0.9604	0.9446	0.9725
		(0.0215)	(0.0921)	(0.0349)
VaR 1%	-2.0927	-2.0998	-2.1020	-2.1074
	[0.5423]	[0.5464]	[0.5500]	[0.5476]
VaR 5%	-1.4804	-1.4816	-1.4823	-1.4858
	[0.2709]	[0.2725]	[0.2734]	[0.2735]
		$T = 10000$		
AR	–	0.8723	0.8667	0.8655
μ	0.0000	0.0031	0.0300	-0.0049
		(0.0100)	(0.0433)	(0.0123)
σ	1.0000	0.9960	1.0053	0.9865
		(0.0071)	(0.0281)	(0.0111)
VaR 1%	-2.1032	-2.0965	-2.0876	-2.0972
	[0.5428]	[0.5427]	[0.5432]	[0.5428]
VaR 5%	-1.4822	-1.4802	-1.4767	-1.4815
	[0.2710]	[0.2712]	[0.2713]	[0.2713]

* True value for parameters, MC mean value for VaRs.

AR: acceptance rate for the independent MH.

CP10%: Censored posterior, threshold 10% sample quantile.

CP0: Censored posterior, threshold 0.

Table 1: Symmetric (correctly specified) i.i.d. zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 1$. Simulation results for posterior without and with censoring (with different thresholds). For the censored posterior the focus is on the left tail. Standard errors in parantheses, MSEs in brackets.

Value	True/MC*	Posterior	CP10%	CP0
		$T = 100$		
AR	–	0.8620	0.6095	0.7790
μ	0.3989	-0.0147	0.9452	0.5321
		(0.1658)	(1.0768)	(0.3171)
σ	2.0000	1.6414	2.5853	2.2724
		(0.1157)	(0.7684)	(0.2935)
VaR 1%	-4.2620	-3.6551	-4.5968	-4.4697
	[0.0066]	[0.5082]	[0.6438]	[0.3506]
VaR 5%	-2.8924	-2.5675	-2.8886	-2.9773
	[0.0017]	[0.1984]	[0.2612]	[0.1402]
		$T = 1000$		
AR	–	0.8706	0.8188	0.8605
μ	0.3989	-0.0103	0.5348	0.3043
		(0.0481)	(0.2895)	(0.0811)
σ	2.0000	1.5229	2.0338	1.9095
		(0.0343)	(0.1872)	(0.0732)
VaR 1%	-4.2476	-3.5549	-4.2739	-4.2701
	[0.0044]	[0.5063]	[0.0527]	[0.0293]
VaR 5%	-2.8921	-2.5101	-2.8895	-2.8882
	[0.0018]	[0.1540]	[0.0158]	[0.0145]
		$T = 10000$		
AR	–	0.8734	0.8672	0.8785
μ	0.3989	0.0334	0.4279	0.4290
		(0.0152)	(0.0901)	(0.0273)
σ	2.0000	1.5125	1.9825	1.9778
		(0.0106)	(0.0568)	(0.0250)
VaR 1%	-4.2610	-3.5654	-4.2610	-4.2583
	[0.0064]	[0.4787]	[0.0098]	[0.0091]
VaR 5%	-2.8940	-2.5226	-2.8919	-2.8917
	[0.0024]	[0.1369]	[0.0031]	[0.0029]

* True value for parameters, MC mean value for VaRs.

AR: acceptance rate for the independent MH.

CP10%: Censored posterior, threshold 10% sample quantile.

CP0: Censored posterior, threshold 0.

Table 2: Asymmetric (misspecified) i.i.d. zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$. Simulation results for posterior without and with censoring (with different thresholds). For the censored posterior the focus is on the left tail. Standard errors in parantheses, MSEs in brackets.

3 Partially Censored Posterior

The previous section explained the advantages of CP with respect to obtaining accurate evaluations of lower quantiles of the predictive posterior distribution. Moreover, Figure demonstrated that compared to the standard posterior the censored posterior provides a better insight into the location of the parameters in the region of interest. The cost of this better location is however a larger variance of the estimate. The reason for that that censoring leads to the analysis based on effectively a smaller sample. In the split normal i.i.d. example we only consider 2 parameters but typically model are characterised by a broader parameters set – not all of which are expected to differ between the whole observation space and the region of interest. For this reason we propose *partial censoring*, where only a selected subset of parameters is being censored while for the remaining ones a standard posterior analysis is carried out.

3.1 AR(1) example

To illustrate the idea and the benefits of partial censoring we employ the simplest dynamic model, i.e. AR(1). We consider the true DGP of the form

$$\begin{aligned} y_t &= \varrho y_{t-1} + \varepsilon_t, \\ \varepsilon_t &\sim \mathcal{SN}(\mu, \sigma_1^2, \sigma_2^2), \\ \mu &:= \frac{\sigma_2 - \sigma_1}{\sqrt{2\pi}}, \end{aligned}$$

with $|\rho| < 1$ so that unconditionally $\mathbb{E}[y_t] = 0$. As previously, we consider a misspecified model for the estimation, which is

$$\begin{aligned} y_t &= \mu + \rho y_{t-1} + \epsilon_t, \\ \epsilon_t &\sim \mathcal{N}(0, \sigma^2), \end{aligned}$$

where ϵ_t are zero-mean for identification of μ. As previously, we consider two time-constant threshold values, with c set equal to the 10th percentile of the generated data and $c = 0$.

The contribution to the censored likelihood of the observation y_t is then

$$p(y_t|y_{1:t-1}, \theta) = \begin{cases} \phi(y_t; \mu + \rho y_{t-1}, \sigma^2), & \text{if } y_t \leq c, \\ 1 - \Phi(\frac{c - (\mu + \rho y_{t-1})}{\sigma}), & \text{if } y_t > c. \end{cases}$$

For μ and σ we set the same priors as in the i.i.d. case, while for ρ adopt a uniform prior over the stationarity region.

3.1.1 Fully Censored Posterior

We first consider one simulated dataset with one-step-ahead forecast and for this dataset we compare the posterior draws with these from the censored posterior, cf. Figure 3.1. Figures C.3 and C.4 in Appendix C.2 provide a broader overview, for the correctly specified and the misspecified case, respectively, for different time series lengths.

As expected, under correct specification censoring only leads to a higher variance of the parameter estimates, which are located correctly (i.e. similarly as by the regular posterior). However, in the misspecified case we can already see the benefits from censoring: the CP locates μ and σ much closer to their true values. Obviously, the regular posterior misses the true values in the tail. Tables 3 and 4 further confirm this finding for the censoring of "little harm" under the correct specification and a huge gain in the misspecified case.

Nevertheless, we can see that even though the CP "gets closer" to the true parameter values for μ and σ, it might overestimate ρ (cf. Tables 3 and 4). This can be attributed to a difficulty of correctly locating the dynamics parameter effectively using few uncensored observations. Hence, a greater variance of the estimates might not be the only cost of adopting the CP approach, as generally the parameters reflecting the dynamic behaviour of the process might get distorted.

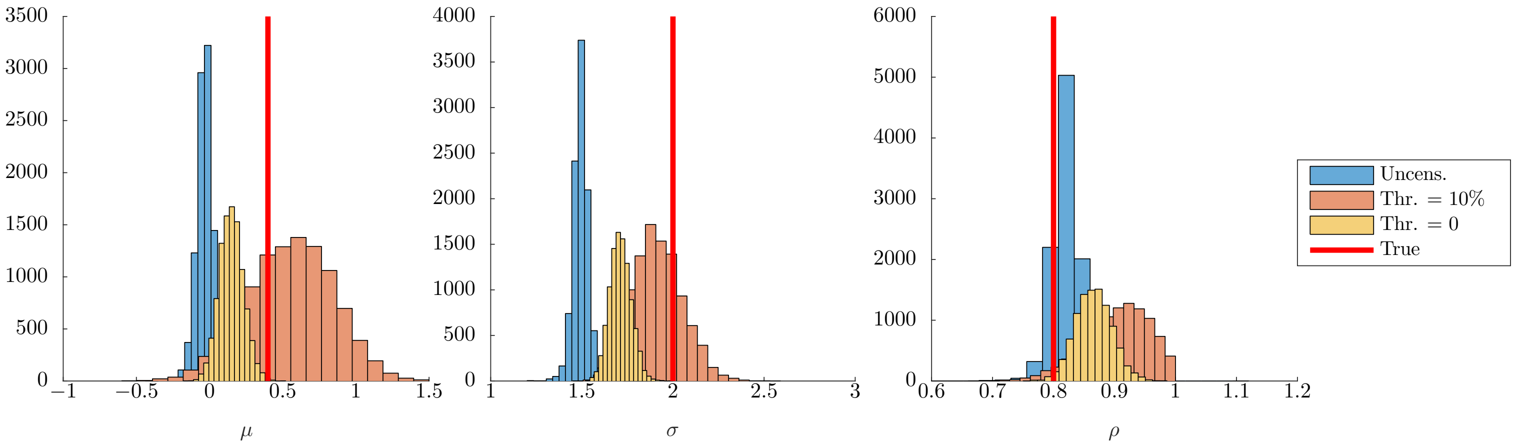


Figure 3.1: AR(1) mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: histograms of posterior draws without and with censoring (with different thresholds), for $T = 1,000$.

Since we know the true DGP we know the true quantiles, so we can compute the mean squared errors (MSEs) of the predicted VaR forecasts. Despite a potential distortion of ρ, the CP is still advantageous when it comes to VaR predictions. Tables 3 and 4 show that in a Monte Carlo simulation study with 100 replications, the MSE for both 1% and 5% one-step-ahead VaR obtained by either of the CP methods are substantially lower than these delivered by the reguar posterior.

Value	True/MC*	Posterior	CP10%	CP0
		$T = 100$		
AR	–	0.9999	0.4311	0.7070
μ	0.0000	0.0198	0.6910	0.0122
		(0.1280)	(0.4464)	(0.1168)
σ	1.0000	0.9984	1.2224	0.9337
		(0.0927)	(0.2812)	(0.1028)
ϕ	0.8000	0.7273	0.9127	0.8259
		(0.0888)	(0.0598)	(0.0680)
VaR 1%	-2.4677	-2.4620	-2.5426	-2.5009
	[0.0013]	[0.0418]	[0.1361]	[0.0527]
VaR 5%	-1.7798	-1.7598	-1.6798	-1.7775
	[0.0004]	[0.0330]	[0.0696]	[0.0331]
		$T = 1000$		
AR	–	0.9999	0.7405	0.8143
μ	0.0000	-0.0298	0.1440	-0.0099
		(0.0390)	(0.1874)	(0.0451)
σ	1.0000	0.9704	1.0390	0.9664
		(0.0277)	(0.0770)	(0.0309)
ϕ	0.8000	0.8169	0.8743	0.8378
		(0.0231)	(0.0592)	(0.0272)
VaR 1%	-2.2554	-2.2559	-2.2564	-2.2641
	[0.0015]	[0.0051]	[0.0184]	[0.0083]
VaR 5%	-1.5722	-1.5720	-1.5426	-1.5700
	[0.0006]	[0.0038]	[0.0177]	[0.0053]
		$T = 10000$		
AR	–	0.8298	0.8181	0.8243
μ	0.0000	-0.0069	0.0606	0.0011
		(0.0100)	(0.0643)	(0.0146)
σ	1.0000	0.9911	1.0082	0.9916
		(0.0070)	(0.0249)	(0.0103)
ϕ	0.8000	0.8075	0.8401	0.8170
		(0.0059)	(0.0218)	(0.0090)
VaR 1%	-2.2316	-2.2295	-2.2379	-2.2321
	[0.0017]	[0.0022]	[0.0034]	[0.0024]
VaR 5%	-1.5438	-1.5454	-1.5404	-1.5418
	[0.0006]	[0.0008]	[0.0025]	[0.0008]

* True value for parameters, MC mean value for VaRs.

AR: acceptance rate for the independent MH.

CP10%: Censored posterior, threshold 10% sample quantile.

CP0: Censored posterior, threshold 0.

Table 3: Symmetric (correctly specified) AR(1) zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 1$. Simulation results for posterior without and with censoring (with different thresholds). For the censored posterior the focus is on the left tail. Standard errors in parentheses, MSEs in brackets.

Value	True/MC*	Posterior	CP10%	CP0
		$T = 100$		
AR	–	0.9999	0.4330	0.6428
μ	0.3989	0.0404	0.9602	0.2872
		(0.1871)	(0.6883)	(0.2417)
σ	2.0000	1.4705	1.9599	1.7384
		(0.1365)	(0.4279)	(0.2037)
ϕ	0.8000	0.7866	0.9212	0.8762
		(0.0796)	(0.0539)	(0.0682)
VaR 1%	-4.5071	-3.7904	-4.5623	-4.2190
	[0.0053]	[0.6295]	[0.4286]	[0.2449]
VaR 5%	-3.1313	-2.7218	-2.9343	-2.9737
	[0.0018]	[0.2672]	[0.2242]	[0.1302]
		$T = 1000$		
AR	–	0.9999	0.7054	0.8116
μ	0.3989	-0.0314	0.5547	0.1497
		(0.0599)	(0.2925)	(0.0817)
σ	2.0000	1.4948	1.9181	1.7119
		(0.0427)	(0.1348)	(0.0578)
ϕ	0.8000	0.8202	0.9042	0.8657
		(0.0229)	(0.0527)	(0.0297)
VaR 1%	-4.1856	-3.4922	-4.1131	-3.8540
	[0.0061]	[0.4988]	[0.0657]	[0.1386]
VaR 5%	-2.8190	-2.4461	-2.7627	-2.6511
	[0.0025]	[0.1552]	[0.0567]	[0.0489]
		$T = 10000$		
AR	–	0.8296	0.8184	0.8244
μ	0.3989	-0.0068	0.4263	0.1360
		(0.0153)	(0.1115)	(0.0256)
σ	2.0000	1.5195	1.9435	1.7216
		(0.0108)	(0.0484)	(0.0185)
ϕ	0.8000	0.8082	0.8512	0.8339
		(0.0059)	(0.0225)	(0.0096)
VaR 1%	-4.0991	-3.3971	-4.0256	-3.7386
	[0.0069]	[0.4776]	[0.0151]	[0.1283]
VaR 5%	-2.7235	-2.3506	-2.7009	-2.5454
	[0.0025]	[0.1394]	[0.0102]	[0.0357]

* True value for parameters, MC mean value for VaRs.

AR: acceptance rate for the independent MH.

CP10%: Censored posterior, threshold 10% sample quantile.

CP0: Censored posterior, threshold 0.

Table 4: Asymmetric (misspecified) AR(1) zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$. Simulation results for posterior without and with censoring (with different thresholds). For the censored posterior the focus is on the left tail. Standard errors in parentheses, MSEs in brackets.

3.1.2 Partially Censored Posterior

Second, we add the partially censored posterior (PCP) to the analysis, where the draws of ρ are kept from the standard posterior. The exact formulation and the details of the algorithm are provided in the next section. Again, we start with one-step-ahead forecasts (to save space we skip the results for the correctly specified case, with $\sigma_2 = 1$).

Figure 3.2 presents the draws from the regular posterior and from the PCP method. By construction, ρ is now located correctly and is characterised by a low variance, as with the standard posterior. However differently than with the regular posterior, the PCP correctly captures the location and the shape of the left tail of the conditional distribution of y_t. Table 7 confirms that the PCP delivers more precise estimates compared to the fully CP, as expected. It also performs better when it comes to the VaR forecasting. We can conclude that generally the PCP outperforms the CP, which outperforms the regular posterior, both in terms of better tail predictions and obtaining more accurate parameter estimates.

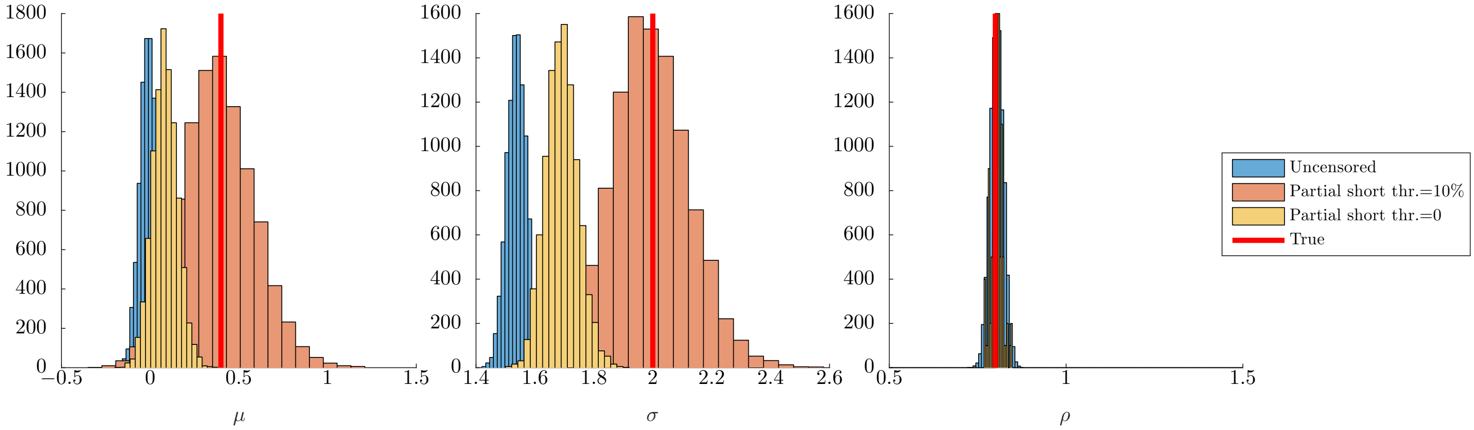


Figure 3.2: AR(1) mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: histograms of posterior draws from the regular posterior and for the PCP, for $T = 1,000$. For the PCP conditional simulation given $\rho^{(i)}$, $i = 1, \ldots, 10,000$, from the posterior, with thresholds the 10% percentile of the simulated data and 0.

Value	True/MC*	Posterior	CP0	CP10%	PCP0	PCP10%
			$T = 100$			
AR	–	0.8668	0.8019	0.7553	0.7805	0.5382
μ	0.3989	0.1417	0.2585	1.1983	0.2680	1.0837
		(0.1498)	(0.2307)	(0.9262)	(0.2315)	(0.8232)
σ	2.0000	1.5257	1.7561	2.3916	1.7671	2.3454
		(0.1103)	(0.2040)	(0.5891)	(0.1987)	(0.5783)
ϕ	0.8000	0.8420	0.8349	0.8721	0.8424	0.8424
		(0.0516)	(0.0723)	(0.0906)	(0.0564)	(0.0564)
VaR 1%	-4.5416	-3.9813	-4.3710	-4.7444	-4.3329	-4.7369
	[0.0050]	[0.5049]	[0.2264]	[0.5155]	[0.2323]	[0.5573]
VaR 5%	-3.1787	-2.8862	-3.1064	-3.1133	-3.0794	-3.1762
	[0.0019]	[0.2161]	[0.1419]	[0.2279]	[0.1412]	[0.2429]
			$T = 1000$			
AR	–	0.8803	0.8811	0.8457	0.7293	0.7327
μ	0.3989	0.0082	0.1056	-0.0270	0.1006	0.2273
		(0.0492)	(0.0847)	(0.3586)	(0.0740)	(0.1976)
σ	2.0000	1.5453	1.7665	1.8609	1.7617	1.9370
		(0.0351)	(0.0603)	(0.1513)	(0.0566)	(0.1253)
ϕ	0.8000	0.8493	0.8515	0.7947	0.8502	0.8502
		(0.0168)	(0.0293)	(0.0684)	(0.0164)	(0.0164)
VaR 1%	-4.1085	-3.4252	-3.7688	-4.0298	-3.7658	-4.0336
	[0.0057]	[0.4714]	[0.1357]	[0.0821]	[0.1307]	[0.0399]
VaR 5%	-2.7380	-2.3666	-2.5656	-2.6815	-2.5740	-2.7191
	[0.0017]	[0.1459]	[0.0510]	[0.0700]	[0.0381]	[0.0169]
			$T = 10000$			
AR	–	0.7908	0.7750	0.7908	0.6963	0.7570
μ	0.3989	0.0146	0.1311	0.2805	0.1179	0.3045
		(0.0155)	(0.0256)	(0.1076)	(0.0237)	(0.0629)
σ	2.0000	1.5317	1.7316	1.9280	1.7238	1.9363
		(0.0109)	(0.0188)	(0.0489)	(0.0179)	(0.0399)
ϕ	0.8000	0.7966	0.8091	0.7910	0.7974	0.7974
		(0.0061)	(0.0100)	(0.0229)	(0.0058)	(0.0058)
VaR 1%	-4.3568	-3.6713	-3.9959	-4.2696	-3.9909	-4.2693
	[0.0068]	[0.4641]	[0.1310]	[0.0190]	[0.1345]	[0.0160]
VaR 5%	-2.9826	-2.6181	-2.8163	-2.9565	-2.8227	-2.9662
	[0.0021]	[0.1378]	[0.0324]	[0.0086]	[0.0294]	[0.0037]

* True value for parameters, MC mean value for VaRs.

AR: acceptance rate for the independent MH.

CP: Censored posterior.

PCP: Partially censored posterior.

(P)CP0: Censoring with threshold 0.

(P)CP10%: Censoring with threshold 10% sample quantile.

Table 5: Asymmetric (misspecified) AR(1) zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: simulation results for standard posterior, censored posterior and partially censored posterior (the latter two with two threshold values). For the censored and the partially censored posterior the focus is on the left tail. Standard errors in parentheses, MSEs in brackets.

3.1.3 $H = 100$ out-of-sample forecasts and the Diebold-Mariano test

Third, to control for the simulation uncertainty, we extend the analysis to 100-step-ahead forecasts, which we repeat in 100 MC experiments. This also allows us to quantify the relative performance between the methods via the Diebold-Mariano test, as discussed below.

Tables 7 and 7 confirm our previous findings that in terms of the parameter estimates the PCP inherits the benefits from both methods, the regular posterior and the full CP. All the estimates of the left tail parameters delivered by the PCP are closer to their true values than these from the standard posterior and the CP. They are also more accurate compared to the fully censored approach. In terms of the MSEs for the 1% and 5% VaR, the PCP also outperforms both benchmark approaches, which can be seen in Tables 7 and 7.

Table 8 presents the results of the Diebold-Mariano test for pairwise method comparison of forecasting performance. They are based on the loss differential vector d constructed using $S = 100$ simulated time series, with the loss function defined as the RMSE over $H = 100$ out-of-sample periods, with the in-sample periods set equal to $T = 100$, $1,000$ and $10,000$. The RMSEs are calculated against the theoretical VaRs. Since the MC experiments are independent, the elements of d are i.i.d., hence no Newey-West-type correction is needed and the test statistics for methods 1 and 2 can be computed as

$$d_{1,2} = \mathrm{RMSE}_1 - \mathrm{RMSE}_2,$$
$$\mathrm{DM}_{1,2} = \frac{\overline{\mathrm{DM}}_{1,2}}{\mathrm{SD}(\mathrm{DM}_{1,2})/\sqrt{S}},$$

where SD denotes the sample standard deviation. The values below and above the diagonal correspond to the 99% and 95% VaR, respectively. A negative number indicates that the corresponding row method provides a better quantile evaluation than the corresponding column method. First, it can be seen that all the censoring-based methods easily outperform the regular posterior, for all the sample sizes and both quantiles (in each block the values below and above the diagonal correspond to the 99% and 95% VaR, respectively). Second, the PCP with threshold equal to 0 (PCP0) significantly outperform its fully censored counterpart. Also for the 5% quantile the PCP based on the 10% percentile of the sample (PCP10%) works significantly better than the CP with the same threshold (only for $T = 100$ there is no significant result, which is understandable given the small sample size and the restrictive censoring).

Value	True/MC*	Posterior	CP0	CP10%	PCP0	PCP10%
			$T = 100$			
AR	–	0.8650	0.8083	0.7435	0.7785	0.5362
μ	0.0000	-0.0091	0.0499	0.4937	0.0268	0.3330
		(0.1001)	(0.1398)	(0.5254)	(0.1350)	(0.4487)
σ	1.0000	0.9939	1.0227	1.2242	1.0108	1.1700
		(0.0717)	(0.1111)	(0.3196)	(0.1075)	(0.2895)
ϕ	0.8000	0.7869	0.8146	0.8547	0.7863	0.7863
		(0.0610)	(0.0783)	(0.1013)	(0.0606)	(0.0606)
VaR 1%	-2.3376	-2.3761	-2.4013	-2.5025	-2.3875	-2.4932
	[0.0014]	[0.0461]	[0.0753]	[0.1694]	[0.0698]	[0.1609]
VaR 5%	-1.6547	-1.6733	-1.6677	-1.6066	-1.6651	-1.6379
	[0.0004]	[0.0299]	[0.0431]	[0.0670]	[0.0394]	[0.0570]
			$T = 1000$			
AR	–	0.8799	0.8778	0.8429	0.7267	0.6912
μ	0.0000	-0.0040	0.0077	0.1242	0.0016	0.0299
		(0.0316)	(0.0463)	(0.1916)	(0.0426)	(0.1075)
σ	1.0000	0.9983	1.0047	1.0445	1.0016	1.0191
		(0.0224)	(0.0334)	(0.0816)	(0.0322)	(0.0664)
ϕ	0.8000	0.7963	0.8034	0.8327	0.7964	0.7964
		(0.0191)	(0.0293)	(0.0654)	(0.0191)	(0.0191)
VaR 1%	-2.3462	-2.3526	-2.3572	-2.3554	-2.3552	-2.3708
	[0.0014]	[0.0060]	[0.0083]	[0.0256]	[0.0070]	[0.0144]
VaR 5%	-1.6637	-1.6682	-1.6682	-1.6363	-1.6680	-1.6703
	[0.0004]	[0.0038]	[0.0055]	[0.0222]	[0.0042]	[0.0068]
			$T = 10000$			
AR	–	0.8522	0.8101	0.7804	0.7084	0.6579
μ	0.0000	0.0012	0.0033	0.0322	0.0014	0.0099
		(0.0099)	(0.0141)	(0.0605)	(0.0140)	(0.0328)
σ	1.0000	0.9997	0.9999	1.0100	0.9993	1.0042
		(0.0070)	(0.0102)	(0.0246)	(0.0105)	(0.0206)
ϕ	0.8000	0.7998	0.8013	0.8084	0.7999	0.7999
		(0.0059)	(0.0088)	(0.0217)	(0.0059)	(0.0059)
VaR 1%	-2.3291	-2.3290	-2.3268	-2.3251	-2.3277	-2.3310
	[0.0014]	[0.0018]	[0.0023]	[0.0046]	[0.0021]	[0.0027]
VaR 5%	-1.6470	-1.6457	-1.6433	-1.6351	-1.6447	-1.6452
	[0.0005]	[0.0008]	[0.0012]	[0.0038]	[0.0009]	[0.0013]

* True value for parameters, MC mean value for the mean (over out-of-sample horizon) VaRs.

AR: acceptance rate for the independent MH (M = 10,000, burn in of 1,000).

CP: Censored posterior.

PCP: Partially censored posterior.

(P)CP0: Censoring with threshold 0.

(P)CP10%: Censoring with threshold 10% sample quantile.

Table 6: Symmetric (correctly specified) AR(1) zero mean split normal model with $\sigma_1 = 1$ **and** $\sigma_2 = 1$: simulation results for standard posterior, censored posterior and partially censored posterior (the latter two with two threshold values) for . For the censored and the partially censored posterior the focus is on the left tail. All results averaged 100 MC replications. Results for the VaRs additionally averaged over out-of-sample horizon of $H = 100$. (Mean) standard errors in parentheses, (Mean) MSEs in brackets.

Value	True/MC*	Posterior	CP0	CP10%	PCP0	PCP10%
			$T = 100$			
AR	–	0.8662	0.8036	0.7529	0.7805	0.5448
μ	0.3989	-0.0127	0.2195	1.1094	0.1761	0.8366
		(0.1552)	(0.2549)	(0.9804)	(0.2462)	(0.8485)
σ	2.0000	1.5228	1.7953	2.3110	1.7651	2.1985
		(0.1100)	(0.2045)	(0.6008)	(0.1974)	(0.5421)
ϕ	0.8000	0.7861	0.8275	0.8550	0.7860	0.7860
		(0.0609)	(0.0782)	(0.1011)	(0.0602)	(0.0602)
VaR 1%	-4.2606	-3.6252	-4.0605	-4.4759	-4.0204	-4.4368
	[0.0057]	[0.6169]	[0.2386]	[0.4906]	[0.2674]	[0.4773]
VaR 5%	-2.8947	-2.5484	-2.7748	-2.8027	-2.7585	-2.8412
	[0.0018]	[0.2688]	[0.1487]	[0.2133]	[0.1656]	[0.2242]
			$T = 1000$			
AR	–	0.8793	0.8805	0.8471	0.7423	0.7264
μ	0.3989	-0.0039	0.1346	0.3876	0.1069	0.2676
		(0.0484)	(0.0805)	(0.3291)	(0.0737)	(0.1993)
σ	2.0000	1.5305	1.7377	1.9471	1.7228	1.9094
		(0.0343)	(0.0594)	(0.1523)	(0.0565)	(0.1237)
ϕ	0.8000	0.8008	0.8223	0.8298	0.8008	0.8008
		(0.0190)	(0.0306)	(0.0657)	(0.0189)	(0.0189)
VaR 1%	-4.2637	-3.5804	-3.9271	-4.1864	-3.9182	-4.1977
	[0.0057]	[0.4824]	[0.1402]	[0.0800]	[0.1398]	[0.0481]
VaR 5%	-2.8985	-2.5318	-2.7355	-2.8499	-2.7368	-2.8856
	[0.0018]	[0.1458]	[0.0453]	[0.0617]	[0.0388]	[0.0214]
			$T = 10000$			
AR	–	0.8372	0.8213	0.7974	0.7074	0.6633
μ	0.3989	-0.0015	0.1201	0.2738	0.1014	0.2528
		(0.0152)	(0.0246)	(0.1062)	(0.0231)	(0.0632)
σ	2.0000	1.5304	1.7302	1.9112	1.7201	1.9085
		(0.0106)	(0.0180)	(0.0470)	(0.0180)	(0.0390)
ϕ	0.8000	0.8002	0.8148	0.8048	0.8001	0.8001
		(0.0060)	(0.0094)	(0.0223)	(0.0060)	(0.0060)
VaR 1%	-4.2375	-3.5435	-3.8877	-4.1634	-3.8830	-4.1707
	[0.0057]	[0.4816]	[0.1276]	[0.0198]	[0.1288]	[0.0184]
VaR 5%	-2.8713	-2.4981	-2.7053	-2.8578	-2.7084	-2.8663
	[0.0018]	[0.1412]	[0.0317]	[0.0107]	[0.0288]	[0.0058]

* True value for parameters, MC mean value for the mean (over out-of-sample horizon) VaRs.

AR: acceptance rate for the independent MH (M = 10,000, burn in of 1,000).

CP: Censored posterior.

PCP: Partially censored posterior.

(P)CP0: Censoring with threshold 0.

(P)CP10%: Censoring with threshold 10% sample quantile.

Table 7: Asymmetric (misspecified) AR(1) zero mean split normal model with $\sigma_1 = 1$ **and** $\sigma_2 = 2$: simulation results for standard posterior, censored posterior and partially censored posterior (the latter two with two threshold values) . For the censored and the partially censored posterior the focus is on the left tail. All results averaged 100 MC replications. Results for the VaRs additionally averaged over out-of-sample horizon of $H = 100$. (Mean) standard errors in parentheses, (Mean) MSEs in brackets.

VaR 1% \ 5%	True	Posterior	CP0	PCP0	CP10%	PCP10%
			T=100			
True	–	-15.6304***	-12.9805***	-12.5644***	-13.5620***	-12.6605***
Posterior	17.4067***	–	2.2375**	2.1633**	6.6691***	6.8225***
CP0	12.3550***	-2.1602**	–	-0.3636	2.4007**	1.8939*
PCP0	11.8415***	-2.4400**	-1.3660	–	2.8128***	2.5647**
CP10%	12.6564***	-9.4565***	-3.8159***	-3.6316***	–	-1.2083
PCP10%	12.3414***	-9.4418***	-3.2849***	-3.1230***	2.1410**	–
			T=1000			
True	–	-34.9567***	-13.9906***	-11.1712***	-17.2601***	-15.7161***
Posterior	50.6764***	–	10.9280***	22.1073***	39.2825***	60.1055***
CP0	13.8532***	-28.7300***	–	8.3932***	1.5407	2.7921***
PCP0	10.5767***	-36.8884***	-7.7151***	–	-6.5617***	-5.1664***
CP10%	22.9236***	-66.7436***	6.9451***	12.6067***	–	5.0673***
PCP10%	22.3444***	-70.3424***	6.7325***	12.3974***	-0.5778	–
			T=10000			
True	–	-97.1968***	-10.9592***	-5.6158***	-37.9503***	-40.2723***
Posterior	143.4896***	–	48.7821***	59.0816***	99.7489***	142.1441***
CP0	11.7504***	-102.3047***	–	6.2508***	-14.3701***	-13.3585***
PCP0	8.2189***	-91.2430***	-2.1248**	–	-20.9393***	-20.6374***
CP10%	63.6761***	-134.8305***	40.8060***	36.9795***	–	4.8930***
PCP10%	73.4391***	-157.0898***	44.2324***	40.3268***	1.2348	–

True: simulation from the correctly specifie model.

CP: Censored posterior.

PCP: Partially censored posterior.

(P)CP0: Censoring with threshold 0.

(P)CP10%: Censoring with threshold 10% sample quantile.

Table 8: Asymmetric (misspecified) AR(1) zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: Diebold-Mariano test statistics for pairwise method comparison of forecasting performance based on loss differential vectors of length $S = 100$, with the loss function defined as the RMSE over $H = 100$ out-of-sample periods. In each block the values below and above the diagonal correspond to the 99% and 95% VaR, respectively. A negative number indicates that the corresponding row method provides a better quantile evaluation than the corresponding column method. Significance: *, **, *** at 10%, 5%, 1% level, respectively.

3.2 Formulation and algorithm

More precisely, we propose a novel estimation method based on combining the standard posterior for the "common" parameters and the censored posterior for the parameters specifying the properties of the region of interest. Consider a parameter θ and suppose that some subset of it, call it θ_2, might benefit from censoring, while the other one, θ_1, should rather not be censored. In other words, we consider a partition $\theta = (\theta_1^T, \theta_2^T)^T$. How this partition is done depends on a model under consideration, however a sensible way is to collect in θ_1 all the parameters reflecting dynamic behaviour of the process (ARMA parameters, (G)ARCH parameters, etc.), while in θ_2 the parameters determining the shape of the conditional distribution (degrees of freedom of Student's-t distribution, shape parameter of Generalized Error Distribution, means, variances and weights in mixture distribution for innovations).

We define the partially censored posterior as

$$p^{pcp}(\theta_1, \theta_2|y) := p(\theta_1|y)p^{cs)}(\theta_2|\theta_1, y),$$

where $p(\theta_1|y)$ is the standard posterior of θ_1 and $p^{cs}(\theta_2|\theta_1, y)$ is the *conditional* censored posterior of θ_2 given θ_1. In the later conditioning can be easily obtained by noting that

$$p^{cs}(\theta_2|\theta_1, y) \propto p(\theta_1, \theta_2)p^{cs}(y|\theta_1, \theta_2)$$

where a fixed value of θ_1 is plugged in the joint censored posterior density kernel. This suggest the following simulation procedure.

1. Simulate $(\theta_1^{(i)}, \theta_2^{(i)})$, $i = 1, \ldots, M$, from posterior $p(\theta_1, \theta_2|y)$.
2. Keep $\theta_1^{(i)}$ and ignore $\theta_2^{(i)}$, $i = 1, \ldots, M$.
3. For each $\theta_1^{(i)}$ simulate $\theta_2^{(i,j)}$, $j = 1, \ldots, N$, from the joint censored posterior $p^{cs}(\theta_1, \theta_2|y)$:
   - 3.1. construct joint candidate density $q(\theta_1, \theta_2)$ that approximates the joint censored posterior (e.g. with mixtures of Student's t distributions);
   - 3.2. use conditional candidate density $q(\theta_2|\theta_1 = \theta_1^{(i)})$ as candidate density to simulate $\theta_2^{(i,j)}$ from $p^{cs}(\theta_2|\theta_1^{(i)}, y)$
     (e.g. with importance sampling or the Metropolis-Hastings (MH) algorithm).

Step 3.1. can be efficiently carried out by employing the Mixture of t by Importance Sampling weighted Expectation Maximization (MitISEM) algorithm of Hoogerheide et al. (2012). To perform conditional sampling in step 3.2. we use the fact that the conditional distribution of Student's t distribution is a Student's t distribution itself, a that this fact extends to mixtures of such distributions (the details are provided in Appendix A).

In principle for every posterior draw of $\theta_1^{(i)}$ we can run a separate chain and take a single draw of $\theta_2^{(i,j)}$ after the burn-in. This, however, might be too time consuming and is not necessary efficient, as there likely will be some repeated values of $\theta_1^{(i)}$ in the posterior sample due to using of the MH algorithm. Hence, we can use every τth draw from posterior to run the MH algorithm and then take τ draws of θ_2 corresponding to this particular value of $\theta_1^{(i)}$, so in result we have the same number of the PCP draws as the posterior draws. The lower the value of τ, the the smaller the thinning and the more diverse sample of $\theta_1^{(i)}$.

In the case when θ_2 is one-dimensional (e.g. degrees of freedom of a Student's t distribution of an error term), an alternative approach to carrying out step 3.2. is the inverse transformation method which is faster than running an MH algorithm for each posterior draw separately (cf. Appendix B for details).

4 Time Varying Threshold

Notice that the region of interest A_t used to define the censored variable in (2.2) is potentially time-varying. However, most of the literature has been limited to a time-constant threshold. Gatarek et al. (2014) set the "censoring boundary" to the 20% or 30% percentile of the estimation window, leaving the topic of a time-varying threshold for further research. Opschoor et al. (2016) focus on the 15% percentile of a two-piece Normal distribution or a certain percentile (15% or 25%) of the empirical distribution of the data. Diks et al. (2011) investigate the impact of a time-varying threshold, which, however, is understood slightly differently. These authors evaluate the forecasting methods using a rolling window scheme and set the time-varying constant equal to the empirical quantile of the observations in the relevant estimation window. Obviously, for a different time series (even the same variable but in a different time frame), a time-constant threshold implied by a certain percentile of a different data window is different.

However, a constant threshold might be suboptimal when we focus on the left tail of the conditional distribution (given past observations), in particular due to its unconditional character[7]. Thus it might appear more appropriate to consider a time-varying threshold, which is then chosen conditionally. For the GARCH-type models we thus suggest to condition the threshold on the quantile of the implied distribution. Since we want to sample from a time-varying-censored distribution, all the draws need to results from the same form of censoring, i.e. the set of observations which are censored needs to be common for all draws so set a priori to sampling. Hence, we need an objective rule for censoring, e.g. the MLE-implied conditional distributions. Consider the standard GARCH(1,1) model of the form

$$\begin{aligned}
y_t &= \mu + \sqrt{h_t}\epsilon_t, \\
h_t &= \omega(1-\alpha-\beta) + \alpha(y_{t-1}-\mu)^2 + \beta h_{t-1}, \\
\epsilon_t &\overset{i.i.d.}{\sim} \mathcal{N}(0,1),
\end{aligned}$$

and suppose we want to censor only if y_t is above q_t, i.e.

$$C_t = \{y_t | y_t > q_t\},$$

with

$$\begin{aligned}
q_t &= \hat{\mu} + \sqrt{\hat{h}_t} \cdot q, \\
\hat{h}_t &= \hat{\omega}(1-\hat{\alpha}-\hat{\beta}) + \hat{\alpha}(y_{t-1}-\hat{\mu})^2 + \hat{\beta}\hat{h}_{t-1},
\end{aligned}$$

where the variables with hats refer to the MLE implied values. We can then consider

$$q = \Phi^{-1}(0.1),$$

as a threshold, which means that we do not censor the observations that are more negative than from the MLE-implied conditional 10% quantile.

Hence, a given q (and the time series y) implies a certain censored likelihood/posterior density, which can be maximised to deliver its own MLE/MAP (Maximum A Posteriori estimate), $\hat{\theta}_q^{cs}$ which can be used as a staring point for a construction of a candidate distribution for sampling, e.g. using the MitISEM algorithm. Obviously, with $q = \Phi^{-1}(1)$ the censored function should boil down to the regular likelihood/posterior, with $\hat{\theta}_q^{cs} = \hat{\theta}$. It shows up that there are no (big) differences in the corresponding

[7]Even if the interest is in the unconditional left tail, so only in the most negative returns, then the time-varying threshold might be still more advantageous than the time-constant one. This is simply because the time-varying threshold provides more information about the left tail of the distribution of the standardized errors compared to the time-constant one. Only if the returns (instead of the standardized errors) are generated in a split way, then the time-constant threshold may be optimal.

MAP (Maximum A Posteriori) Estimates for a time-varying threshold CP as a function of quantile (threshold) for censoring

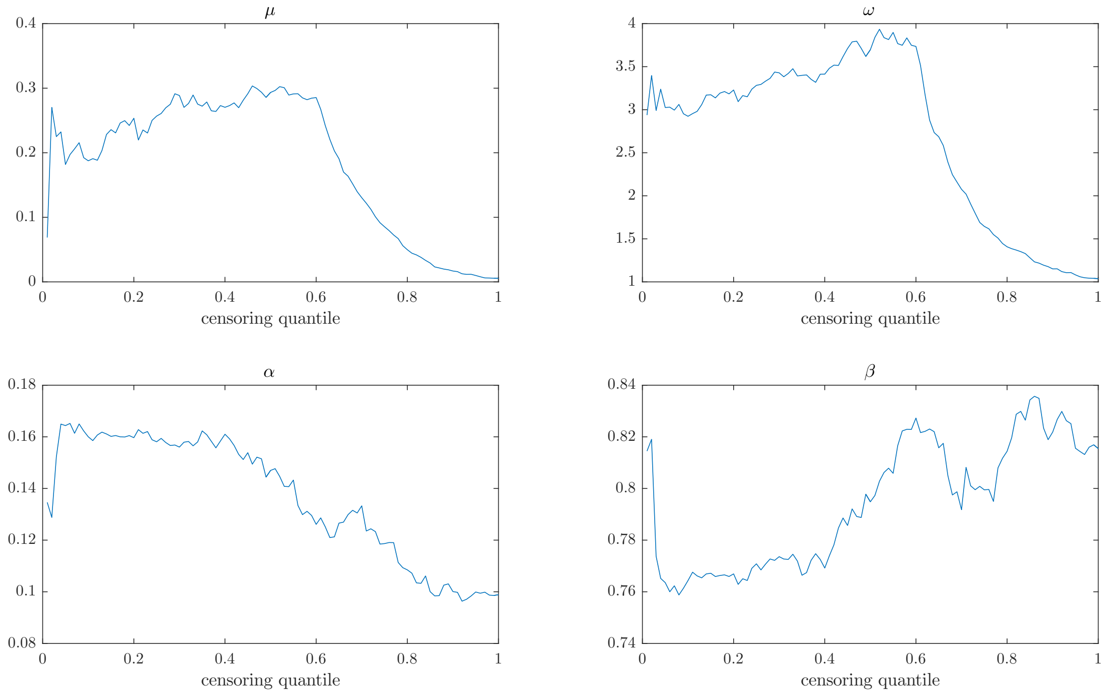


Figure 4.1: $\hat{\theta}_q^{cs}$ for 100 different quantiles q for the MLE implied time varying censored likelihood.

$\hat{\theta}_q^{cs}$ for quantiles roughly below 60%, as it can be seen in Figure 4.1. Above this quantile we can see that the estimates "converge" to the uncensored MAP. This is of interest for choosing a preferred threshold for time-varying censoring.

A potential drawback of performing MLE-based conditioning is that under misspecification the MLE would be biased. Moreover, as noticed by Gatarek et al. (2014), another potential disadvantage of censoring based on a model-implied percentile of the conditional distribution of y_t is that the choice of a particular model for censoring may affect the performance of different models in terms of forecasting. Hence, we also investigate an alternative, simplistic way of imposing time-varying threshold (which we refer to as an "ad hoc") is simply to define

$$C_t = \{y_t | y_t > y_{t-1}\},$$

which roughly should correspond to censoring of all the right part of the unconditional distribution.

4.1 AR(1) with time varying threshold

We again start with the simple AR(1) model, with $\mu = 0$ and $\rho = 0.8$. Suppose that $y_{t-1} = 1$ and $y_t = 0.5$ so $\mathbb{E}(y_t | y_{t-1}, \theta) = 0.8$, hence the time t realisation is below the conditional mean. If we censor unconditionally, with a threshold of e.g. 0 (or lower), then we do not perform censoring at time t, even though from time $t-1$ respective the current value y_t is *relatively* bad. Following the reasoning above, since

$$\frac{y_t - \mu - \rho y_{t-1}}{\sigma} \sim \mathcal{N}(0, 1),$$

we want to censor only if y_t is above q_t, i.e.

$$C_t = \{y_t | y_t > q_t\},$$

		μ	σ	ρ
True		0.3989	2.000	0.8000
Posterior	$T = 1000$	0.0019	1.5230	0.7974
	$T = 100$	-0.0371	1.4995	0.7571
CP10%	$T = 1000$	0.1605	1.8482	0.7807
	$T = 100$	-0.2061	1.6734	0.6503
CP0	$T = 1000$	0.1219	1.7175	0.8115
	$T = 100$	0.0679	1.6660	0.7655
CP var ah*	$T = 1000$	0.3327	1.9239	0.7850
	$T = 100$	0.2450	1.8366	0.7242
CP var mle	$T = 1000$	0.3675	1.9627	0.7932
	$T = 100$	0.3494	1.9085	0.7391

CP0: Censoring with threshold 0.

CP10%: Censoring with threshold 10% sample quantile.

CP var ah: Time Varying Censoring, ad hoc method, with $C_t = \{y_t | y_t > \bar{y} + y_{t-1}\}$.

CP var mle: Time Varying Censoring, MLE based method.

Table 9: AR(1) split normal $\sigma_2 = 2$, for $T = 100$ and $T = 1000$. Maxima of the standard posterior (MAPs) and the censored posterior densities with different thresholds, averaged over 100 simulations.

with

$$q_t = \hat{\mu} + \hat{\rho} y_{t-1} + \hat{\sigma} q,$$

where the variables with hats denote the MLE. Table compares the maximisation results for objective functions specified as the regular posterior and the CP with different thresholds, averaged over 100 replications. For robustness we consider two sample sizes, $T = 100$ and $T = 1000$. Not surprisingly, all the censored methods deliver the values much closer to the true ones. Within the censored methods, one can see these with time varying threshold perform much better than the time-constant threshold ones. Moreover, the MLE-based threshold seems to be more robust to the ad hoc one in small samples.

Table 10 compares the performance of the time-varying threshold methods with that of the time-constant threshold ones. In terms of VaR forecasting clearly the time-varying MLE-based PCP works the best, easily outperforming the regular posterior and improving upon the remaining censoring-based methods. As far as the parameter estimates are concerned, as expected the PCP approach increase their precision compared to the full CP.

Value	True/MC*	Posterior	CP0	PCP0	CP10%	PCP10%	CP var ah	PCP var ah	CP var mle	PCP var mle
					$T = 1000$					
AR	–	0.8764	0.8793	0.6827	0.8573	0.6436	0.8777	0.7578	0.8709	0.7675
μ	0.3989	0.0042	0.1387	0.1142	0.2536	0.2621	0.3524	0.3411	0.4027	0.3876
		(0.0483)	(0.0804)	(0.0739)	(0.3319)	(0.1993)	(0.0837)	(0.0820)	(0.2316)	(0.2290)
σ	2.0000	1.5282	1.7321	1.7188	1.9026	1.9057	1.9364	1.9287	1.9888	1.9800
		(0.0343)	(0.0596)	(0.0566)	(0.1517)	(0.1240)	(0.0738)	(0.0727)	(0.1584)	(0.1566)
ϕ	0.8000	0.7939	0.8134	0.7939	0.7911	0.7939	0.7763	0.7927	0.7868	0.7927
		(0.0192)	(0.0310)	(0.0192)	(0.0696)	(0.0192)	(0.0295)	(0.0192)	(0.0402)	(0.0192)
VaR 1%	-4.3149	-3.6159	-3.9615	-3.9496	-4.2684	-4.2449	-4.1373	-4.1266	-4.2117	-4.1990
	[0.0056]	[0.5010]	[0.1482]	[0.1517]	[0.0700]	[0.0485]	[0.0576]	[0.0511]	[0.0501]	[0.0379]
VaR 5%	-2.9500	-2.5690	-2.7724	-2.7719	-2.9612	-2.9357	-2.8076	-2.8037	-2.8438	-2.8401
	[0.0018]	[0.1545]	[0.0466]	[0.0424]	[0.0553]	[0.0207]	[0.0308]	[0.0235]	[0.0272]	[0.0163]

* True value for parameters, MC mean value for the mean (over out-of-sample horizon) VaRs.

AR: acceptance rate for the independent MH (M = 10,000, burn in of 1,000).

CP: Censored posterior.

PCP: Partially censored posterior.

(P)CP0: Censoring with threshold 0.

(P)CP10%: Censoring with threshold 10% sample quantile.

(P)CP var ah: Time Varying Censoring, ad hoc method.

(P)CP var mle: Time Varying Censoring, MLE based method.

Table 10: Simulation results for standard posterior, censored posterior and partially censored posterior (the latter two with two time-constant and two time-varying thresholds) for the ar1 zero mean split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$. For the censored and the partially censored posterior the focus is on the left tail. All results averaged 100 MC replications. Results for the VaRs additionally averaged over out-of-sample horizon of $H = 100$. (Mean) standard errors in parentheses, (Mean) MSEs in brackets.

4.2 A-ARCH(1) with time varying threshold

In the standard GARCH model there is no finite set of parameters to control the exact shape of the tail because the time-varying conditional distribution is infinitely dimensional, $y_t|y_{t-1} \sim p_t$, with p_t standing for a distribution with mean 0 and variance $\sqrt{h_t}$. This is different than in the AR(1) model example where single parameter σ can characterise the tail.

Therefore, for illustrative purposes we consider a highly synthetic example. We build upon the AR(1) model with the aim to perfectly fit the shape of the tail of the conditional distribution in a GARCH-type model. We consider the Asymmetric ARCH(1) model (A-ARCH(1)), in spirit of the asymmetric GARCH of Engle and Ng (1993), which is characterised by two means, one for the observation process and another one for the volatility process. By removing the autoregressive component of the volatility from the GARCH model, we eliminate the problem of "self-feeding" of the volatility so that the shape of the tail can be controlled for. As the DGP we consider the ARCH(1) with split normal distribution for the error term.

$$
\begin{aligned}
y_t &= \sqrt{(\kappa^{-1}h_t)}\varepsilon_t,\\
h_t &= \omega(1-\alpha) + \alpha y_{t-1}^2,\\
\varepsilon_t &\sim \mathcal{SN}(\mu, \sigma_1^2, \sigma_2^2),\\
\mu &:= \frac{\sigma_2 - \sigma_1}{\sqrt{2\pi}},\\
\kappa &:= \frac{1}{2}\left(\left(\sigma_1^2 + \sigma_2^2\right) - \frac{(\sigma_2 - \sigma_1)^2}{\pi}\right),
\end{aligned}
$$

κ – the variance of ε_t (for the stationarity of the volatility process). We estimate a misspecified model given by

$$
\begin{aligned}
y_t &= \mu_1 + \sqrt{h_t}\epsilon_t,\\
h_t &= \omega(1-\alpha) + \alpha(y_{t-1} - \mu_2)^2,\\
\epsilon_t &\sim \mathcal{N}(0,1).
\end{aligned}
$$

We set $\omega = 1$, $\alpha = 0.1$, $\beta = 0.8$, and for ε_t we again choose $\sigma_1 = 1$ and $\sigma_2 = 2$. Then $\kappa \approx 2.34$, which effectively implies the standard deviation of the right and the left tail of around 0.65 and 1.3, respectively. We consider 1% and 5% one-step-ahead VaR forecasting over a horizon of 100 days (50 MC replications).

Table 11 present the results. For the 1% quantile every censored method outperforms the regular posterior, and partial censoring outperforms regular censoring. With the 5% quantile it gets harder for the censored methods to provide forecasts much better than the regular posterior. This is because the normal distribution used to generate the data is light-tailed, so that the 5% quantile is not really hard to capture (it can be seen how greatly the regular posterior improves for 5% vs 1%). But the censored methods are not far off, with the *ad hoc* PCP clearly outperforming the regular posterior. Finally, all PCP methods are better than their CP counterparts.

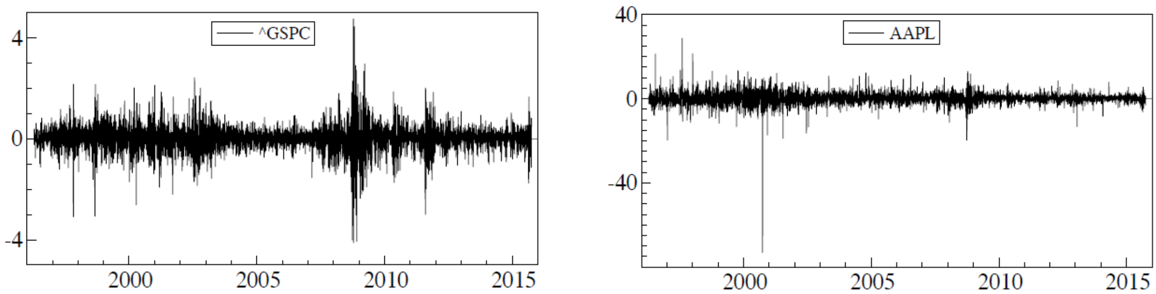
5 Empirical Application of GARCH-t and GAS-t models to S&P500 and Apple Returns

In this section we compare the forecasting performance of the regular posterior with the partially censored one. We consider daily logreturns of the S&P 500 Index (ˆGSPC) and of the stock of Apple Inc. (AAPL).

	1%		5%	
Method	$T = 1000$	$T = 2500$	$T = 1000$	$T = 2500$
Posterior	0.1035	0.1013	0.0188	0.0148
CP0	0.0689	0.0428	0.0432	0.0310
PCP0	0.0483	0.0383	0.0332	0.0284
CP10%	0.0916	0.0531	0.0443	0.0318
PCP10%	0.0590	0.0443	0.0334	0.0289
CP var ah	0.0477	0.0325	0.0288	0.0204
PCP var ah	0.0252	0.0164	0.0142	0.0084
CP var mle	0.0960	0.0492	0.0445	0.0886
PCP var mle	0.0705	0.0407	0.0341	0.0270

Table 11: A-ARCH(1) split normal $\sigma_2 = 2$ model: MSEs for $M = 50$ replications of one-step-ahead forecasting of 1% and 5% VaR over $H = 100$ days.

The former represents the performance of the broader market, while the latter is one of the most heavily traded stocks. We consider period from April 15, 1996 to October 5, 2015 (4903 observations, Figure 5.1).



(a) S&P 500

(b) Apple Inc.

Figure 5.1: The daily logreturns from April 15, 1996 to October 5, 2015 of (a) S&P 500, (b) Apple Inc.

We analyse two benchmark models of volatility, commonly employed by practitioners, the Generalized Autoregressive Conditional Heteroscedasticity model (GARCH, Engle, 1982; Bollerslev, 1986) and the Generalised Autoregressive Score model (GAS, Creal et al., 2013), both with Student's t innovations. The specification for the former is as follows

$$\begin{aligned} y_t &= \mu + \sqrt{\rho h_t}\varepsilon_t, \\ \varepsilon_t &\sim t(\nu), \\ \rho &:= \frac{\nu - 2}{\nu}, \\ h_t &= \omega + \alpha y_{t-1}^2 + \beta h_{t-1}, \end{aligned}$$

GARCH-t			GAS-t		
Parameter	AAPL	ˆGSPC	Parameter	AAPL	ˆGSPC
ω	0.030	0.003	ω	0.030	0.003
	(0.011)	(0.001)		(0.011)	(0.001)
β	0.959	0.901	B	0.997	0.988
	(0.009)	(0.010)		(0.002)	(0.004)
α	0.037	0.087	A	0.044	0.083
	(0.008)	(0.009)		(0.007)	(0.008)
ν	4.920	8.041	ν	5.114	8.573
	(0.300)	(0.865)		(0.332)	(1.045)
ν^{pcp}	4.873	5.256	ν^{pcp}	5.069	5.354
	(0.304)	(0.376)		(0.350)	(0.399)

Table 12: Regular posterior estimates and the PCP estimates for the degrees of freedom for the GARCH(1,1)-t and GAS(1,1)-t for the daily log returns of S&P500 and Apple Inc. (standard errors in parentheses).

while for the latter

$$\begin{aligned}
y_t &= \mu + \sqrt{\rho h_t}\varepsilon_t,\\
\varepsilon_t &\sim t(\nu),\\
\rho &:= \frac{\nu-2}{\nu},\\
h_t &= \omega + A\frac{\nu+3}{\nu}\Big(C_{t-1}(y_t-\mu)^2 - h_{t-1}\Big) + Bh_{t-1},\\
C_t &= \frac{\nu+1}{\nu-2}\left(1+\frac{(y_{t-1}-\mu)^2}{(\nu-2)h_{t-1}}\right)^{-1}.
\end{aligned}$$

In both cases we put flat priors and impose the standard variance positivity and stationarity restrictions (i.e. $\omega > 0$, $\alpha \in (0,1)$, $\beta \in (0,1)$ with $\alpha+\beta < 1$ for GARCH, $\omega > 0$ and $B \in (0,1)$ for GAS), except for the degrees of freedom, for which we set an uninformative yet proper prior $\nu - 2 \sim \text{Exp}(0.01)$.

As a benchmark and the starting point for the PCP approach, we first carry out the standard posterior analysis for both models. We run $M = 10,000$ iterations (after a burn-in of $1,000$) of the independent Metropolis-Hastings using as a candidate the Student's-t distribution with the mode set to the MLE, the scale matrix set to minus the inverse of the Hessian matrix of the log-posterior at the mode and with 4 degrees of freedom for the robustness of the sampler. Next, given the posterior draws of the dynamics parameters, we perform conditional sampling of the degrees of freedom using the inverse transform method as discussed in Appendix B (on a grid of $N = 180$ points, $2.1, 2.2, \ldots, 20.0$).

Table 12 presents the estimation results for both models obtained with the regular posterior and for the degrees of freedom additionally with the PCP approach. We observe that all the PCP estimates for ν turn out to be lower than those from the regular posterior. We further investigate the impact of these differences on the forecasting performance.

In our forecasting study we adopt a rolling window approach: for the datasets of 4903 observations all the rolling windows consist of 2903 observations, corresponding to $H = 2000$ out-of-sample density forecasts. We estimate our models based on the first 2903 observations in-sample for each series, i.e. until October 2007. We forecast the densities until the end of our complete sample. Hence our out-of-sample size is equal to 2000 in this case. Again, we re-estimate our parameters annually to perform the density forecasts, resulting in a rolling window optimization with eight different windows.

We compare the density forecasting performance between the PCP and the regular posterior, where the

	AAPL	ˆGSPC
Model	CSR	CSR
GARCH-t	-1.377	-2.478**
GAS-t	-1.740*	-3.058***

Table 13: Empirical study: Diebold-Mariano test statistics for pairwise method comparison of forecasting performance based on the average loss differential of the KLIC over $H = 2000$ out-of-sample density forecasts. The loss differential is for the comparison of the PCP approach versus the regular (uncensored) approach, where we assess the quality using the censored likelihood of Diks et al. (2011) (CSR). A negative value indicates that the PCP performs better than the regular posterior. Significance: *, **, *** at 10%, 5%, 1% level, respectively.

comparison is based on the censored likelihood scoring rule(CSR) of Diks et al. (2011), where we set the threshold to the 30% percentile. Table 13 presents the results of the DM test and the plots of the loss differentials used in the KLIC-based DM tests are provided in Figure D.1 in Appendix D. We record that for both time series and both volatility models, the PCP outperforms the regular posterior, providing better density forecasts in all cases.

6 Conclusions

We have proposed a novel approach to inference for a specific region of interest of the predictive distribution. Our Partially Censored Posterior method falls outside the framework of the purely Bayesian statistics as we do not work with the regular likelihood but with the censored likelihood originating from Diks et al. (2011). This allows us to keep all the merits of the regular Bayesian analysis, e.g. dealing with the parameter uncertainty, and at the same time to make the region-focused inference robust against a potential model misspecification. The latter is vital for risk management, where the shape of the left tail of the conditional distribution is of crucial importance.

Partitioning of the parameter set into two subsets, only of of which is likely to benefit from censoring, increases the precision of the parameter estimates compared to the fully censored posterior of Gatarek et al. (2014) and allows us to obtain better tail forecast. Finally, we have developed novel ways of time-varying censoring, which are beneficial from the tail prediction perspective. We have demonstrated the usefulness of out methods in extensive simulation and empirical studies.

To further exploit the power of our quasi-Bayesian framework, in future research we intend to employ the PCP in the context of forecast combination via Bayesian Model Averaging. Also extending of the classical approach of Opschoor et al. (2016) based on the so-called pooling seems to be a relevant in this regard. Another interesting extension will be to investigate the impact of using the smooth-censored likelihood of Diks et al. (2011) in our PCP setting.

Bibliography

Amisano, G. and R. Giacomini (2007), "Comparing Density Forecasts via Weighted Likelihood Ratio Tests." *Journal of Business and Economic Statistics*, 25, 177–190.

Artzner, P., F. Delbaen, J. M. Eber, and D. Heath (1999), "Coherent Measures of Risk." *Mathematical Finance*, 9, 203–228.

Bollerslev, T. (1986), "Generalised Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 51, 307–327.

Creal, D., S. J. Koopman, and A. Lucas. (2013), "Generalized Autoregressive Score Models with Applications." *Journal of Applied Econometrics*, 28, 777–795.

Cross Validated (2017), "Why should I be Bayesian when my model is wrong?" `https://stats.stackexchange.com/questions/274815/why-should-i-be-bayesian-when-my-model-is-wrong`. Accessed: 2017-07-18.

de Roon, F. and P. Karehnke (2016), "A Simple Skewed Distribution with Asset Pricing Applications." *Review of Finance*, 1–29.

Diebold, F. X. and R. S. Mariano (1995), "Comparing Predictive Accuracy." *Journal of Business and Economic Statistic*, 13, 253–263.

Diks, C., V. Panchenko, and D. Van Dijk (2011), "Likelihood-based Scoring Rules for Comparing Density Forecasts in Tails." *Journal of Econometrics*, 163, 215–230.

Engle, R. F. (1982), "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of the United Kingdom Inflation." *Econometrica*, 50, 987–1007.

Engle, R. F. and V. K. Ng (1993), "Measuring and Testing the Impact of News on Volatility." *Journal of Finance*, 48, 1749–1778.

Gatarek, L. T., L. F. Hoogerheide, K. Hooning, and H. K. van Dijk (2014), "Censored Posterior and Predictive Likelihood in Bayesian Left-tail Prediction for Accurate Value at Risk Estimation." Technical Report TI 2013-060/III, Tinbergen Institute Discussion Paper.

Geweke, J. (1989), "Bayesian Inference in Econometric Models using Monte Carlo Integration." *Econometrica*, 57, 1317–1739.

Hoogerheide, L. F., A. Opschoor, and H. K. van Dijk (2012), "A Class of Adaptive Importance Sampling Weighted EM Algorithms for Efficient and Robust Posterior and Predictive Simulation." *Journal of Econometrics*, 171, 101–120.

Kleijn, B. J. K. and A. W. van der Vaart (2006), "Misspecification in Infinite-Dimensional Bayesian Statistics." *The Annals of Statistics*, 34, 837–877.

McNeil, A. J. and R. Frey (2000), "Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: an Extreme Value Approach." *Journal of Empirical Finance*, 7, 271–300.

McNeil, A. J., R. Frey, and P. Embrechts (2015), *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

Müller, U. K. (2013), "Misspecification in Infinite-Dimensional Bayesian Statistics." *Econometrica*, 81, 1805–1849.

Opschoor, A., D. van Dijk, and M. van der Wel (2016), "Combining Density Forecasts using Focused Scoring Rules." *Tinbergen Institute Discussion Paper*, 14-090/III.

Robert, C. (2017), "Why should I be Bayesian when my model is wrong?" `https://xianblog.wordpress.com/2017/05/09/why-should-i-be-bayesian-when-my-model-is-wrong/`. Accessed: 18 July 2017.

Roth, M. (2013), "On the Multivariate t Distribution." Technical Report LiTH-ISY-R-3059, Automatic Control Group at Linköpings Universitet.

Zellner, A. (1996), *An Introduction to Bayesian Inference in Econometrics*. Willey.

A Conditional density of (mixtures of) multivariate Student's t distributions

Let $x \in \mathbb{R}^d$ follow the Student's t distribution with mode μ, scale matrix Σ and ν degrees of freedom, denoted $t(x; \mu, \Sigma, \nu)$, where we assume $\nu > 2$ for $\mathrm{Var}[x] < \infty$. Then, the pdf of x is given by (cf. Zellner, 1996; Roth, 2013)

$$p(x) = \frac{\Gamma\left(\frac{\nu+d}{2}\right)}{\Gamma\left(\frac{d}{2}\right)(\pi\nu)^{\frac{d}{2}}} |\Sigma|^{-\frac{1}{2}} \left(1 + \frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{\nu}\right)^{-\frac{d+\nu}{2}}.$$

Conditional density Next, consider a partition of x into $x = (x_1^T, x_2^T)^T$ with x_1 and x_2 of dimensions d_1 and d_2, respectively. The corresponding parameter partitionings are then

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Then, the conditional density of x_2 given x_1 is given by

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} = t(x_2; \mu_{2|1}, \Sigma_{2|1}, \nu_{2|1}),$$

with

$$\begin{aligned}
\mu_{2|1} &= \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \\
\Sigma_{2|1} &= \frac{\nu + (x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1)}{\nu + d_1} \left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right), \\
\nu_{2|1} &= \nu + d_1.
\end{aligned}$$

Mixtures The above result extends to mixtures of Students' t distributions. Now let x follow an H component mixture of Student's t distributions $t(x; \mu_h, \Sigma_h, \nu_h)$, with component probabilities η_h, $h = 1, \ldots, H$, so that its pdf is given by

$$p(x) = \sum_{h=1}^{H} \eta_h t(x; \mu_h, \Sigma_h, \nu_h).$$

Let z denote a (latent) H-dimensional vector indicating from which component the observation x stems from: if x was generated from the hth component then $z = e_h$, the hth vector of the standard basis of $\mathbb{R}^H$, i.e. $z_h = 1$ and $z_l = 0$ for $l \neq h$. Obviously, unconditionally $\mathbb{P}(z = e_h) = \eta_h$. The conditional probability of x stemming from the hth component is

$$\begin{aligned}
\mathbb{P}[z = e_h | x] &= \frac{p(z = e_h, x)}{p(x)} \\
&= \frac{\mathbb{P}[z = e_h] p(x | z = e_h)}{\sum_{m=1}^{H} \mathbb{P}[z = e_m] p(x | z = e_m)} \\
&= \frac{\eta_h t(x; \mu_h, \Sigma_h, \nu_h)}{\sum_{m=1}^{H} \eta_m t(x; \mu_m, \Sigma_m, \nu_m)}.
\end{aligned}$$

Then, the conditional density of x_2 given x_1 is given by

$$p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{\sum_{h=1}^{H} \eta_h t(x; \mu_h, \Sigma_h, \nu_h)}{\sum_{h=1}^{H} \eta_h t(x_1; \mu_{h,1}, \Sigma_{h,1}, \nu_h)} = \sum_{h=1}^{H} \eta_{h,2|1} t(x_2; \mu_{h,2|1}, \Sigma_{h,2|1}, \nu_{h,2|1}),$$

with

$$\begin{aligned}
\mu_{h,2|1} &= \mu_{h,2} + \Sigma_{h,21}\Sigma_{h,11}^{-1}(x_1 - \mu_{h,1}), \\
\Sigma_{h,2|1} &= \frac{\nu_h + (x_1 - \mu_{h,1})^T \Sigma_{h,11}^{-1}(x_1 - \mu_{h,1})}{\nu_h + d_1} \left(\Sigma_{h,22} - \Sigma_{h,21}\Sigma_{h,h,11}^{-1}\Sigma_{h,12} \right), \\
\nu_{h,2|1} &= \nu_h + d_1,
\end{aligned}$$

and with adjusted component probabilities

$$\eta_{h,2|1} := \mathbb{P}[z = e_h|x] = \frac{\eta_h t(x_1; \mu_{h,1}, \Sigma_{h,11}, \nu_h)}{\sum_{m=1}^{H} \eta_m t(x_1; \mu_{m,1}, \Sigma_{m,11}, \nu_m)}.$$

B The inverse transform method for conditional simulation

In the notation from Section 3.2, suppose that θ_2 is one-dimensional and consider its conditional posterior distribution $p(\theta_2|\theta_1, y)$ in the joint posterior distribution

$$p(\theta_1, \theta_2|y) = p(\theta_2|\theta_1, y)p(\theta_1|y).$$

It can be rewritten as

$$p(\theta_2|\theta_1, y) = \frac{p(y|\theta_1, \theta_2)p(\theta_2|\theta_1)}{p(\theta_1, y)},$$

and under the assumption of prior parameter independence further simplified to

$$p(\theta_2|\theta_1, y) \propto p(y|\theta_1, \theta_2)p(\theta_2).$$

To perform step 3.2. in the PCP algorithm, we consider a fixed grid of points $\vartheta_2^{(i,j)}$, $j = 1, \ldots, N$, given $\theta_1 = \theta_1^{(i)}$, with the grid range covering the plausible values of $\theta_2|\theta_1, y$. We then perform the following algorithm to conditionally sample θ_2 given $\theta_1 = \theta_1^{(i)}$.

1. Evaluate the conditional distribution $p(\vartheta_2^{(i,j)}|\theta_1^{(i)}, y)$ for each grid point and store the values in a vector $\Theta_2^{(i)} = \{\Theta_2^{(i,j)}\}_{j=1}^N$ with $\Theta_2^{(i,j)} = p(\vartheta_2^{(i,j)}|\theta_1^{(i)}, y)$.

2. Use the vector $\Theta_2^{(i)}$ to construct the "empirical" cdf on the grid

$$P_{pcp}^{(i,j)} \equiv P(\vartheta_2^{(i,j)}|\theta_1^{(i)}, y) = \frac{\sum_{k=1}^{j} \Theta_2^{(i,k)}}{\sum_{l=1}^{N} \Theta_2^{(i,l)}}.$$

3. Preform the inverse transformation method on the grid: draw $u \sim U(0, 1)$ and calculate

$$k = \sum_{j=1}^{N} \mathbb{1}\{P_{pcp}^{(i,j)} < u\}.$$

4. Set as the conditional draw of θ_2 either to the grid value corresponding to the index k

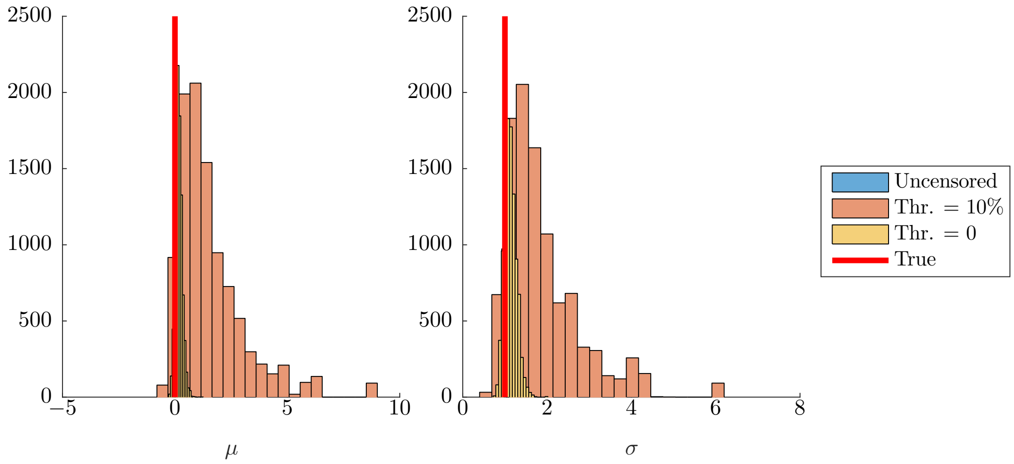
$$\theta_2^{(i,j)} = \vartheta^{(i,k)},$$

or interpolate between this value and the next one:

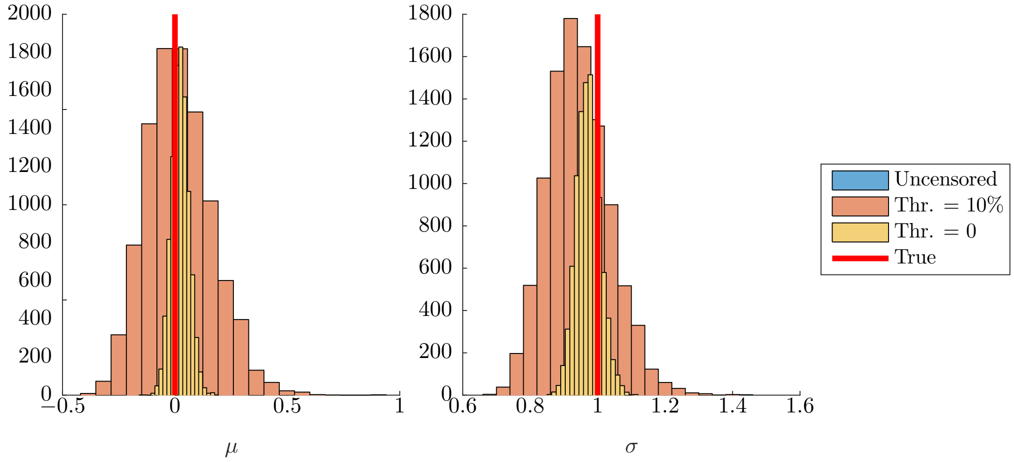
$$\theta_2^{(i,j)} = \vartheta^{(i,k)} + (\vartheta^{(i,k+1)} - \vartheta^{(i,k)}) \frac{u - P_{pcp}^{(i,k)}}{P_{pcp}^{(i,k+1)} - P_{pcp}^{(i,k)}}.$$

C Posterior draws plots

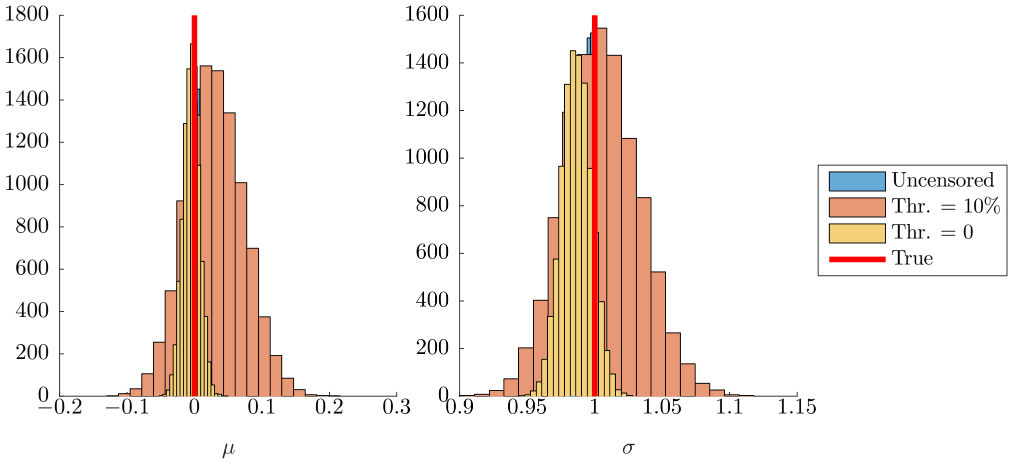
C.1 I.i.d.



(a) $T = 100$

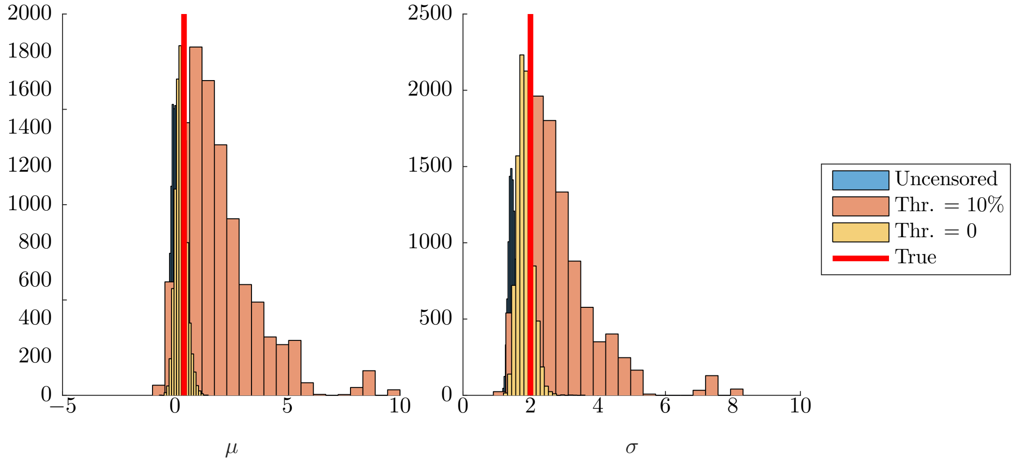


(b) $T = 1,000$

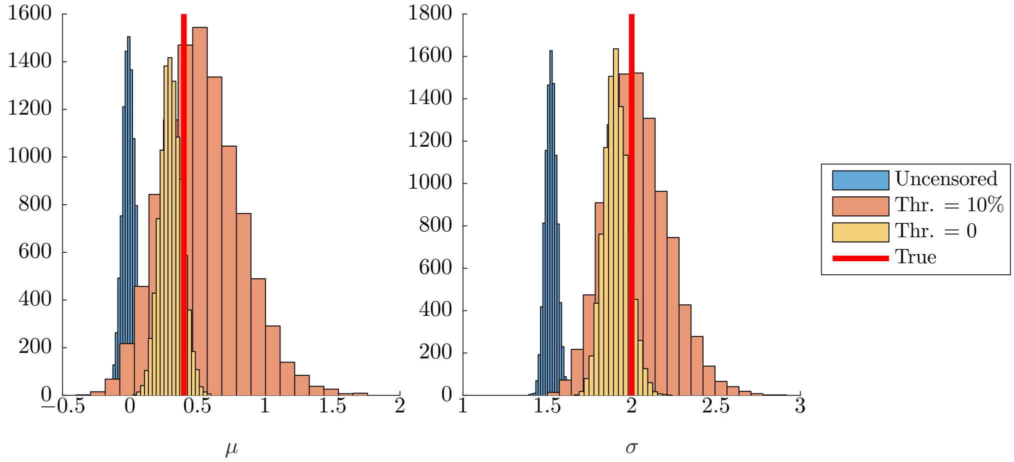


(c) $T = 10,000$

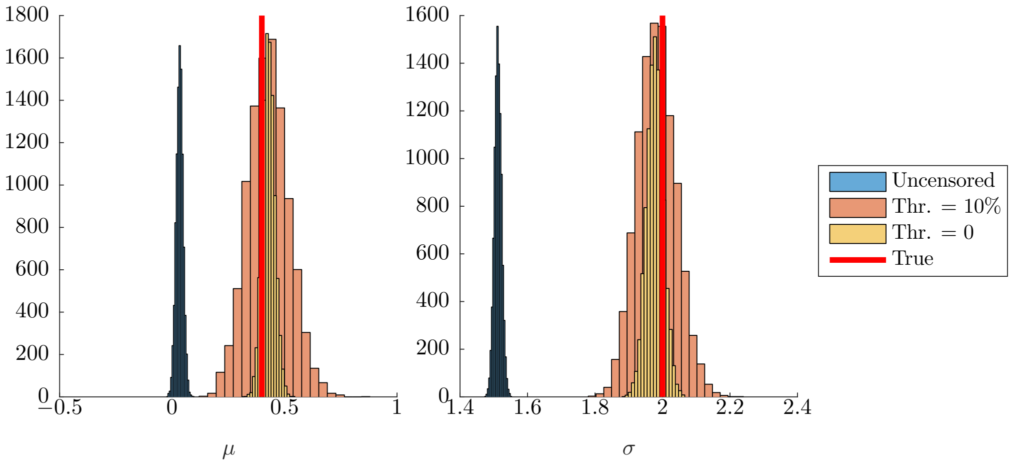
Figure C.1: Symmetric (correctly specified) i.i.d. mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 1$: histograms of posterior draws without and with censoring (with different thresholds).



(a) $T = 100$



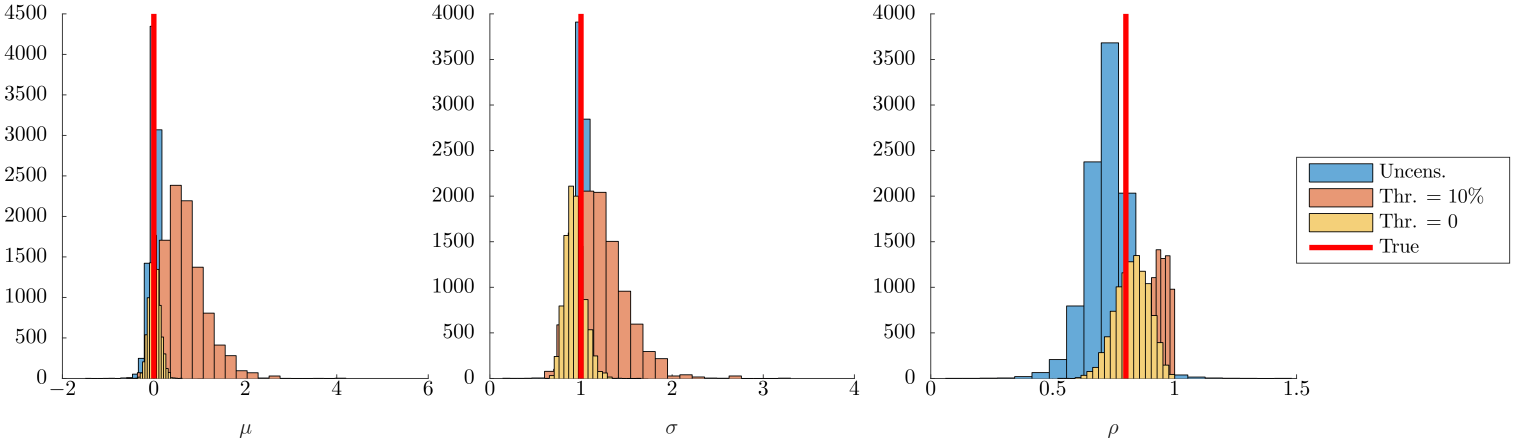
(b) $T = 1,000$



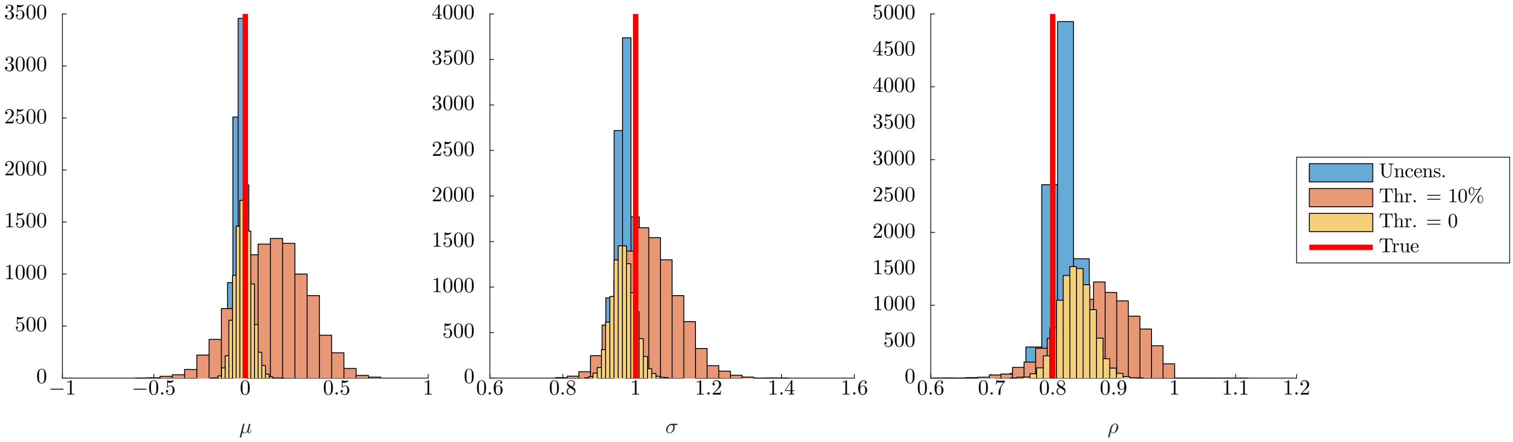
(c) $T = 10,000$

Figure C.2: Asymmetric (misspecified) i.i.d. mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: histograms of posterior draws without and with censoring (with different thresholds).

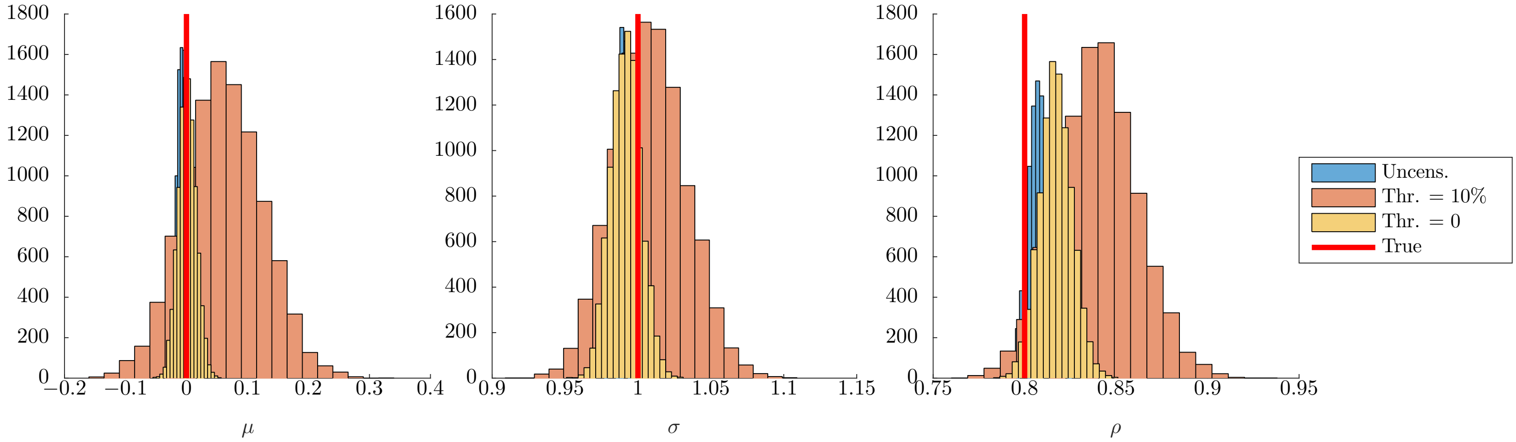
C.2 AR(1)



(a) $T = 100$

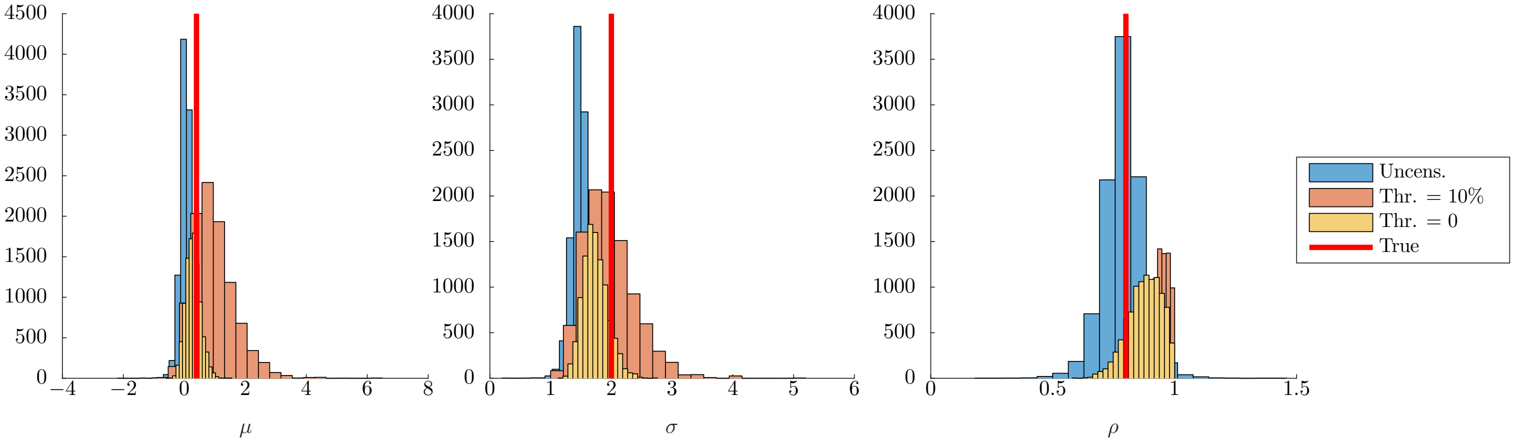


(b) $T = 1,000$

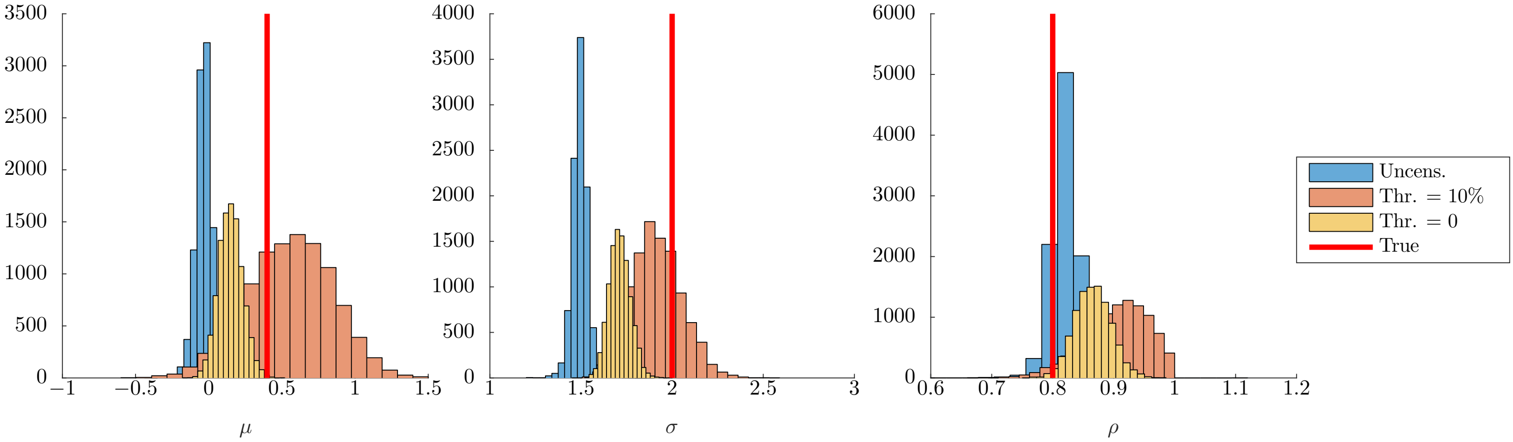


(c) $T = 10,000$

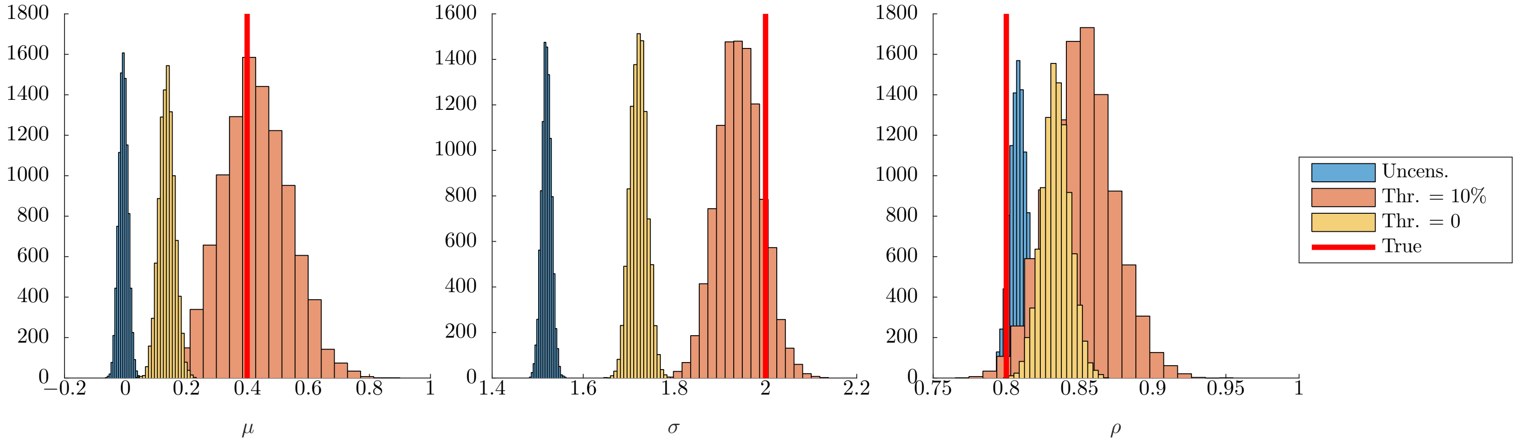
Figure C.3: Symmetric (correctly specified) AR(1) mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 1$: histograms of posterior draws without and with censoring (with different thresholds).



(a) $T = 100$



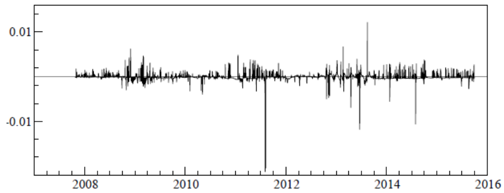
(b) $T = 1,000$



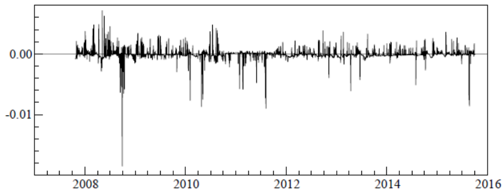
(c) $T = 10,000$

Figure C.4: Asymmetric (misspecified) AR(1) mean zero split normal model with $\sigma_1 = 1$ and $\sigma_2 = 2$: histograms of posterior draws without and with censoring (with different thresholds).

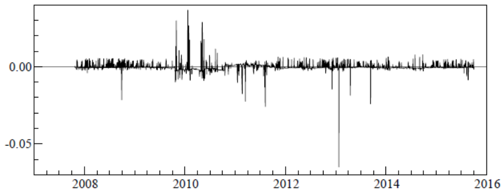
D Loss differential plots



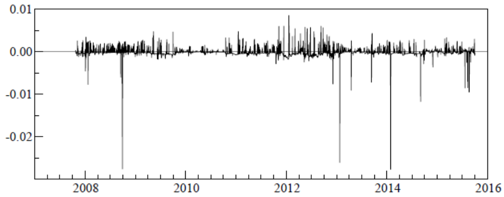
(a) S&P500, GARCH(1,1)-t.



(b) S&P500, GAS(1,1)-t.



(c) Apple Inc., GARCH(1,1)-t.



(d) Apple Inc., GAS(1,1)-t.

Figure D.1: Loss differentials, for 2000 one-step-ahead density forecasts, based on the censored likelihood (CSR).