ONLINE APPENDIX C.

Toru Kitagawa[1], José-Luis Montiel-Olea[2], Jonathan Payne[3] and Amilcar Velez[4]

1. MAX$\{\theta_1, \theta_2\}$

In this Appendix we illustrate Theorem 2 with an alternative example. Let $(X_1, \dots X_n)$ be an i.i.d sample of size n from the statistical model:

$$X_i \sim \mathcal{N}_2(\theta, \Sigma), \quad \theta = (\theta_1, \theta_2)' \in \mathbb{R}^2,\ \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

where Σ is assumed known. Consider the family of priors:

$$\theta \sim \mathcal{N}_2(\mu, (1/\lambda^2)\Sigma), \quad \mu = (\mu_1, \mu_2)' \in \mathbb{R}^2$$

indexed by the location parameter μ and the precision parameter $\lambda^2 > 0$. The object of interest is the transformation:

$$g(\theta) = \max\{\theta_1, \theta_2\}.$$

Relation to the main assumptions: The transformation g is Lipschitz continuous everywhere and differentiable everywhere except at $\theta_1 = \theta_2$ where it has directional derivative $g'_\theta(h) = \max\{h_1, h_2\}$. This implies that Assumption 1 is satisfied.

Once again, we take $\widehat{\theta}_n$ to be the maximum likelihood estimator given by $\widehat{\theta}_{\rm n} = (1/n)\sum_{i=1}^n X_i$ and so $\sqrt{n}(\widehat{\theta}_{\rm n} - \theta) \sim Z \sim \mathcal{N}_2(0, \Sigma)$. Thus, Assumption 2 is satisfied.

The posterior distribution for θ is given by Gelman, Carlin, Stern, and Rubin (2009), p. 89:

$$\theta_n^{P*} | X^n \sim \mathcal{N}_2\Big(\frac{n}{n+\lambda^2}\widehat{\theta}_n + \frac{\lambda^2}{n+\lambda^2}\mu\,,\ \frac{1}{n+\lambda^2}\Sigma\Big).$$

[1]University College London, Department of Economics, and Kyoto University, Department of Economics. E-mail: t.kitagawa@ucl.ac.uk.
[2]Columbia University, Department of Economics. E-mail: montiel.olea@gmail.com.
[3]New York University, Department of Economics. E-mail: jep459@nyu.edu.
[4]Central Bank of Peru, Macro Modelling Division. E-mail: velzs.ae@pucp.pe.

and so by an analogous argument to the absolute value example we have that:

$$\beta(\sqrt{n}(\theta_n^{P*} - \hat{\theta}_n), \mathcal{N}_2(0, \Sigma)); X^n) \xrightarrow{p} 0,$$

which implies that Assumption 3 holds.

Finally, since g is directionally differentiable, Remark 2 (and Lemma 4) imply that Assumption 4 (i) is satisfied by function:

$$\begin{aligned} h_{\theta_0}(Z, X^n) &= g'_{\theta_0}(Z + Z_n) - g'_{\theta_0}(Z_n) \\ &= \max\{Z_1 + Z_{n,1}, Z_2 + Z_{n,2}\} - \max\{Z_{n,1}, Z_{n,2}\} \end{aligned}$$

Define the random variable $Y \equiv h_{\theta_0}(Z, X^n) = \max\{Z_1 + Z_{n,1}, Z_2 + Z_{n,2}\} - M_n$, where $M_n \equiv \max\{Z_{n,1}, Z_{n,2}\}$. Based on the results of Nadarajah and Kotz (2008), the (conditional) density of Y, denoted $f_{\theta_0}(y|X^n)$, is given by:

$$\begin{aligned} &\frac{1}{\sigma_1}\phi\left(\frac{Z_{n,1} - y - M_n}{\sigma_1}\right)\Phi\left(\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\rho(Z_{n,1} - y - M_n)}{\sigma_1} + \frac{y + M_n - Z_{n,2}}{\sigma_2}\right)\right) \\ + \quad &\frac{1}{\sigma_2}\phi\left(\frac{Z_{n,2} - y - M_n}{\sigma_2}\right)\Phi\left(\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\rho(Z_{n,2} - y - M_n)}{\sigma_2} + \frac{y + M_n - Z_{n,1}}{\sigma_1}\right)\right), \end{aligned}$$

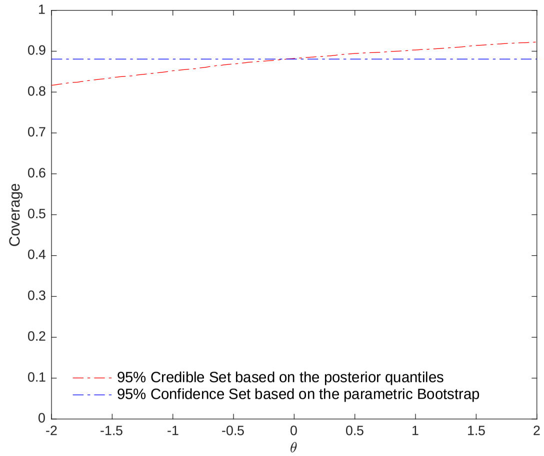
where $\rho = \sigma_{12}/\sigma_1\sigma_2$ and ϕ, Φ are the p.d.f. and the c.d.f. of a standard normal. It follows that:

$$f_{\theta_0}(y|Z_n) \leq \frac{1}{\sqrt{2\pi}}\left(\frac{1}{\sigma_1} + \frac{1}{\sigma_2}\right).$$

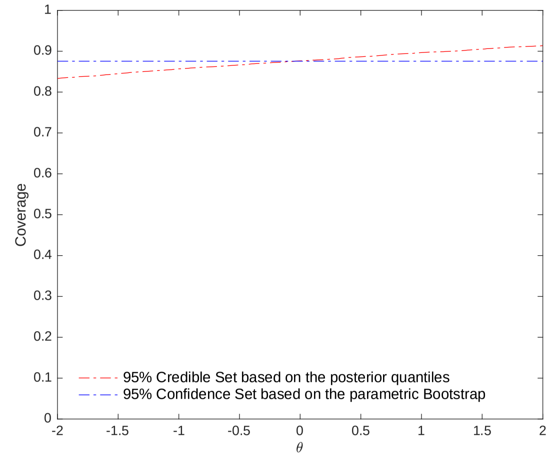
and so, by an analogous argument to the absolute value case, $F_{\theta_0}(y|X^n)$ is Lipschitz continuous with Lipschitz constant independent of Z_n and so Assumption 4(ii) holds.

Graphical illustration of coverage failure: Theorem 2 implies that credible sets based on the quantiles of $g(\theta_n^{P*})$ will effectively have the same asymptotic coverage properties as confidence sets based on quantiles of the bootstrap. For the transformation $g(\theta) = \max\{\theta_1, \theta_2\}$, this means that both methods lead to deficient frequentist coverage at the points in the parameter space in which $\theta_1 = \theta_2$. This is illustrated in Figure 2, which depicts the coverage of a nominal 95% bootstrap confidence set and different 95% credible sets. The coverage is evaluated assuming $\theta_1 = \theta_2 = \theta \in [-2, 2]$ and $\Sigma = \mathbb{I}_2$. The sample sizes considered are $n \in \{100, 200, 300, 500\}$. A prior characterized by $\mu = 0$ and $\lambda^2 = 1$ is used to calculate the credible sets. The credible sets and confidence sets have similar coverage as n becomes large and neither achieves 95% probability coverage for all $\theta \in [-2, 2]$.

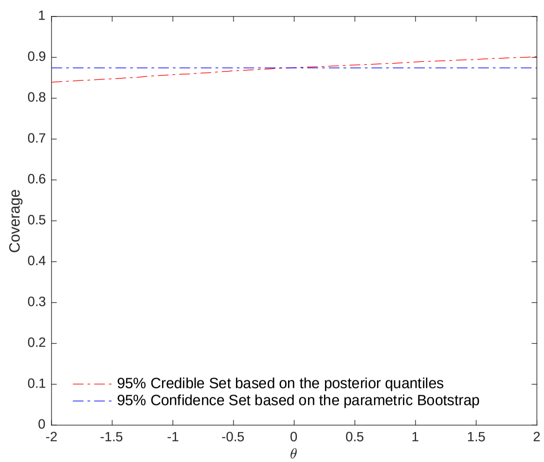
Figure 1: Coverage probability of 95% Credible Sets and Parametric Bootstrap Confidence Intervals.



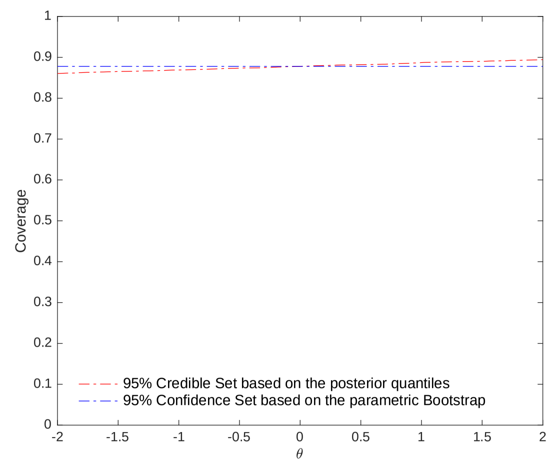
(a) $n = 100$



(b) $n = 200$



(c) $n = 300$



(d) $n = 500$

Description of Figure 2: Coverage probabilities of 95% bootstrap confidence intervals and 95% Credible Sets for $g(\theta) = \max\{\theta_1, \theta_2\}$ at $\theta_1 = \theta_2 = \theta \in [-2, 2]$ and $\Sigma = \mathbb{I}_2$ based on data from samples of size $n \in \{100, 200, 300, 500\}$. (Blue, Dotted Line) Coverage probability of 95% confidence intervals based on the quantiles of the parametric bootstrap distribution of $g(\widehat{\theta}_n)$; that is, $g(N_2(\widehat{\theta}_n, \mathbb{I}_2/n))$. (Red, Dotted Line) 95% credible sets based on quantiles of the posterior distribution of $g(\theta)$; that is $g(\mathcal{N}_2(\frac{n}{n+\lambda^2}\widehat{\theta}_n + \frac{\lambda^2}{n+\lambda^2}\mu\,,\ \frac{1}{n+\lambda^2}\mathbb{I}_2))$ for a prior characterized by $\mu = 0$ and $\lambda^2 = 1$.

Remark 1 Dümbgen (1993) and Hong and Li (2015) have proposed re-scaling the bootstrap to conduct inference about a directionally differentiable parameter. More specifically, the re-scaled bootstrap in Dümbgen (1993) and the numerical delta-method in Hong and Li (2015) can be implemented by constructing a new random variable:

$$y_n^* \equiv n^{1/2-\delta}\left(g\left(\frac{1}{n^{1/2-\delta}}Z_n^* + \widehat{\theta}_n\right) - g(\widehat{\theta}_n)\right),$$

where $0 \leq \delta \leq 1/2$ is a fixed parameter and Z_n^* could be either Z_n^{P*} or Z_n^{B*}. The suggested confidence interval is of the form:

$$CS_n^H(1-\alpha) = \left[g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{1-\alpha/2},\ g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{\alpha/2}\right] \tag{1.1}$$

where c_β^* denote the β-quantile of y_n^*. Hong and Li (2015) have recently established the pointwise validity of the confidence interval above.

Whenever (1.1) is implemented using posterior draws; i.e., by relying on draws from:

$$Z_n^{P*} \equiv \sqrt{n}(\theta_n^{P*} - \widehat{\theta}_n),$$

it seems natural to use the same posterior distribution to evaluate the credibility of the proposed confidence set. Figure 2 reports both the frequentist coverage and the Bayesian credibility of (1.1), assuming that the Hong and Li (2015) procedure is implemented using the posterior:

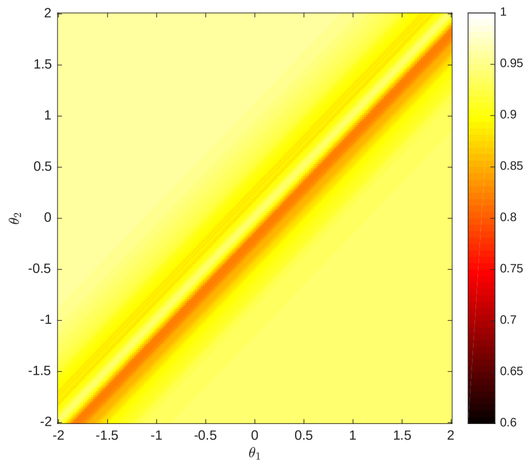
$$\theta_n^{P*}|X^n \sim \mathcal{N}_2\Big(\frac{n}{n+1}\widehat{\theta}_n\ ,\ \frac{1}{n+1}\mathbb{I}_2\Big).$$

The following figure shows that at least in this example fixing coverage comes at the expense of distorting Bayesian credibility.[1]

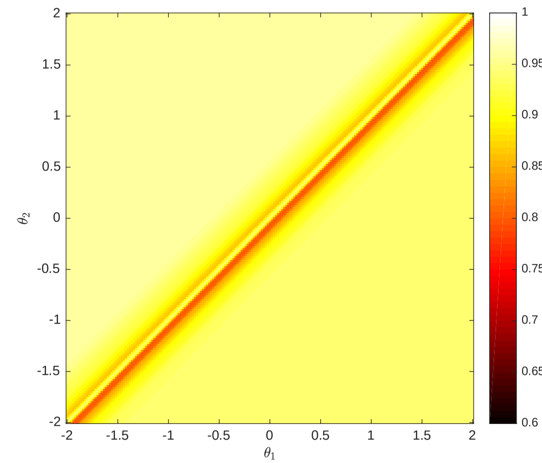
[1]The Bayesian credibility of $CS_n^H(1-\alpha)$ is given by:

$$\begin{aligned}&\mathbb{P}^*(g(\theta_n^{P*}) \in CS_n^H(1-\alpha)|X^n)\\ &\quad = \mathbb{P}^*\left(g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{1-\alpha/2}(X^n) \leq g(\theta_n^{P*}) \leq g(\widehat{\theta}_n) - \frac{1}{\sqrt{n}}c^*_{\alpha/2}(X^n)\ \Big|\ X^n\right)\end{aligned}$$

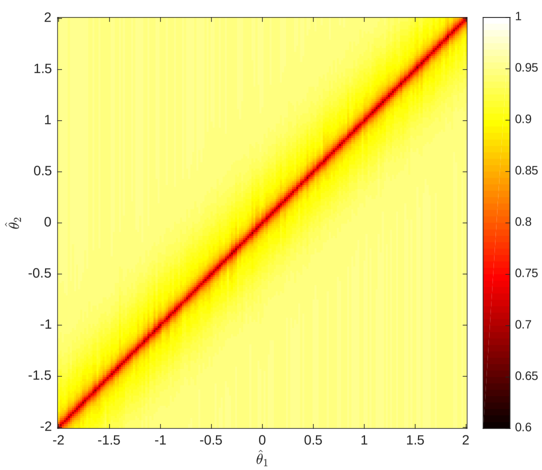
Figure 2: Coverage probability and Credibility of 95% Confidence Sets based on y_n^*



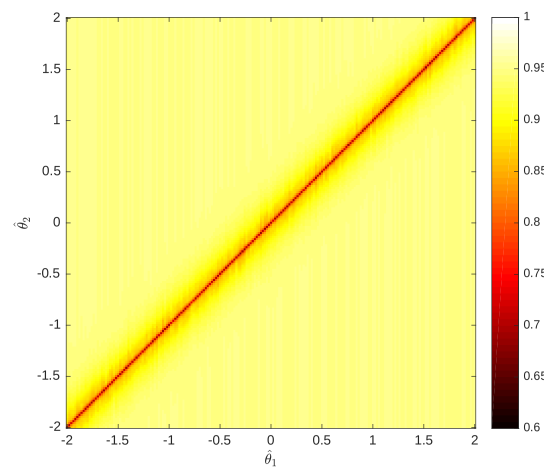
(a) Coverage Probability ($n = 100$)



(b) Coverage Probability ($n = 1000$)



(c) Credibility ($n = 100$)



(d) Credibility ($n = 1000$)

Description of Figure 2: Plots (a) and (b) show heat maps depicting the coverage probability of confidence sets based on the scaled random variable y_n^* for sample sizes $n \in \{100, 1000\}$ when $\theta_1, \theta_2 \in [-2, 2]$ and $\Sigma = \mathbb{I}_2$. Plots (c) and (d) show heat maps depicting the credibility of confidence sets based on the scaled random variable y_n^* for sample sizes $n \in \{100, 1000\}$ when $\theta = 0$, $\Sigma = \mathbb{I}_2$, Z_n^* is approximated by $N_2(0, \Sigma)$ for computing the quantiles of y_n^* and $\widehat{\theta}_{n,1}, \widehat{\theta}_{n,2} \in [-2, 2]$.

REFERENCES

Dümbgen, L. (1993): "On nondifferentiable functions and the bootstrap," *Probability Theory and Related Fields*, 95, 125–140.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2009): *Bayesian data analysis*, vol. 2 of *Texts in Statistical Science*, Taylor & Francis.

Hong, H. and J. Li (2015): "The numerical delta-method," *Working Paper, Stanford University.*

Nadarajah, S. and S. Kotz (2008): "Exact distribution of the max/min of two Gaussian random variables," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 16, 210–212.