Calibrating and Visualizing Some Bootstrap Confidence Regions

W.A. Dhanushka M. Welagedara and David J. Olive *
Southern Illinois University

January 15, 2024

Abstract

When the bootstrap sample size is moderate, bootstrap confidence regions tend to have undercoverage. Improving the coverage is known as calibrating the confidence region. Several bootstrap confidence regions are also prediction regions for a future value of a bootstrap statistic. Then a simple prediction region calibration technique can be applied to some bootstrap confidence regions to improve the coverage. The DD plot for visualizing prediction regions can also be used to visualize some bootstrap confidence regions.

KEY WORDS: Data Splitting, Prediction Regions.

1 INTRODUCTION

When the bootstrap sample size B is small or moderate, bootstrap confidence regions, including bootstrap confidence intervals, tend to have undercoverage: the probability that the confidence region contains the $p \times 1$ parameter vector $\boldsymbol{\theta}$ is less than the nominal large sample coverage probability $1 - \delta$. Then coverage can be increased by increasing the nominal coverage of the large sample bootstrap confidence region. For example, if the undercoverage of the nominal large sample 95% bootstrap confidence region with $B = 1000$ is 2%, increase the coverage to 97%. This procedure is known as calibrating the confidence region. Calibration tends to be difficult since the amount of undercoverage is usually unknown. This paper provides a simple method to improve the coverage, and provides a method to visualize some bootstrap confidence regions.

Section 1.1 reviews prediction intervals, prediction regions, confidence intervals, and confidence regions. Several bootstrap confidence intervals and regions are obtained by applying prediction intervals and regions to the bootstrap sample. Section 1.2 reviews a bootstrap theorem that shows some bootstrap confidence regions are asymptotically equivalent.

*

Section 2 gives a new bootstrap confidence region with a simple correction factor, while Section 3 shows how to visualize some bootstrap confidence regions. Section 4 gives some simulations.

1.1 Prediction Regions and Confidence Regions

Consider predicting a future test value Y_f given past training data $Y_1, ..., Y_n$ where $Y_1, ..., Y_n, Y_f$ are independent and identically distributed (iid). A large sample $100(1-\delta)\%$ prediction interval (PI) for Y_f is $[L_n, U_n]$ where the coverage $P(L_n \leq Y_f \leq U_n) = 1 - \delta_n$ is eventually bounded below by $1 - \delta$ as $n \to \infty$. We often want $1 - \delta_n \to 1 - \delta$ as $n \to \infty$. A large sample $100(1-\delta)\%$ PI is *asymptotically optimal* if it has the shortest asymptotic length: the length of $[\hat{L}_n, \hat{U}_n]$ converges to $U_s - L_s$ as $n \to \infty$ where $[L_s, U_s]$ is the *population shorth*: the shortest interval covering at least $100(1-\delta)\%$ of the mass.

Let the data $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$ have joint probability density function or probability mass function $f(\boldsymbol{y}|\theta)$ with parameter space Θ and support $\mathcal{Y}$. Let $L_n = L_n(\boldsymbol{Y})$ and $U_n = U_n(\boldsymbol{Y})$ be statistics such that $L_n(\boldsymbol{y}) \leq U_n(\boldsymbol{y})$, $\forall \boldsymbol{y} \in \mathcal{Y}$. Then the interval $[L_n(\boldsymbol{y}), U_n(\boldsymbol{y})]$ is a large sample $100(1-\delta)\%$ CI for θ if

$$P_\theta(L_n(\boldsymbol{Y}) \leq \theta \leq U_n(\boldsymbol{Y}))$$

is eventually bounded below by $1 - \delta$ for all $\theta \in \Theta$ as the sample size $n \to \infty$.

Consider predicting a $p \times 1$ future test value $\boldsymbol{x}_f$, given past training data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ where $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{x}_f$ are iid. A *large sample* $100(1-\delta)\%$ *prediction region* is a set $\mathcal{A}_n$ such that $P(\boldsymbol{x}_f \in \mathcal{A}_n)$ is eventually bounded below by $1 - \delta$ as $n \to \infty$. A prediction region is *asymptotically optimal* if its volume converges in probability to the volume of the minimum volume covering region or the highest density region of the distribution of $\boldsymbol{x}_f$.

A *large sample* $100(1-\delta)\%$ *confidence region* for a $p \times 1$ vector of parameters $\boldsymbol{\theta}$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\theta} \in \mathcal{A}_n)$ is eventually bounded below by $1 - \delta$ as $n \to \infty$.

For prediction intervals, let $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$ be the order statistics of the training data. Open intervals need more regularity conditions than closed intervals. For the following prediction interval, if the open interval $(Y_{(k_1)}, Y_{(k_2)})$ was used, we would need to add the regularity condition that the population percentiles $Y_{\delta/2}$ and $Y_{1-\delta/2}$ are continuity points of the cumulative distribution function $F_Y(y)$. See Frey (2013) for references.

Let $k_1 = \lceil n\delta/2 \rceil$ and $k_2 = \lceil n(1-\delta/2) \rceil$ where $0 < \delta < 1$. A large sample $100(1-\delta)\%$ *percentile prediction interval* for Y_f is

$$[Y_{(k_1)}, Y_{(k_2)}]. \tag{1}$$

The bootstrap percentile confidence interval given by Equation (2) is obtained by applying the percentile prediction interval (1) to the bootstrap sample $T_1^*, ..., T_B^*$. See Efron (1982).

A large sample $100(1-\delta)\%$ *bootstrap percentile confidence interval* for θ is an interval $[T^*_{(k_L)}, T^*_{(k_U)}]$ containing $\approx \lceil B(1-\delta) \rceil$ of the T_i^*. Let $k_1 = \lceil B\delta/2 \rceil$ and $k_2 = \lceil B(1-\delta/2) \rceil$.

A common choice is

$$[T^*_{(k_1)}, T^*_{(k_2)}]. \tag{2}$$

The shorth(c) estimator of the population shorth is useful for making asymptotically optimal prediction intervals. For a large sample $100(1-\delta)\%$ PI, the nominal coverage is $100(1-\delta)\%$. Undercoverage occurs if the actual coverage is below the nominal coverage. For example, if the actual coverage is 0.93 for a large sample 95% PI, then the undercoverage is 0.02. Consider intervals that contain c cases $[Y_{(1)}, Y_{(c)}], [Y_{(2)}, Y_{(c+1)}], ..., [Y_{(n-c+1)}, Y_{(n)}]$. Compute $Y_{(c)} - Y_{(1)}, Y_{(c+1)} - Y_{(2)}, ..., Y_{(n)} - Y_{(n-c+1)}$. Then the estimator shorth(c) $=$ $[Y_{(s)}, Y_{(s+c-1)}]$ is the interval with the shortest length. The shorth(c) interval is a large sample $100(1-\delta)\%$ PI if $c/n \to 1-\delta$ as $n \to \infty$, that often has the asymptotically shortest length. Let $k_n = \lceil n(1-\delta) \rceil$. Frey (2013) showed that for large $n\delta$ and iid data, the large sample $100(1-\delta)\%$ shorth(k_n) prediction interval has maximum undercoverage $\approx 1.12\sqrt{\delta/n}$, and used the large sample $100(1-\delta)\%$ PI shorth(c) $=$

$$[Y_{(s)}, Y_{(s+c-1)}] \;\; \text{with} \;\; \mathrm{c} = \min(\mathrm{n}, \lceil \mathrm{n}[1 - \delta + 1.12\sqrt{\delta/\mathrm{n}}\ ]\ \rceil). \tag{3}$$

The shorth confidence interval is a practical implementation of the Hall (1988) shortest bootstrap percentile interval based on all possible bootstrap samples, and is obtained by applying shorth PI (3) to the bootstrap sample $T_1^*, ..., T_B^*$. See Pelawa Watagoda and Olive (2021). The large sample $100(1-\delta)\%$ shorth(c) CI $=$

$$[T^*_{(s)}, T^*_{(s+c-1)}] \;\; \text{where} \;\; \mathrm{c} = \min(\mathrm{B}, \lceil \mathrm{B}[1 - \delta + 1.12\sqrt{\delta/\mathrm{B}}\ ]\ \rceil). \tag{4}$$

To describe the Olive (2013) nonparametric prediction region, Mahalanobis distances will be useful. Let the $p \times 1$ column vector T be a multivariate location estimator, and let the $p \times p$ symmetric positive definite matrix $\boldsymbol{C}$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T, \boldsymbol{C}) = D^2_{\boldsymbol{x}_i}(T, \boldsymbol{C}) = (\boldsymbol{x}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{x}_i - T) \tag{5}$$

for each observation $\boldsymbol{x}_i$, where $i = 1, ..., n$. Notice that the Euclidean distance of $\boldsymbol{x}_i$ from the estimate of center T is $D_i(T, \boldsymbol{I}_p)$ where $\boldsymbol{I}_p$ is the $p \times p$ identity matrix. The classical Mahalanobis distance D_i uses $(T, \boldsymbol{C}) = (\overline{\boldsymbol{x}}, \boldsymbol{S})$, the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i \;\; \text{and} \;\; \boldsymbol{S} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})(\boldsymbol{x}_{\mathrm{i}} - \overline{\boldsymbol{x}})^{\mathrm{T}}. \tag{6}$$

Let the $p \times 1$ location vector be $\boldsymbol{\mu}$, often the population mean, and let the $p \times p$ dispersion matrix be $\boldsymbol{\Sigma}$, often the population covariance matrix. If $\boldsymbol{x}$ is a random vector, then the population squared Mahalanobis distance is

$$D^2_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}). \tag{7}$$

Like prediction intervals, prediction regions often need correction factors. For iid data from a distribution with a $p \times p$ nonsingular covariance matrix, it was found that the simulated maximum undercoverage of prediction region (9) without the correction factor was about 0.05 when $n = 20p$. Hence the correction factor (8) is used to give better coverage for small n. Let $q_n = \min(1 - \delta + 0.05, 1 - \delta + p/n)$ for $\delta > 0.1$ and

$$q_n = \min(1 - \delta/2, 1 - \delta + 10\delta p/n), \quad \text{otherwise.} \tag{8}$$

If $1 - \delta < 0.999$ and $q_n < 1 - \delta + 0.001$, set $q_n = 1 - \delta$. Let $D_{(U_n)}$ be the $100q_n$th sample quantile of the D_i where $i = 1, ..., n$. Olive (2013) suggests $n \geq 50p$ may be needed for the following prediction region to have a good volume, and $n \geq 20p$ for good coverage. Of course for any n there are distributions that will have severe undercoverage.

The large sample $100(1 - \delta)\%$ *nonparametric prediction region* for a future value $\boldsymbol{x}_f$ given iid data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ is

$$\{\boldsymbol{z} : (\boldsymbol{z} - \overline{\boldsymbol{x}})^T \boldsymbol{S}^{-1} (\boldsymbol{z} - \overline{\boldsymbol{x}}) \leq D^2_{(U_n)}\} = \{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \leq D^2_{(U_n)}\}. \tag{9}$$

The Olive (2017ab, 2018) prediction region method confidence region applies prediction region (9) to the bootstrap sample. Let the bootstrap sample be $T_1^*, ..., T_B^*$. Let $\overline{T}^*$ and $\boldsymbol{S}_T^*$ be the sample mean and sample covariance matrix of the bootstrap sample.

The large sample $100(1 - \delta)\%$ prediction region method confidence region for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{w} : (\boldsymbol{w} - \overline{T}^*)^T [\boldsymbol{S}_T^*]^{-1} (\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}_T^*) \leq D^2_{(U_B)}\} \tag{10}$$

where the cutoff $D^2_{(U_B)}$ is the $100q_B$th sample quantile of the $D_i^2 = (T_i^* - \overline{T}^*)^T [\boldsymbol{S}_T^*]^{-1} (T_i^* - \overline{T}^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if $(\overline{T}^* - \boldsymbol{\theta}_0)^T [\boldsymbol{S}_T^*]^{-1} (\overline{T}^* - \boldsymbol{\theta}_0) > D^2_{(U_B)}$.

The Olive (2017ab, 2018) large sample $100(1 - \delta)\%$ modified Bickel and Ren (2001) confidence region is

$$\{\boldsymbol{w} : (\boldsymbol{w} - T_n)^T [\boldsymbol{S}_T^*]^{-1} (\boldsymbol{w} - T_n) \leq D^2_{(U_{BT})}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{S}_T^*) \leq D^2_{(U_{BT})}\} \tag{11}$$

where the cutoff $D^2_{(U_{BT})}$ is the $100q_B$th sample quantile of the $D_i^2 = (T_i^* - T_n)^T [\boldsymbol{S}_T^*]^{-1} (T_i^* - T_n)$. Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if $(T_n - \boldsymbol{\theta}_0)^T [\boldsymbol{S}_T^*]^{-1} (T_n - \boldsymbol{\theta}_0) > D^2_{(U_{BT})}$.

Shift region (9) to have center T_n, or equivalently, change the cutoff of region (11) to $D^2_{(U_B)}$ to get the Pelawa Watagoda and Olive (2021) large sample $100(1 - \delta)\%$ hybrid confidence region: $\{\boldsymbol{w} : (\boldsymbol{w} - T_n)^T [\boldsymbol{S}_T^*]^{-1} (\boldsymbol{w} - T_n) \leq D^2_{(U_B)}\} =$

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{S}_T^*) \leq D^2_{(U_B)}\}. \tag{12}$$

Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if
$(T_n - \boldsymbol{\theta}_0)^T [\boldsymbol{S}_T^*]^{-1} (T_n - \boldsymbol{\theta}_0) > D^2_{(U_B)}$.

Rajapaksha and Olive (2022) gave the following two confidence regions. The names of these confidence regions were chosen since they are similar to the Bickel and Ren and prediction region method confidence regions.

The large sample $100(1-\delta)\%$ BR confidence region is

$$\{\boldsymbol{w} : n(\boldsymbol{w} - T_n)^T \boldsymbol{C}_n^{-1}(\boldsymbol{w} - T_n) \leq D^2_{(U_{BT})}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{C}_n/n) \leq D^2_{(U_{BT})}\} \tag{13}$$

where the cutoff $D^2_{(U_{BT})}$ is the $100q_B$th sample quantile of the $D_i^2 = n(T_i^* - T_n)^T \boldsymbol{C}_n^{-1}(T_i^* - T_n)$. Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if $n(T_n - \boldsymbol{\theta}_0)^T \boldsymbol{C}_n^{-1}(T_n - \boldsymbol{\theta}_0) > D^2_{(U_{BT})}$.

The large sample $100(1-\delta)\%$ PR confidence region for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{w} : n(\boldsymbol{w} - \overline{T}^*)^T \boldsymbol{C}_n^{-1}(\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{C}_n/n) \leq D^2_{(U_B)}\} \tag{14}$$

where $D^2_{(U_B)}$ is the $100q_B$th sample quantile of the $D_i^2 = n(T_i^* - \overline{T}^*)^T \boldsymbol{C}_n^{-1}(T_i^* - \overline{T}^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if $n(\overline{T}^* - \boldsymbol{\theta}_0)^T \boldsymbol{C}_n^{-1}(\overline{T}^* - \boldsymbol{\theta}_0) > D^2_{(U_B)}$.

Assume that $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{x}_f$ are iid $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{x}})$. Then the Chew (1966) large sample $100(1-\delta)\%$ *classical prediction region* for multivariate normal data is

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{x}}, \boldsymbol{S}) \leq \chi^2_{p,1-\delta}\}. \tag{15}$$

The next bootstrap confidence region is similar to what would be obtained if the classical prediction region (15) for multivariate normal data was applied to the bootstrap sample. The large sample $100(1-\delta)\%$ standard bootstrap confidence region for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{w} : (\boldsymbol{w} - T_n)^T [\boldsymbol{S}^*_T]^{-1}(\boldsymbol{w} - T_n) \leq D^2_{1-\delta}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{S}^*_T) \leq D^2_{1-\delta}\} \tag{16}$$

where $D^2_{1-\delta} = \chi^2_{p,1-\delta}$ or $D^2_{1-\delta} = d_n F_{p,d_n,1-\delta}$ where $d_n \to \infty$ as $n \to \infty$.

If $p = 1$, then a hyperellipsoid is an interval, and confidence intervals are special cases of confidence regions. Suppose the parameter of interest is θ, and there is a bootstrap sample $T_1^*, ..., T_B^*$ where the statistic T_n is an estimator of θ based on a sample of size n. Let $a_i = |T_i^* - \overline{T}^*|$ and let $b_i = |T_i^* - T_n|$. Let $\overline{T}^*$ and S_T^{2*} be the sample mean and variance of the T_i^*. Then the squared Mahalanobis distance $D^2_\theta = (\theta - \overline{T}^*)^2 / S_T^{*2} \leq D^2_{(U_B)}$ is equivalent to $\theta \in [\overline{T}^* - S^*_T D_{(U_B)}, \overline{T}^* + S^*_T D_{(U_B)}] = [\overline{T}^* - a_{(U_B)}, \overline{T}^* + a_{(U_B)}]$, which is an interval centered at $\overline{T}^*$ just long enough to cover U_B of the T_i^*. Efron (2014) used a similar large sample $100(1-\delta)\%$ confidence interval assuming that $\overline{T}^*$ is asymptotically normal. Then the large sample $100(1-\delta)\%$ PR CI is $[\overline{T}^* - a_{(U_B)}, \overline{T}^* + a_{(U_B)}]$. The large sample $100(1-\delta)\%$ BR CI is $[T_n - b_{(U_{BT})}, T_n + b_{(U_{BT})}]$ is an interval centered at T_n just long enough to cover U_{BT} of the T_i^*. The large sample $100(1-\delta)\%$ hybrid CI is $[T_n - a_{(U_B)}, T_n + a_{(U_B)}]$.

The following prediction region will be used to develop a new correction factor for bootstrap confidence regions. See Section 2. Data splitting divides the training data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ into two sets: H and the validation set V where H has n_H of the cases and V has the remaining $n_V = n - n_H$ cases $i_1, ..., i_{n_V}$.

The estimator $(T_H, \boldsymbol{C}_H)$ is computed using the data set H. Then the squared validation distances $D_j^2 = D^2_{\boldsymbol{x}_{i_j}}(T_H, \boldsymbol{C}_H) = (\boldsymbol{x}_{i_j} - T_H)^T \boldsymbol{C}_H^{-1}(\boldsymbol{x}_{i_j} - T_H)$ are computed for the

$j = 1, ..., n_V$ cases in the validation set V. Let $D^2_{(U_V)}$ be the U_Vth order statistic of the D^2_j where

$$U_V = \min(n_V, \lceil (n_V + 1)(1 - \delta) \rceil). \tag{17}$$

The Haile, Zhang, and Olive (2024) large sample $100(1-\delta)\%$ data splitting prediction region for $\boldsymbol{x}_f$ is

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(T_H, \boldsymbol{C}_H) \le D^2_{(U_V)}\}. \tag{18}$$

1.2 Some Confidence Region Theory

Some large sample theory for bootstrap confidence regions is given in the references in Section 1.1. The following Pelawa Watagoda and Olive (2021) theorem and proof is useful.

Theorem 1. a) Suppose i) $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, and ii) $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \boldsymbol{u}$ with $E(\boldsymbol{u}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{u}) = \boldsymbol{\Sigma_u}$. Then iii) $\sqrt{n}(\overline{T}^* - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, iv) $\sqrt{n}(T_i^* - \overline{T}^*) \xrightarrow{D} \boldsymbol{u}$, and v) $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{P} \boldsymbol{0}$.

b) Then the prediction region method gives a large sample confidence region for $\boldsymbol{\theta}$ provided that the sample percentile $\hat{D}^2_{1-\delta}$ of the $D^2_{T_i^*}(\overline{T}^*, \boldsymbol{S}^*_T) = \sqrt{n}(T_i^* - \overline{T}^*)^T (n\boldsymbol{S}^*_T)^{-1} \sqrt{n}(T_i^* - \overline{T}^*)$ is a consistent estimator of the percentile $D^2_{n,1-\delta}$ of the random variable $D^2_{\boldsymbol{\theta}}(\overline{T}^*, \boldsymbol{S}^*_T) = \sqrt{n}(\boldsymbol{\theta} - \overline{T}^*)^T (n\boldsymbol{S}^*_T)^{-1} \sqrt{n}(\boldsymbol{\theta} - \overline{T}^*)$ in that $\hat{D}^2_{1-\delta} - D^2_{n,1-\delta} \xrightarrow{P} 0$.

Proof. With respect to the bootstrap sample, T_n is a constant and the $\sqrt{n}(T_i^* - T_n)$ are iid for $i = 1, ..., B$. Fix B. Then

$$\begin{bmatrix} \sqrt{n}(T_1^* - T_n) \\ \vdots \\ \sqrt{n}(T_B^* - T_n) \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_B \end{bmatrix}$$

where the $\boldsymbol{v}_i$ are iid with the same distribution as $\boldsymbol{u}$. For fixed B, the average of the $\sqrt{n}(T_i^* - T_n)$ is

$$\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{D} \frac{1}{B} \sum_{i=1}^{B} \boldsymbol{v}_i \sim AN_g\left(\boldsymbol{0}, \frac{\boldsymbol{\Sigma_u}}{B}\right)$$

by Continuous Mapping Theorem where $\boldsymbol{z} \sim AN_g(\boldsymbol{0}, \boldsymbol{\Sigma})$ is an asymptotic multivariate normal approximation. Hence as $B \to \infty$, $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{P} \boldsymbol{0}$, and iii), iv), and v) hold. Hence b) follows. $\square$

Under regularity conditions, Bickel and Ren (2001), Olive (2017b, 2018), and Pelawa Watagoda and Olive (2021) proved that (10), (11), and (12) are large sample confidence regions. For Theorem 1, usually i) and ii) are proven using large sample theory. Then

$$D_1^2 = D^2_{T_i^*}(\overline{T}^*, \boldsymbol{C}_n/n) = \sqrt{n}(T_i^* - \overline{T}^*)^T \boldsymbol{C}_n^{-1} \sqrt{n}(T_i^* - \overline{T}^*),$$

$$D_2^2 = D^2_{\boldsymbol{\theta}}(T_n, \boldsymbol{C}_n/n) = \sqrt{n}(T_n - \boldsymbol{\theta})^T \boldsymbol{C}_n^{-1} \sqrt{n}(T_n - \boldsymbol{\theta}),$$

$$D_3^2 = D^2_{\boldsymbol{\theta}}(\overline{T}^*, \boldsymbol{C}_n/n) = \sqrt{n}(\overline{T}^* - \boldsymbol{\theta})^T \boldsymbol{C}_n^{-1} \sqrt{n}(\overline{T}^* - \boldsymbol{\theta}), \quad \text{and}$$

$$D_4^2 = D^2_{T_i^*}(T_n, \boldsymbol{C}_n/n) = \sqrt{n}(T_i^* - T_n)^T \boldsymbol{C}_n^{-1} \sqrt{n}(T_i^* - T_n),$$

are well behaved. If $\boldsymbol{C}_n^{-1} \xrightarrow{P} \boldsymbol{C}^{-1}$, then $D_j^2 \xrightarrow{D} D^2 = \boldsymbol{u}^T \boldsymbol{C}^{-1} \boldsymbol{u}$, and (13) and (14) are large sample confidence regions. If $\boldsymbol{C}_n^{-1}$ is "not too ill conditioned," then $D_j^2 \approx \boldsymbol{u}^T \boldsymbol{C}_n^{-1} \boldsymbol{u}$ for large n, and the confidence regions (13) and (14) will have coverage near $1 - \delta$. See Rajapaksha and Olive (2022).

2 The Two Sample Bootstrap

Correction factors for calibrating confidence regions and prediction regions are often difficult to obtain. For prediction regions, see Barndorff-Nielsen and Cox (1996), Beran (1990), Fonseca, Giummole, and Vidoni (2012), Frey (2013), Hall, Peng, and Tajvidi (1999), and Ueki and Fueda (2007). Simulation was used to obtain correction factor (8). The bootstrap confidence regions (2), (4), and (10) were obtained by applying prediction regions (1), (3), and (9), respectively, on the bootstrap sample. By Theorem 1, bootstrap confidence regions (11) and (12) are asymptotically equivalent to (10). Hence these large sample confidence regions for $\boldsymbol{\theta}$ are also large sample prediction regions for a future value of the bootstrap statistic T_F^*.

Haile, Zhang, and Olive (2024) proved that the data splitting prediction regions (18) have coverage $\geq \min(n_V, \lceil (n_V+1)(1-\delta) \rceil)/(n_V+1)$, with equality if the probability of ties is zero. Hence data splitting can be used to calibrate prediction regions. The new confidence region gets $(T_H, \boldsymbol{C}_H)$ from the bootstrap data set $T_1^*, ..., T_B^*$ using $n_H = B$. For example, $(T_H, \boldsymbol{C}_H) = (\overline{T}^*, \boldsymbol{S}_T^*)$. Then a second bootstrap sample $T_{2,1}^*, ..., T_{2,n_V}^*$ is drawn. Then the new large sample $100(1-\delta)\%$ two sample bootstrap confidence region is

$$\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_H, \boldsymbol{C}_H) \leq D^2_{(U_V)}\}. \tag{19}$$

For a large sample 95% confidence region, we recommend $n_v = 49, 99$, or B. Then as a prediction region for T_F^*, (19) has coverage probability nearly equal to 0.95. Hence the coverage probability for $\boldsymbol{\theta}$ will be near 0.95 if n, B, and n_V are large enough.

The two sample bootstrap confidence region applies the data splitting prediction region on $T_1^*, ..., T_B^*, T_{2,1}^*, ..., T_{2,n_V}^*$ with $n_H = B$ and $n_V = n_V$ where H uses the first B cases and V uses the remaining n_V cases. Random selection of cases is not needed since the T^*s are iid with respect to the bootstrap sample. For (19) to be a large sample $100(1-\delta)\%$ confidence region, the region applied to the first sample H needs to be both a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ and a large sample $100(1-\delta)\%$ prediction region for T_f^*. Using $(T_H, \boldsymbol{C}_H) = (\overline{T}^*, \boldsymbol{S}_T^*)$ corresponds to (10) while using $(T_H, \boldsymbol{C}_H) = (T_n, \boldsymbol{S}_T^*)$ corresponds to (11). Thus the two sample bootstrap confidence region corresponding to (10) is

$$\{\boldsymbol{w} : (\boldsymbol{w} - \overline{T}^*)^T [\boldsymbol{S}_T^*]^{-1} (\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_V)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}_T^*) \leq D^2_{(U_V)}\}.$$

Hence the sample percentile $D^2_{(U_B)}$ in (10) gets replaced by the order statistic $D^2_{(U_V)}$.

3 Visualizing Some Bootstrap Confidence Regions

Olive (2013) showed how to visualize the nonparametric prediction region (9) with the Rousseeuw and Van Driessen (1999) DD plot of classical distances versus robust distances on the vertical axis. Hence the exact same method can be used to visualize the bootstrap confidence region (10).

If a good robust estimator is used, Olive (2002) showed that the plotted points in a DD plot cluster about the identity line with zero intercept and unit slope if the $\boldsymbol{x}_i$ are iid from a multivariate normal distribution with nonsingular covariance matrix, while the plotted points cluster about some other line through the origin if the $\boldsymbol{x}_i$ are iid from a large family of nonnormal elliptically contoured distributions. For the robust estimator of multivariate location and dispersion, we recommend the RFCH or RMVN estimator. See Olive (2017b), Olive and Hawkins (2010), and Zhang, Olive, and Ye (2012). These two estimators $(T_n, \boldsymbol{C}_n)$ are such that $\boldsymbol{C}_n$ is a $\sqrt{n}$ consistent estimator of $a\text{Cov}(\boldsymbol{x})$ for a large class of elliptically contoured distributions where the constant $a > 0$ depends on the elliptically contoured distribution and the estimator RFCH or RMVN, and $a = 1$ for the multivariate normal distribution with nonsingular covariance matrix. We used the RMVN estimator in the simulations.

Example 1, in the following section, shows how to use the DD plot to visualize some bootstrap confidence regions. Often $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_T)$, $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_T)$, and $\sqrt{n}(T_i^* - \overline{T}^*) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_T)$. Then the plotted points in the DD plot tend to cluster about the identity line in the DD plot. Note that $\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}^*_T) \le D^2_{(U_B)}\} = \{\boldsymbol{w} : D_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}^*_T) \le D_{(U_B)}\}$. Hence T_i^* such that $D_{T_i^*}(\overline{T}^*, \boldsymbol{S}^*_T) \le D_{(U_B)}$ are in the confidence region (10). These T_i^* correspond to the points to the left of the vertical line $MD = D_{(U_B)}$ in the DD plot.

4 Example and Simulations

Example 1. We generated $\boldsymbol{x}_i \sim N_4(\boldsymbol{0}, \boldsymbol{I})$ for $i = 1, ..., 250$. The coordinatewise median was the statistic T_n. The nonparametric bootstrap was used with $B = 1000$ for the 90% confidence region (10). Then the $100q_B$th sample quantile of the D_i is the 90.4% quantile. The DD plot of the bootstrap sample is shown in Figure 1. This bootstrap sample was a rather poor sample: the plotted points cluster about the identity line, but for most bootstrap samples, the clustering is tighter. The vertical line MD = 2.9098 is the cutoff for the prediction region method 90% confidence region (10). Hence the points to the left of the vertical line correspond to T_i^* that are inside the confidence region (10), while the points to the right of the vertical line correspond to T_i^* than are outside of the confidence region (10). The long horizontal line RD = 3.0995 is the cutoff using the robust estimator. When $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_T)$, then under mild regularity conditions, $\sqrt{n}(T_n - \overline{T}^*_n) \xrightarrow{P} \boldsymbol{0}$. The short horizontal line is RD = 2.8074, and MD = $2.8074 = \sqrt{\chi^2_{4,0.904}}$ is approximately the cutoff $\sqrt{\chi^2_{4,0.9}} = 2.7892$ that would be used by the standard bootstrap confidence region (mentally drop a vertical line from where the short horizontal line ends at the identity line). Variability in DD plots increases as RD increases.

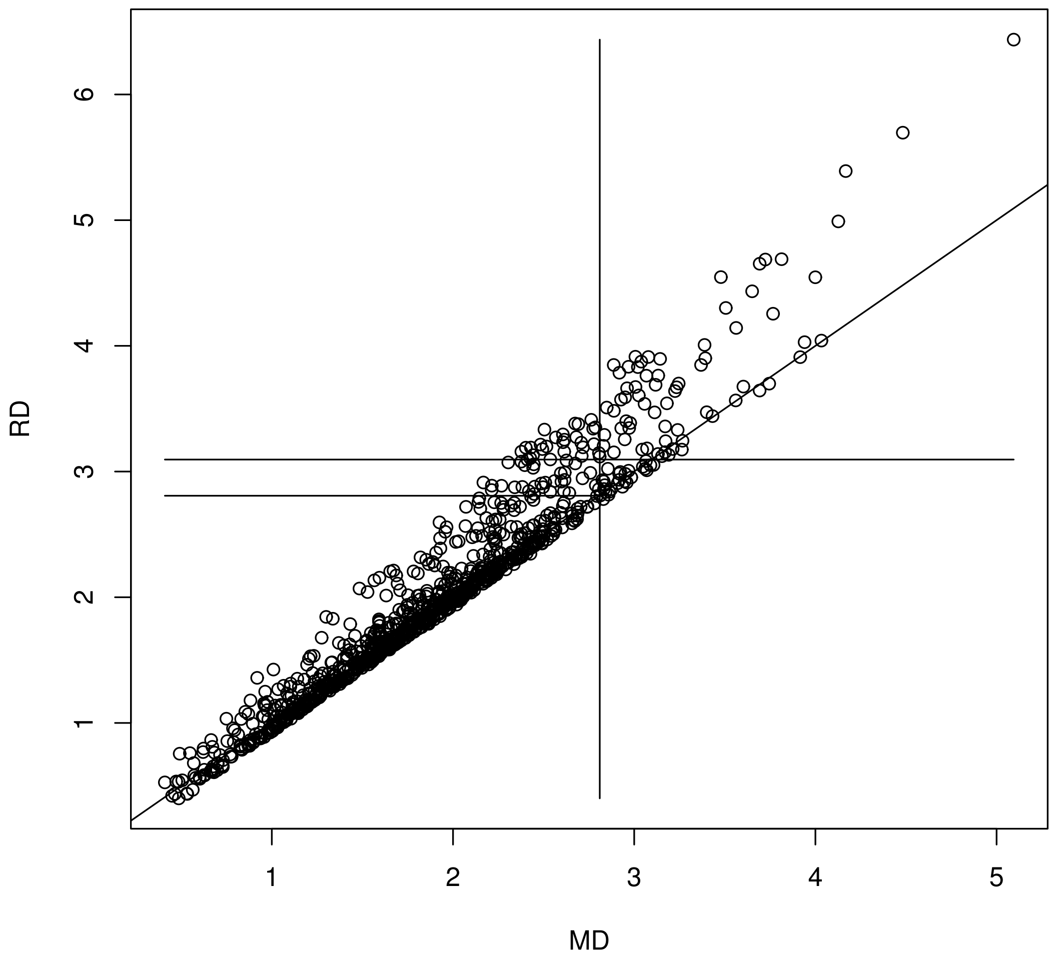

Figure 1: Visualizing the Confidence Region with a DD Plot

A small simulation study was done on large sample 95% confidence regions. The coordinatewise median was used since that statistic is moderately difficult to bootstrap. We used 5000 runs. Then coverage within [0.94,0.96] suggests that the true coverage is near the nominal coverage 0.95. The simulation used 10 distribution where xtype = 1 for $N_p(\mathbf{0}, \boldsymbol{I})$, xtype = 2, 3, 4 and 5 for $(1-\delta)N_p(\mathbf{0}, \boldsymbol{I}) + \delta N_p(\mathbf{0}, 25\boldsymbol{I})$, xtype = 6, 7, 8 and 9 for a multivariate t_d with d = 3, 5, 19 or d given by the user, and xtype=10 for a lognormal distribution shifted to have coordinatewise median = $\mathbf{0}$. If $\boldsymbol{w}$ corresponds to one of the above distributions, then $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{w}$ with $\boldsymbol{A} = diag(\sqrt{1}, \sqrt{2}, ..., \sqrt{p})$. Then the population coordinatewise median is $\mathbf{0}$ for each distribution. Table 1 shows the coverages and average cutoff for four large sample confidence regions: (10), (19) with $n_V = B = 1000$, (19) with $n_V = B = 49$, and (19) with $n_V = B = 99$. The coverage is the proportion of times the confidence region contained $\boldsymbol{\theta} = \mathbf{0}$ where $\boldsymbol{\theta}$ is a $p \times 1$ vector. Each confidence region has a cutoff, $D = \sqrt{D^2}$, that depends on the bootstrap sample, and the average of the 5000 cutoffs is given. Here $D^2 = D^2_{(U_B)}$ for confidence region (10), while $D^2 = D^2_{(U_V)}$ for confidence region (19), where the cutoff also depends on n_V. The coverages were usually between 0.94 and 0.96. The average cutoffs for the prediction region method large sample 95% confidence region tended to be very close to the average cutoffs for confidence region (19) with $n_V = B = 1000$. Note that $\sqrt{\chi^2_{2,0.95}} = 2.4477$ and $\sqrt{\chi^2_{4,0.95}} = 3.0802$ are the cutoffs for the standard bootstrap confidence region (15). The ratio of volumes of the two confidence regions is volume (10)/ volume (19) $= (D_{(U_B)}/D_{(U_V)})^p$.

Table 1: Coverages and Average Cutoffs for Some Large Sample 95% Confidence Regions, B=1000

n	p	dist	CR (10)	(19), $n_V = 1000$	(19), $n_V = 49$	(19), $n_V = 99$
100	2	N	(0.9430,2.4931)	(0.9450,2.5015)	(0.9536,2.7127)	(0.9452,2.5351)
100	2	LN	(0.9494,2.5025)	(0.9488,2.5088)	(0.9598,2.7401)	(0.9500,2.5539)
100	4	N	(0.9386,3.1738)	(0.9384,3.1795)	(0.9522,3.3922)	(0.9384,3.2177)
100	4	LN	(0.9456,3.2012)	(0.9466,3.2046)	(0.9598,3.4512)	(0.9468,3.2543)
200	4	N	(0.9476,3.1489)	(0.9480,3.1575)	(0.9590,3.3510)	(0.9490,3.1948)
200	4	LN	(0.9432,3.1673)	(0.9440,3.1700)	(0.9554,3.3861)	(0.9440,3.2065)

5 CONCLUSIONS

The bootstrap is due to Efron (1979). Also see Efron (1982) and Bickel and Freedman (1981). Ghosh and Polansky (2014) is a useful reference for bootstrap confidence regions. Visualizing a bootstrap confidence region is useful for checking whether the asymptotic normal approximation for the statistic is good since the plotted points will then tend to cluster tightly about the identity line. Making five plots corresponding to five bootstrap samples can be used to check the variability of the plots and the probability of getting a bad sample. For Example 1, most of the bootstrap samples produced plots that had tighter clustering about the identity line than the clustering in Figure 1.

Calibrating a bootstrap confidence region is useful for several reasons. For simulations, computation time can be reduced if B can be reduced. Using the correction factor (8) is faster than using the two sample bootstrap of Section 2, but the two sample bootstrap can be used to check the accuracy of (8), as in Table 1 with $n_V = B$. For a nominal 95% prediction region, the correction factor (8) increases the coverage to at most 97% of the training data. Coverage for test data $\boldsymbol{x}_f$ tends to be worse than coverage for training data. Using the (8) cutoff $D^2_{(U_B)}$ gives better coverage than using cutoff $D^2_{(U)}$ with $U = \lceil B(1-\delta) \rceil$. The two calibration methods in this paper were first applied to prediction regions, and work for bootstrap confidence regions (10) and (11) since those two regions are also prediction regions for T^*_f.

Plots and simulations were done in *R*. See R Core Team (2020). Welagedara (2023) lists some *R* functions for bootstrapping several statistics. Programs are in the collection of functions *slpack.txt*. The function `ddplot4` applied to the bootstrap sample can be used to visualize the bootstrap prediction region method confidence region. The function `medbootsim` was used for Table 1.

References

Barndorff-Nielsen, O.E. and Cox, D.R. (1996), "Prediction and Asymptotics," *Bernoulli,* 2, 319-340.

Beran, R. (1990), "Calibrating Prediction Regions," *Journal of the American Statistical Association*, 85, 715-723.

Bickel, P.J., and Freedman, D.A. (1981), "Some Asymptotic Theory for the Bootstrap," *The Annals of Statistics*, 9, 1196-1217.

Bickel, P.J., and Ren, J.–J. (2001), "The Bootstrap in Hypothesis Testing," in *State of the Art in Probability and Statistics: Festschrift for William R. van Zwet*, eds. de Gunst, M., Klaassen, C., and van der Vaart, A., The Institute of Mathematical Statistics, Hayward, CA, 91-112.

Chew, V. (1966), "Confidence, Prediction and Tolerance Regions for the Multivariate Normal Distribution," *Journal of the American Statistical Association,* 61, 605-617.

Efron, B. (1979), "Bootstrap Methods, Another Look at the Jackknife," *The Annals of Statistics*, 7, 1-26.

Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans,* SIAM, Philadelphia, PA.

Efron, B. (2014), "Estimation and Accuracy after Model Selection," (with discussion), *Journal of the American Statistical Association,* 109, 991-1007.

Fonseca, G., Giummole, F., and Vidoni, P. (2012), "A Note about Calibrated Prediction Regions and Distributions," *Journal of Statistical Planning and Inference*, 142, 2726-2734.

Frey, J. (2013), "Data-Driven Nonparametric Prediction Intervals," *Journal of Statistical Planning and Inference*, 143, 1039-1048.

Ghosh, S., and Polansky, A.M. (2014), "Smoothed and Iterated Bootstrap Confidence Regions for Parameter Vectors," *Journal of Multivariate Analysis*, 132, 171-182.

Haile, M.G., Zhang, L., and Olive, D.J. (2024), "Predicting Random Walks and a Data Splitting Prediction Region," *Stats*, 7, 23-33.

Hall, P. (1988), "Theoretical Comparisons of Bootstrap Confidence Intervals," (with discussion), *The Annals of Statistics*, 16, 927-985.

Hall, P., Peng, L., and Tajvidi, N. (1999), "On Prediction Intervals Based on Predictive Likelihood or Bootstrap Methods," *Biometrika*, 86, 871-880.

Olive, D.J. (2002), "Applications of Robust Distances for Regression," *Technometrics,* 44, 64-71.

Olive, D.J. (2013), "Asymptotically Optimal Regression Prediction Intervals and Prediction Regions for Multivariate Data," *International Journal of Statistics and Probability*, 2, 90-100.

Olive, D.J. (2017a), *Linear Regression*, Springer, New York, NY.

Olive, D.J. (2017b), *Robust Multivariate Analysis*, Springer, New York, NY.

Olive, D.J. (2018), “Applications of Hyperellipsoidal Prediction Regions,” *Statistical Papers*, 59, 913-931.

Olive, D.J., and Hawkins, D.M. (2010), “Robust Multivariate Location and Dispersion,” unpublished manuscript, see (http://parker.ad.siu.edu/Olive/pphbmld.pdf).

Pelawa Watagoda, L.C.R., and Olive, D.J. (2021), “Bootstrapping Multiple Linear Regression after Variable Selection,” *Statistical Papers,* 62, 681-700.

R Core Team (2020), “R: a Language and Environment for Statistical Computing,” R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

Rajapaksha, K.W.G.D.H., and Olive, D.J. (2022), “Wald Type Tests with the Wrong Dispersion Matrix,” *Communications in Statistics: Theory and Methods*, to appear.

Rousseeuw, P.J., and Van Driessen, K. (1999), “A Fast Algorithm for the Minimum Covariance Determinant Estimator,” *Technometrics,* 41, 212-223.

Ueki, M., and Fueda, K. (2007), “Adjusting Estimative Prediction Limits,” *Biometrika*, 94, 509-511.

Welagedara, W.A.D.M. (2023), “Model Selection, Data Splitting for ARMA Time Series, and Visualizing Some Bootstrap Confidence Regions,” Ph.D. Thesis, Southern Illinois University. See (http://parker.ad.siu.edu/Olive/swelagedara.pdf).

Zhang, J., Olive, D.J., and Ye, P. (2012), “Robust Covariance Matrix Estimation with Canonical Correlation Analysis,” *International Journal of Statistics and Probability*, 1, 119-136.