Wald Type Tests with the Wrong Dispersion Matrix

Kosman W.G.D.H. Rajapaksha and David J. Olive*

Department of Mathematics and Statistics, Washburn University, Topeka, Kansas 66621, kosman.rajapaksha@washburn.edu

School of Mathematical & Statistical Sciences, Southern Illinois University, Carbondale, Illinois 62901-4408, dolive@siu.edu

* Corresponding Author

Keywords Bootstrap; confidence region; MANOVA; weighted least squares.

Mathematics Subject Classification Primary 62H15; Secondary 62J10.

Abstract

A Wald type test with the wrong dispersion matrix is used when the dispersion matrix is not a consistent estimator of the asymptotic covariance matrix of the test statistic. One class of such tests occurs when there are p groups and it is assumed that the population covariance matrices from the p groups are equal, but the common covariance matrix assumption does not hold. The pooled t test, one-way ANOVA F test, and one-way MANOVA F test are examples of this class. Another class of such tests is used for weighted least squares. Two bootstrap confidence regions are modified to obtain large sample Wald type tests with the wrong dispersion matrix.

1. Introduction

This section reviews Wald type tests and Wald type tests with the wrong dispersion matrix. Consider testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where a $g \times 1$ statistic T_n satisfies $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u} \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma})$. If $\hat{\boldsymbol{\Sigma}}^{-1} \xrightarrow{P} \boldsymbol{\Sigma}^{-1}$ and H_0 is true, then

$$D_n^2 = D^2_{\boldsymbol{\theta}_0}(T_n, \hat{\boldsymbol{\Sigma}}/n) = n(T_n - \boldsymbol{\theta}_0)^T \hat{\boldsymbol{\Sigma}}^{-1} (T_n - \boldsymbol{\theta}_0) \xrightarrow{D} \boldsymbol{u}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{u} \sim \chi^2_g$$

as $n \to \infty$. Then a Wald type test rejects H_0 at significance level δ if $D_n^2 > \chi^2_{g,1-\delta}$ where

$P(X \leq \chi^2_{g,1-\delta}) = 1 - \delta$ if $X \sim \chi^2_g$, a chi-square distribution with g degrees of freedom.

It is common to implement a Wald type test using

$$D_n^2 = D^2_{\boldsymbol{\theta}_0}(T_n, \boldsymbol{C}_n/n) = n(T_n - \boldsymbol{\theta}_0)^T \boldsymbol{C}_n^{-1}(T_n - \boldsymbol{\theta}_0) \xrightarrow{D} \boldsymbol{u}^T \boldsymbol{C}^{-1} \boldsymbol{u}$$

as $n \to \infty$ if H_0 is true, where the $g \times g$ symmetric positive definite matrix $\boldsymbol{C}_n \xrightarrow{P} \boldsymbol{C} \neq \boldsymbol{\Sigma}$. Hence $\boldsymbol{C}_n$ is the wrong dispersion matrix, and $\boldsymbol{u}^T \boldsymbol{C}^{-1} \boldsymbol{u}$ does not have a χ^2_g distribution when H_0 is true. Often $\boldsymbol{C}_n$ is a regularized estimator of $\boldsymbol{\Sigma}$, or $\boldsymbol{C}_n^{-1}$ is a regularized estimator of the precision matrix $\boldsymbol{\Sigma}^{-1}$, such as $\boldsymbol{C}_n = diag(\hat{\boldsymbol{\Sigma}})$ or $\boldsymbol{C}_n = \boldsymbol{I}_g$, the $g \times g$ identity matrix. Another example is $\boldsymbol{C}_n = \boldsymbol{S}_p$, where $\boldsymbol{S}_p$ is a pooled covariance matrix, and it is assumed that the p groups have the same covariance matrix $\boldsymbol{\Sigma}$. When this assumption is violated, $\boldsymbol{C}_n$ is usually not a consistent estimator of $\boldsymbol{\Sigma}$. When the bootstrap is used, often $\boldsymbol{C}_n = n\boldsymbol{S}^*_T$ where $\boldsymbol{S}^*_T$ is the sample covariance matrix of the bootstrap sample $T_1^*, ..., T_B^*$. The assumption that $n\boldsymbol{S}^*_T$ is a consistent estimator of $\boldsymbol{\Sigma}$ is strong. See, for example, Machado and Parente (2005).

Some examples include the pooled t test and one-way ANOVA test. Rupasinghe Arachchige Don and Pelawa Watagoda (2018) and Rupasinghe Arachchige Don and Olive (2019) gave Wald type tests for analogs of the two sample Hotelling's T^2 and one-way MANOVA tests using a consistent estimator $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$. These tests could greatly outperform the classical tests that used the pooled covariance matrix when the sample sizes were large enough to give good estimates of the covariance matrix of each group, but for small sample sizes, the classical tests (with the wrong dispersion matrix) sometimes did better in the simulations.

If $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ and $\sqrt{n}(T_n^* - T_n) \xrightarrow{D} \boldsymbol{u}$, then the percentiles of $n(T_n - \boldsymbol{\theta}_0)^T \boldsymbol{C}_n^{-1}(T_n - \boldsymbol{\theta}_0)$ can be estimated with the sample percentiles of $n(T_n^* - T_n)^T \boldsymbol{C}_n^{-1}(T_n^* - T_n)$. Section 2 shows how to use this idea with bootstrap confidence regions. Section 3 reviews large sample theory for one-way MANOVA type tests, Section 4 considers weighted least squares, and Section 5 gives some simulation results.

2. Bootstrap Confidence Regions

This section modifies the Bickel and Ren (2001) and Olive (2017ab, 2018) confidence regions to work if $\boldsymbol{C}_n^{-1} \xrightarrow{P} \boldsymbol{C}^{-1} \neq \boldsymbol{\Sigma}^{-1}$. Consider testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is a known $g \times 1$ vector. Then a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$

is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\theta} \in \mathcal{A}_n)$ is eventually bounded below by $1-\delta$ as the sample size $n \to \infty$. Then reject H_0 if $\boldsymbol{\theta}_0$ is not in the confidence region $\mathcal{A}_n$.

For a confidence region, let the $g \times 1$ vector T_n be an estimator of the $g \times 1$ parameter vector $\boldsymbol{\theta}$. Let $T_1^*, ..., T_B^*$ be the bootstrap sample for T_n. Let $\boldsymbol{A}$ be a full rank $g \times q$ constant matrix where $g \leq q$, and consider testing $H_0 : \boldsymbol{A\mu} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{A\mu} \neq \boldsymbol{\theta}_0$ with $\boldsymbol{\theta} = \boldsymbol{A\mu}$ where often $\boldsymbol{\theta}_0 = \boldsymbol{0}$. Then let $T_n = \boldsymbol{A}\hat{\boldsymbol{\mu}}$, and let $T_i^* = \boldsymbol{A}\hat{\boldsymbol{\mu}}^*$ for $i = 1, ..., B$.

For a bootstrap confidence region, Mahalanobis distances will be useful. Let the $g \times 1$ column vector T be a multivariate location estimator, and let the $g \times g$ symmetric positive definite matrix $\boldsymbol{C}$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

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

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

$$\overline{\boldsymbol{z}} = \frac{1}{B} \sum_{i=1}^{B} \boldsymbol{z}_i \quad \text{and} \quad \boldsymbol{S} = \frac{1}{\mathrm{B}-1} \sum_{\mathrm{i}=1}^{\mathrm{B}} (\boldsymbol{z}_\mathrm{i} - \overline{\boldsymbol{z}})(\boldsymbol{z}_\mathrm{i} - \overline{\boldsymbol{z}})^\mathrm{T}. \tag{2}$$

Correction factors are often used to help prevent undercoverage. For example, suppose the limiting distribution is N(0,1) or χ^2_p. Then often a t_{d_n} or pF_{p,d_n} cutoff is used where $d_n \to \infty$ as $n \to \infty$. These t and F tests are asymptotically correct since $t_{d_n} \xrightarrow{D} N(0,1)$ and $pF_{p,d_n} \xrightarrow{D} \chi^2_p$ as $n \to \infty$. Obtaining correction factors for good coverage can be complicated. See Hall and Rieck (2001) and Ueki and Fueda (2007). For the correction factor below, and a nominal 95% confidence region, instead of using $D^2_{(\lceil 0.95B \rceil)}$ as the cutoff where $D^2_{(c)}$ is the cth order statistic of the D_i^2, the $100q_B$th sample quantile of the D_i^2, denoted by $D^2_{(U_B)}$, is used where $0.95B \leq U_B \leq 0.975B$ and $U_B \to 0.95B$ as B increases. Let $q_B = \min(1-\delta+0.05, 1-\delta+b/B)$ for $\delta > 0.1$ and

$$q_B = \min(1-\delta/2, 1-\delta+10\delta b/B), \quad \text{otherwise.} \tag{3}$$

If $1-\delta < 0.999$ and $q_B < 1-\delta+0.001$, set $q_B = 1-\delta$. We often use $b = g$ or $b = q$ if

$\boldsymbol{\theta} = \boldsymbol{A\mu}$ and $\boldsymbol{\mu}$ is a $q \times 1$ vector. This correction factor helps reduce undercoverage when $B \geq 50b$.

For the following *two new confidence regions*, let a statistic $T = T_n$ estimate $\boldsymbol{\theta}$. Assume $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ and $\sqrt{n}(T_n^* - T_n) \xrightarrow{D} \boldsymbol{u}$. 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 names of these confidence regions were chosen since they are similar to the Bickel and Ren and prediction region method confidence regions described after the following paragraph. The large sample $100(1-\delta)\%$ BR confidence region is

$$\{\boldsymbol{w} : n(\boldsymbol{w} - T_n)^T \boldsymbol{C}_n^{-1} (\boldsymbol{w} - T_n) \leq D^2_{(U_{BT})}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{C}_n/n) \leq D^2_{(U_{BT})}\} \tag{4}$$

where the cutoff $D^2_{(U_{BT})}$ is the $100q_B$th sample quantile of the $D_i^2 = n(T_i^* - T_n)^T \boldsymbol{C}_n^{-1}(T_i^* - T_n)$ where q_B is found from (3) with $\boldsymbol{z}_i = T_i^*$. 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

$$\{\boldsymbol{w} : n(\boldsymbol{w} - \overline{T}^*)^T \boldsymbol{C}_n^{-1} (\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{C}_n/n) \leq D^2_{(U_B)}\} \tag{5}$$

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

If $n\boldsymbol{C}_n^{-1} = [\boldsymbol{S}_T^*]^{-1}$, then (4) and (5) are the modified Bickel and Ren (2001) and Olive (2017ab, 2018) prediction region method large sample $100(1-\delta)\%$ confidence regions for $\boldsymbol{\theta}$. The hybrid confidence region replaces $D^2_{(U_{BT})}$ by $D^2_{(U_B)}$ in the modified Bickel and Ren confidence region. Under regularity conditions, Bickel and Ren (2001) and Olive (2017b, 2018) proved that (4) and (5) are large sample confidence regions when $n\boldsymbol{C}_n^{-1} = [\boldsymbol{S}_T^*]^{-1}$. Pelawa Watagoda and Olive (2021) gave simpler proofs. The Smaga (2017) bootstrap method replaces $\boldsymbol{C}_n$ by $\boldsymbol{C}_n^*$ and uses a different cutoff, but $\boldsymbol{C}_n^* = \boldsymbol{I}$ if $\boldsymbol{C}_n = \boldsymbol{I}$.

The ratio of the volumes of regions (5) and (4) is

$$\left(\frac{D_{(U_B)}}{D_{(U_{BT})}} \right)^g . \tag{6}$$

Hence region (5) has smaller volume than region (4) if $D_{(U_B)} < D_{(U_{BT})}$. See Johnson and Wichern (1988, p. 103) and Olive (2017b, p. 33) for the volume of a hyperellipsoid.

The theory for confidence regions (4) and (5) is simple. Pelawa Watagoda and Olive (2021) showed that under reasonable regularity conditions, i) $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, ii) $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \boldsymbol{u}$, iii) $\sqrt{n}(\overline{T}^* - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, and iv) $\sqrt{n}(T_i^* - \overline{T}^*) \xrightarrow{D} \boldsymbol{u}$. Usually i) and ii) are proven using large sample theory. If $\boldsymbol{u} \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{u}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ nonsingular, then Pelawa Watagoda and Olive (2021) showed $\sqrt{n}(T_n - \overline{T}^*) \xrightarrow{P} \boldsymbol{0}$. Thus iii) and iv) hold if i) and ii) hold. If T_n is the sample mean or sample coordinatewise median, then see Bickel and Freedman (1981) and Rupasinghe Arachchige Don and Olive (2019). 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 (4) and (5) 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 (4) and (5) will have coverage near $1 - \delta$.

Confidence regions (4) and (5) use sample percentiles of D_4^2 or D_1^2 from a bootstrap sample to get better cutoffs for Wald type tests that use the wrong dispersion matrix. If $g = 1$, then confidence intervals are special cases of confidence regions. Suppose there is a bootstrap sample $T_1^*, ..., T_B^*$ where the statistic $T = T_n$ is an estimator of θ based on a sample of size n. Let $T^*_{(1)}, ..., T^*_{(n)}$ be the order statistics of the bootstrap sample. The large sample $100(1-\delta)\%$ percentile confidence interval (CI) for θ is an interval $[T^*_{(k_L)}, T^*_{(K_U)}]$ containing $U_B \approx \lceil B(1-\delta) \rceil$ 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

$$[T^*_{(k_1)}, T^*_{(k_2)}]. \tag{7}$$

See Efron (1982, p. 78) and Chen (2016).

The large sample $100(1-\delta)\%$ *shorth*(c) *CI*

$$[T^*_{(s)}, T^*_{(s+c-1)}] \tag{8}$$

uses the interval $[T^*_{(1)}, T^*_{(c)}], [T^*_{(2)}, T^*_{(c+1)}], ..., [T^*_{(B-c+1)}, T^*_{(B)}]$ of shortest length. Here

$$c = \min(B, \lceil B[1 - \delta + 1.12\sqrt{\delta/B}\ ]\ \rceil) \tag{9}$$

for the shorth CI was obtained by applying the Frey (2013) prediction interval to the bootstrap sample. The shorth CI is the shortest percentile CI covering $c = c_n$ cases, and the shorth CI can be regarded as the shortest large sample $100(1-\delta)\%$ percentile CI, asymptotically. Hence the shorth CI is a practical implementation of the Hall (1988) shortest bootstrap interval based on all possible bootstrap samples. Olive (2014: p. 283, 2017b: p. 168, 2018) recommended using the shorth CI for the percentile CI.

If $a_i = |T^*_i - \overline{T}^*|$, then the CI corresponding to (5) is $[\overline{T}^* - a_{(U_B)}, \overline{T}^* + a_{(U_B)}]$, which is a percentile CI 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. The CI $[T_n - b_{(U_{BT})}, T_n + b_{(U_{BT})}]$ corresponding to (4) is a percentile CI centered at T_n just long enough to cover U_{BT} of the T^*_i with $b_i = |T^*_i - T_n|$.

Note that the two CIs corresponding to (4) and (5) can be computed without finding $\boldsymbol{C}_n$, $D_{(U_B)}$, or $D_{(U_{BT})}$. Hence these CIs correspond to the prediction region method CI and the modified Bickel and Ren CI. Suppose $\sqrt{n}(\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{D} N_g(\boldsymbol{0}, \boldsymbol{\Sigma})$. Then confidence regions (4) and (5) do not depend on whether $\boldsymbol{C}_n^{-1}$ or $d_n \boldsymbol{C}_n^{-1}$ is used if the scalar $d_n > 0$. Let $\theta = \boldsymbol{a}^T \boldsymbol{\mu}$ and $T_n = \boldsymbol{a}^T \hat{\boldsymbol{\mu}}$. Then $\boldsymbol{a}^T \boldsymbol{C}_n^{-1} \boldsymbol{a} = d_n \boldsymbol{a}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{a}$ where $d_n = \boldsymbol{a} \boldsymbol{C}_n^{-1} \boldsymbol{a} / \boldsymbol{a}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{a}$. Hence the confidence intervals do not depend on whether the wrong dispersion matrix is used.

As noted by Pelawa Watagoda and Olive (2021), if $g = 1$, if $\sqrt{n}(T_n - \theta) \xrightarrow{D} U$, and if $\sqrt{n}(T^*_i - T_n) \xrightarrow{D} U$ where U has a unimodal probability density function symmetric about zero with $E(U) = 0$, then the confidence intervals from the two confidence regions (4) and (5), the shorth confidence interval (8), and the "usual" percentile method confidence interval (7) are asymptotically equivalent (use the central proportion of the bootstrap sample, asymptotically).

3. One-Way MANOVA Type Tests

One-way MANOVA type tests give a large class of Wald type tests and Wald type tests with the wrong dispersion matrix. Using double subscripts will be useful for describing these models. Suppose there are independent random samples of size n_i from p different populations (treatments), or n_i cases are randomly assigned to p treatment groups. Then $n = \sum_{i=1}^p n_i$ and the group sample sizes are n_i for $i = 1, ..., p$. Assume that m response variables $\boldsymbol{y}_{ij} = (Y_{ij1}, ..., Y_{ijm})^T$ are measured for the ith treatment group and the jth case in the group. Hence $i = 1, ..., p$ and $j = 1, ..., n_i$. Assume the p treatments have possibly different population location vectors $\boldsymbol{\mu}_i$, such as $E(\boldsymbol{y}_{ij}) = \boldsymbol{\mu}_i$. Coordinatewise population medians and coordinatewise population trimmed means are also useful. Then a one-way MANOVA type test is used to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_p$ versus the alternative that not all of the $\boldsymbol{\mu}_i$ are equal.

Large sample theory can be used to derive Wald type tests, although large sample theory is not the only solution. Let $\text{Cov}(\boldsymbol{y}_{ij}) = \boldsymbol{\Sigma}_i$ be the nonsingular population covariance matrix of the ith treatment group or population. To simplify the large sample theory, assume $n_i = \pi_i n$ where $0 < \pi_i < 1$ and $\sum_{i=1}^p \pi_i = 1$. Let T_i be a multivariate location estimator such that $\sqrt{n_i}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$, and $\sqrt{n}(T_i - \boldsymbol{\mu}_i) \xrightarrow{D} N_m\left(\boldsymbol{0}, \frac{\boldsymbol{\Sigma}_i}{\pi_i}\right)$. Let $\boldsymbol{T} = (T_1^T, T_2^T, ..., T_p^T)^T$, $\boldsymbol{\nu} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, ..., \boldsymbol{\mu}_p^T)^T$, and $\boldsymbol{A}$ be a full rank $r \times mp$ matrix with rank r, then a large sample test of the form $H_0 : \boldsymbol{A\nu} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{A\nu} \neq \boldsymbol{\theta}_0$ uses

$$\boldsymbol{A}\sqrt{n}(\boldsymbol{T} - \boldsymbol{\nu}) \xrightarrow{D} \boldsymbol{u} \sim N_r\left(\boldsymbol{0}, \boldsymbol{A}\ diag\left(\frac{\boldsymbol{\Sigma}_1}{\pi_1}, \frac{\boldsymbol{\Sigma}_2}{\pi_2}, ..., \frac{\boldsymbol{\Sigma}_p}{\pi_p}\right)\ \boldsymbol{A}^T\right). \tag{10}$$

Let the Wald type statistic

$$t_0 = [\boldsymbol{AT} - \boldsymbol{\theta}_0]^T \left[\boldsymbol{A}\ diag\left(\frac{\hat{\boldsymbol{\Sigma}}_1}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}_2}{n_2}, ..., \frac{\hat{\boldsymbol{\Sigma}}_p}{n_p}\right)\ \boldsymbol{A}^T\right]^{-1} [\boldsymbol{AT} - \boldsymbol{\theta}_0]. \tag{11}$$

These results prove the following theorem.

Theorem 1. Under the above conditions, $t_0 \xrightarrow{D} \chi^2_r$ if H_0 is true.

A useful fact for the F and chi-square distributions is $d_n F_{g,d_n,1-\delta} \to \chi^2_{g,1-\delta}$ as $d_n \to \infty$. Here $P(X \leq F_{g,d_n,1-\delta}) = 1 - \delta$ if $X \sim F_{g,d_n}$. Reject H_0 if $t_0/r > F_{g,d_n,1-\delta}$ where $d_n = min(n_i) = min(n_1, ..., n_p)$.

This one-way MANOVA type test was used by Rupasinghe Arachchige Don and Olive (2019), and a special case was used by Zhang and Liu (2013) and Konietschke et al. (2015) with $T_i = \overline{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{\Sigma}}_i = \boldsymbol{S}_i$, the sample covariance matrix corresponding to the ith treatment group. The $p = 2$ case gives analogs to the two sample Hotelling's T^2 test. See Rupasinghe Arachchige Don and Pelawa Watagoda (2018).

Several tests use the common covariance matrix assumption $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$ for $i = 1, ..., p$. These tests are Wald type tests with the wrong dispersion matrix if the common covariance matrix assumption is wrong. Examples include the pooled t test with $m = p = 1$, the one-way ANOVA test with $m = 1$, the two sample Hotelling's T^2 test (with common covariance matrix) with $p = 2$, and the one-way MANOVA test.

For the Rupasinghe Arachchige Don and Olive (2019) one-way MANOVA type test, let $\boldsymbol{A}$ be the $m(p-1) \times mp$ block matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \dots & -\boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} & \dots & -\boldsymbol{I} \\ \vdots & \vdots & \vdots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \dots & \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix}.$$

Let $\boldsymbol{\mu}_i \equiv \boldsymbol{\mu}$, let $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ or, equivalently, $H_0 : \boldsymbol{A\nu} = \boldsymbol{0}$, and let

$$\boldsymbol{w} = \boldsymbol{AT} = \begin{bmatrix} T_1 - T_p \\ T_2 - T_p \\ \vdots \\ T_{p-2} - T_p \\ T_{p-1} - T_p \end{bmatrix}. \tag{12}$$

Then $\sqrt{n}\boldsymbol{w} \xrightarrow{D} N_{m(p-1)}(\boldsymbol{0}, \boldsymbol{\Sigma_w})$ if H_0 is true with $\boldsymbol{\Sigma_w} = (\boldsymbol{\Sigma}_{ij})$ where $\boldsymbol{\Sigma}_{ij} = \dfrac{\boldsymbol{\Sigma}_p}{\pi_p}$ for $i \neq j$, and $\boldsymbol{\Sigma}_{ii} = \dfrac{\boldsymbol{\Sigma}_i}{\pi_i} + \dfrac{\boldsymbol{\Sigma}_p}{\pi_p}$ for $i = j$. Hence

$$t_0 = n\boldsymbol{w}^T \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1} \boldsymbol{w} = \boldsymbol{w}^T \left(\frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}}{n} \right)^{-1} \boldsymbol{w} \xrightarrow{D} \chi^2_{m(p-1)}$$

as the $n_i \to \infty$ if H_0 is true. Here $\dfrac{\hat{\boldsymbol{\Sigma}}\boldsymbol{w}}{n}$ is a block matrix where the off diagonal block entries equal $\hat{\boldsymbol{\Sigma}}_p/n_p$ and the ith diagonal block entry is $\dfrac{\hat{\boldsymbol{\Sigma}}_i}{n_i} + \dfrac{\hat{\boldsymbol{\Sigma}}_p}{n_p}$ for $i = 1, ..., (p-1)$. Reject H_0 if

$$t_0 > m(p-1)F_{m(p-1),d_n}(1-\delta) \tag{13}$$

where $d_n = \min(n_1, ..., n_p)$. This Wald type test may start to outperform the one-way MANOVA test if $n \geq (m+p)^2$ and $n_i \geq 40m$ for $i = 1, ..., p$.

If $H_0 : \boldsymbol{A\nu} = \boldsymbol{\theta}_0$ is true, if the $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$ for $i = 1, ..., p$, and if $\hat{\boldsymbol{\Sigma}}$ is a consistent estimator of $\boldsymbol{\Sigma}$, then by Theorem 1

$$t_0 = [\boldsymbol{AT} - \boldsymbol{\theta}_0]^T \left[\boldsymbol{A}\ diag\left(\frac{\hat{\boldsymbol{\Sigma}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}}{n_2}, ..., \frac{\hat{\boldsymbol{\Sigma}}}{n_p}\right)\ \boldsymbol{A}^T\right]^{-1} [\boldsymbol{AT} - \boldsymbol{\theta}_0] \xrightarrow{D} \chi^2_r.$$

If H_0 is true but the $\boldsymbol{\Sigma}_i$ are not equal, then we get a bootstrap cutoff by using

$$t^*_{0i} = [\boldsymbol{AT}^*_i - \boldsymbol{AT}]^T \left[\boldsymbol{A}\ diag\left(\frac{\hat{\boldsymbol{\Sigma}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}}{n_2}, ..., \frac{\hat{\boldsymbol{\Sigma}}}{n_p}\right)\ \boldsymbol{A}^T\right]^{-1} [\boldsymbol{AT}^*_i - \boldsymbol{AT}] =$$

$$D^2_{\boldsymbol{AT}^*_i}\left(\boldsymbol{AT}, \boldsymbol{A}\ diag\left(\frac{\hat{\boldsymbol{\Sigma}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}}{n_2}, ..., \frac{\hat{\boldsymbol{\Sigma}}}{n_p}\right) \boldsymbol{A}^T\right).$$

Let $F_0 = t_0/r$. Then we can get a bootstrap cutoff using $F^*_{0i} = t^*_{0i}/r$. For $T_i = \overline{\boldsymbol{y}}_i$, let $\hat{\boldsymbol{\Sigma}}$ be the usual pooled covariance matrix estimator.

For Theorem 2, $(n-p)U = t_0 \xrightarrow{D} \chi^2_{m(p-1)}$ follows trivially from Theorem 1, under the equal covariance matrix assumption. Fujikoshi (2002) also showed $(n-p)U \xrightarrow{D} \chi^2_{m(p-1)}$. Kakizawa (2009) also gave large sample theory for some MANOVA tests. Lengthy calculations show $(n-p)U = t_0$. See Rajapaksha (2021) for details.

Theorem 2. For the one-way MANOVA test using $\boldsymbol{\theta}_0 = \boldsymbol{0}$, $\boldsymbol{A}$ as defined above Equation (12), and $T_i = \overline{\boldsymbol{y}}_i$,

$$(n-p)U = t_0 = [\boldsymbol{AT}]^T \left[\boldsymbol{A}\ diag\left(\frac{\hat{\boldsymbol{\Sigma}}}{n_1}, \frac{\hat{\boldsymbol{\Sigma}}}{n_2}, ..., \frac{\hat{\boldsymbol{\Sigma}}}{n_p}\right)\ \boldsymbol{A}^T\right]^{-1} [\boldsymbol{AT}]$$

where U is the Hotelling Lawley trace statistic. Hence if the $\boldsymbol{\Sigma}_i \equiv \boldsymbol{\Sigma}$ and $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ is true, then $(n-p)U = t_0 \xrightarrow{D} \chi^2_{m(p-1)}$.

4. Weighted Least Squares

The weighted least squares (WLS) model is $Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i$ for $i = 1, ..., n$ where the e_i are independent with $E(e_i) = 0$ and $V(e_i) = \sigma_i^2$. In matrix form, this model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e},$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. Also $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma_e} = diag(\sigma_i^2) = diag(\sigma_1^2, ..., \sigma_n^2)$ is an $n \times n$ positive definite matrix. A common assumption is that the $e_i = \sigma_i \tilde{e}_i$ where the $\tilde{e}_i$ are independent and identically distributed with $V(\tilde{e}_i) = 1$.

Under regularity conditions, the least squares estimator $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ can be shown to be a consistent estimator of $\boldsymbol{\beta}$ with $\text{Cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma_e}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ and $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. See, for example, White (1980). Assume $n\text{Cov}(\hat{\boldsymbol{\beta}}) \to \boldsymbol{V}$ as $n \to \infty$. If $\boldsymbol{X}^T\boldsymbol{X}/n \to \boldsymbol{W}^{-1}$ and $\boldsymbol{X}^T\boldsymbol{\Sigma_e}\boldsymbol{X}/n \to \boldsymbol{U}$, then $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{U}\boldsymbol{W}$. We assume that a constant β_1 corresponding to $x_1 \equiv 1$ is in the model so that the OLS residuals sum to 0.

A sandwich estimator is $\widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{OLS}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$. Often $\hat{\boldsymbol{D}}$ is not a consistent estimator of $\boldsymbol{\Sigma_e}$, but often $\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X}/n \xrightarrow{P} \boldsymbol{U}$ under regularity conditions. For the wild bootstrap, we will use $\hat{\boldsymbol{D}}_W = n\ diag(r_1^2, ..., r_n^2)/(n-p)$ where the r_i are the OLS residuals. Often $\hat{\boldsymbol{D}} = diag(d_i^2 r_i^2)$, where $\hat{\boldsymbol{D}}_W$ uses $d_i^2 = n/(n-p)$.

The *nonparametric bootstrap* = *pairs bootstrap* samples the cases $(Y_i, \boldsymbol{x}_i)$ with replacement, and uses

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\hat{\boldsymbol{\beta}} + \boldsymbol{e}^*$$

with $\boldsymbol{e}^* = \boldsymbol{r}^*$ where $(Y_i, \boldsymbol{x}_i, r_i)$ are selected with replacement to form $\boldsymbol{Y}^*, \boldsymbol{X}^*$, and $\boldsymbol{r}^*$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}^* = \hat{\boldsymbol{\beta}} + (\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^* = \hat{\boldsymbol{\beta}} + \boldsymbol{b}^*$ is obtained from the OLS regression of $\boldsymbol{Y}^*$ on $\boldsymbol{X}^*$. Thus $E(\hat{\boldsymbol{\beta}}^*) = \hat{\boldsymbol{\beta}} + E[(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^*] = \hat{\boldsymbol{\beta}} + \boldsymbol{b}$ where the expectation is with respect to the bootstrap distribution and the bias vector $\boldsymbol{b} = E(\boldsymbol{b}^*)$. Freedman (1981) showed that the nonparametric bootstrap can be useful for the WLS model with the e_i independent, suggesting that $\boldsymbol{b}^* = o_p(n^{-1/2})$ or $\boldsymbol{b}^* =$

$O_p(n^{-1/2})$. With respect to the bootstrap distribution, $\text{Cov}(\hat{\boldsymbol{\beta}}^*) = \text{Cov}[(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^*] = E[(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}\boldsymbol{r}^*\boldsymbol{r}^{*T}\boldsymbol{X}^*(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}] - \boldsymbol{b}\boldsymbol{b}^T$.

A version of the *wild bootstrap* uses

$$\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}^*$$

with $e_i^* = W_i c_n r_i$ where $P(W_i = \pm 1) = 0.5$, $E(W_i) = 0$, $V(W_i) = 1$ and $c_n = \sqrt{n/(n-p)}$. Note that $W_i = 2Z_i - 1$ where $Z_i \sim \text{binomial}(m = 1, p = 0.5) \sim \text{Bernoulli}(p = 0.5)$. See Flachaire (2005). With respect to the bootstrap distribution, the $c_n r_i$ are constants, and the e_i^* are independent with $E(e_i^*) = E(W_i)c_n r_i = 0$, and $V(e_i^*) = E(e_i^{*2}) = E(W_i^2)c_n^2 r_i^2 = c_n^2 r_i^2$. Thus $E(\boldsymbol{e}^*) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}^*) = \hat{\boldsymbol{D}}_W$. Then $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^*$ with $E(\hat{\boldsymbol{\beta}}^*) = \hat{\boldsymbol{\beta}}$ and $\text{Cov}(\hat{\boldsymbol{\beta}}^*) = \widehat{\text{Cov}}(\hat{\boldsymbol{\beta}}_{OLS}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\hat{\boldsymbol{D}}_W\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$, a sandwich estimator. Note that $\text{Cov}(\hat{\boldsymbol{\beta}}^*) = \text{Cov}(\hat{\boldsymbol{\beta}}) + (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T[\hat{\boldsymbol{D}}_W - \boldsymbol{\Sigma}_{\boldsymbol{e}}]\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

There is a large literature on WLS and sandwich estimators. See, for example, Buja et al. (2019), Eicker (1963, 1967), Hinkley (1977), Huber (1967), Long and Ervin (2000), MacKinnon and White (1985), Romano and Wolf (2017), White (1980), and Wu (1986). For more on the wild bootstrap, see Mammen (1992, 1993) and Wu (1986). Flachaire (2005) compares the wild and nonparametric bootstrap.

The following method is new. For the OLS model, $V(e_i) = V(Y_i|\boldsymbol{x}_i) = V(Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta}) = \sigma^2$. Hence $Y_i = Y_i|\boldsymbol{x}_i = Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta} = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ with $V(e_i) = \sigma^2$. For the WLS model, $Y_i = Y_i|\boldsymbol{x}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ with $V(e_i) = \sigma_i^2$, while $Y_i = Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta}_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + \epsilon_i$ with $V(\epsilon_i) = \tau_i^2$. The τ_i^2 can be estimated as follows. Divide the ordered $\boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$ into m_s slices each containing approximately n/m_s cases, and find the variance of the residuals v_j^2 in the jth slice for $j = 1, ..., m_s$. Then $\hat{\tau}_i^2 = nv_j^2/(n-p)$ if case i is in the jth slice. If the $\boldsymbol{x}_i$ are bounded, the maximum slice width $\to 0$, if $V(Y|\boldsymbol{x}^T\boldsymbol{\beta})$ is smooth, and the number of cases in each slice $\to \infty$ as $n \to \infty$, then $\hat{\tau}_i^2$ is a consistent estimator of τ_i^2. This method acts as if the variance τ_j^2 is constant within each slice j, and replaces $\hat{\boldsymbol{D}}_W = n\ diag(r_1^2, ..., r_n^2)/(n-p)$ by $diag(\hat{\tau}_1^2, ..., \hat{\tau}_n^2)$, a smoothed version of $\hat{\boldsymbol{D}}_W$. Another option would use a scatterplot smoother in a plot of $\hat{Y}_i$ vs. r_i^2.

The *parametric bootstrap* **does not assume** that the e_i are normal, but uses

$$\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}^*$$

where the $e_i^* \sim N(0, \hat{\tau}_i^2)$ are independent. Hence $\hat{\boldsymbol{\beta}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}^* \sim$

$$N_p[\hat{\boldsymbol{\beta}}, (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \ diag(\hat{\tau}_1^2, ..., \hat{\tau}_n^2) \ \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}].$$

5. Simulations

This section simulates one-way MANOVA type tests and WLS tests. Rajapaksha (2021) has a much larger simulation (including simulations for analogs of the pooled t test, two sample Hotelling's T^2 test, and one-way ANOVA test), and has some real data examples. Rajapaksha (2021) sometimes used coordinatewise medians in addition to sample means. 5000 runs were used, and B was the number of bootstrap replications used.

One-Way MANOVA

We used 3 groups for the one-way MANOVA type tests. Four types of data distributions $\boldsymbol{w}_i$ were considered that were identical for $i = 1, 2$, and 3. Then $\boldsymbol{y}_1 = \sigma_1 \boldsymbol{C}\boldsymbol{w}_1 + \delta_1 \boldsymbol{1}$, $\boldsymbol{y}_2 = \sigma_2 \boldsymbol{C}\boldsymbol{w}_2 + \delta_2 \boldsymbol{1}$, and $\boldsymbol{y}_3 = \sigma_3 \boldsymbol{C}\boldsymbol{w}_3 + \delta_3 \boldsymbol{1}$ or $\boldsymbol{y}_3 = \boldsymbol{w}_3$ where $\boldsymbol{1} = (1, .., 1)^T$ is a vector of ones and $\boldsymbol{C} = diag(1, \sqrt{2}, ..., \sqrt{m})$. The $\boldsymbol{w}_i$ distributions were the multivariate normal distribution $N_m(\boldsymbol{0}, \boldsymbol{I})$, the mixture distribution $0.6N_m(\boldsymbol{0}, \boldsymbol{I}) + 0.4N_m(\boldsymbol{0}, 25\boldsymbol{I})$, the multivariate t distribution with 4 degrees of freedom, and the multivariate lognormal distribution shifted to have zero mean. If $\sigma_1 = 1$ and $\delta_i = 0$ for $i = 1, 2, 3$, note that $\text{Cov}(\boldsymbol{y}_2) = \sigma_2^2 \ \text{Cov}(\boldsymbol{y}_1)$, and $E(\boldsymbol{y}_i) = E(\boldsymbol{w}_i) = \boldsymbol{0}$. If $\boldsymbol{y}_3 = \boldsymbol{w}_3$ then $\text{Cov}(\boldsymbol{y}_3) = c\boldsymbol{I}_m$ for some constant $c > 0$. If $\sigma_1 = 1$ and $\boldsymbol{y}_3 = \sigma_3 \boldsymbol{C}\boldsymbol{w}_3 + \delta_3 \boldsymbol{1}$, then $\text{Cov}(\boldsymbol{y}_3) = \sigma_3^2 \ \text{Cov}(\boldsymbol{y}_1)$.

Tables 1-3 give the coverage = proportion of times the test failed to reject H_0. The classical test (mancov), bootstrap classical test (bootcov) described above Theorem 2, large sample test (manLScov), and bootstrap test with $\boldsymbol{C}_n = \boldsymbol{I}$ where the PR (prcv) and BR (brcv) confidence regions were used. For power, group i has mean $\boldsymbol{\mu}_i = \delta_i \boldsymbol{1}$ where $\delta_2 = 2 \ \delta_1$ and $\delta_3 = 3 \ \delta_1$. When δ_1 increases, the distance between the mean vectors increases. The nominal coverage was 0.95. With 5000 runs, observed coverage between 0.94 and 0.96 suggests coverage is close to the nominal value.

Since the classical test uses the wrong dispersion matrix unless the $\boldsymbol{\Sigma}_i$ are equal, sometimes the test statistic tends to be smaller than the cutoff, resulting in higher than nominal coverage or a conservative test. This result sometimes happened in Tables 1 and 2. Sometimes the test statistic tends to be larger than the cutoff, resulting in lower than nominal coverage or a liberal test: H_0 is rejected too often when H_0 is true. This result often happened in Table 3 and sometimes in Table 2.

The bootstrap classical test had coverage near the nominal in the tables, and sometimes outperformed the large sample test for skewed data, as in Table 3. This good performance occurred because the n_i were rather large. When the n_i are not large enough, the bootstrap classical test can have undercoverage: the bootstrap cutoff is poor.

The large sample test tends to have a confidence region with smaller volume than the other tests if the n_i are large enough. For sample means, the lognormal distribution is a distribution with all moments that is known to need large sample sizes when the covariance matrix is estimated. See Hesterberg (2015).

The tests with $\boldsymbol{C}_n = \boldsymbol{I}$ controlled the type I error very well, at the expense of using a high volume hypersphere instead of a smaller volume hyperellipsoid, resulting in lower power. For skewed data, estimating covariance matrices is much more difficult than estimating means. In the simulations as the n_i approached $m \geq 50$ (need $n_i > m$ to compute the large sample test), the $\boldsymbol{C}_n = \boldsymbol{I}$ tests became conservative (not shown). The $\boldsymbol{C}_n = \boldsymbol{I}$ tests depend on the units of measurement.

WLS

Next, we describe a small WLS simulation study that is similar to that for the full OLS model done by Pelawa Watagoda and Olive (2021). The simulation used $p = 4$ and 8, $\psi = 0, 0.5, 1/\sqrt{p}$, and 0.9; and $k = 1$ and $p - 2$ where k and ψ are defined in the following paragraph.

Let $\boldsymbol{x} = (1 \ \boldsymbol{u}^T)^T$ where $\boldsymbol{u}$ is the $(p-1) \times 1$ vector of nontrivial predictors. In the simulations, for $i = 1, ..., n$, we generated $\boldsymbol{w}_i \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$ where the $m = p - 1$ elements of the vector $\boldsymbol{w}_i$ are independent and identically distributed (iid) N(0,1). Let the $m \times m$

Table 1: One-Way MANOVA Type Test, Coverage for MVN data with $\boldsymbol{\Sigma}_3 \neq c\boldsymbol{I}$

m	n1,n2,n3	B	σ_2, σ_3	mancov	bootcov	manLScov	prcv	brcv
5	200,200,200	400	1,1	0.955	0.960	0.954	0.946	0.947
		1000	1,1	0.952	0.949	0.947	0.945	0.946
		400	2,3	0.931	0.959	0.952	0.946	0.946
		1000	2,3	0.932	0.950	0.950	0.941	0.943
5	200,400,600	400	1,1	0.958	0.962	0.954	0.949	0.949
		1000	1,1	0.954	0.954	0.952	0.961	0.960
		400	2,3	0.996	0.958	0.954	0.950	0.953
		1000	2,3	0.993	0.948	0.946	0.957	0.957
10	400,400,400	800	1,1	0.953	0.957	0.948	0.948	0.947
		2000	1,1	0.947	0.947	0.939	0.949	0.947
		800	2,3	0.925	0.952	0.942	0.936	0.937
		2000	2,3	0.931	0.949	0.953	0.949	0.950
10	400,800,1200	800	1,1	0.955	0.966	0.952	0.946	0.948
		2000	1,1	0.961	0.959	0.958	0.952	0.952
		800	2,3	0.998	0.960	0.956	0.948	0.950
		2000	2,3	0.998	0.947	0.957	0.949	0.947
20	800,800,800	1600	1,1	0.950	0.954	0.947	0.948	0.950
		4000	1,1	0.947	0.947	0.943	0.950	0.950
		1600	2,3	0.923	0.954	0.937	0.949	0.951
		4000	2,3	0.930	0.951	0.949	0.949	0.947
20	800,1600,2400	1600	1,1	0.952	0.962	0.945	0.950	0.949
		4000	1,1	0.957	0.958	0.954	0.956	0.955
		1600	2,3	1	0.958	0.951	0.948	0.951
		4000	2,3	0.999	0.948	0.944	0.945	0.944

Table 2: One-Way MANOVA Type Test, Coverage for lognormal data with $\boldsymbol{\Sigma}_3 \neq c\boldsymbol{I}$

m	n1,n2,n3	B	σ_2, σ_3	mancov	bootcov	manLScov	prcv	brcv
5	200,200,200	400	1,1	0.963	0.974	0.957	0.963	0.965
		1000	1,1	0.957	0.962	0.941	0.948	0.951
		400	2,3	0.929	0.954	0.909	0.952	0.955
		1000	2,3	0.925	0.945	0.912	0.959	0.960
5	200,400,600	400	1,1	0.952	0.963	0.925	0.955	0.955
		1000	1,1	0.956	0.961	0.943	0.958	0.958
		400	2,3	0.995	0.964	0.956	0.956	0.957
		1000	2,3	0.986	0.948	0.941	0.950	0.951
10	400,400,400	800	1,1	0.951	0.966	0.938	0.952	0.952
		2000	1,1	0.952	0.957	0.939	0.954	0.955
		800	2,3	0.921	0.955	0.920	0.958	0.959
		2000	2,3	0.920	0.937	0.908	0.962	0.962
10	400,800,1200	800	1,1	0.942	0.957	0.928	0.948	0.946
		2000	1,1	0.951	0.952	0.934	0.956	0.957
		800	2,3	0.997	0.963	0.949	0.957	0.958
		2000	2,3	0.996	0.949	0.946	0.943	0.944
20	800,800,800	1600	1,1	0.946	0.957	0.932	0.946	0.947
		4000	1,1	0.951	0.956	0.940	0.951	0.953
		1600	2,3	0.920	0.949	0.899	0.952	0.953
		4000	2,3	0.917	0.944	0.907	0.953	0.954
20	800,1600,2400	1600	1,1	0.958	0.972	0.950	0.955	0.956
		4000	1,1	0.953	0.958	0.937	0.953	0.953
		1600	2,3	0.999	0.964	0.955	0.956	0.956
		4000	2,3	0.999	0.944	0.944	0.947	0.947

Table 3: One-Way MANOVA Type Test, Coverage for lognormal data with $\boldsymbol{\Sigma}_3 = c\boldsymbol{I}$

m	n1,n2,n3	B	σ_2, σ_3	mancov	bootcov	manLScov	prcv	brcv
5	200,200,200	400	1,1	0.951	0.966	0.936	0.963	0.963
		1000	1,1	0.951	0.963	0.929	0.963	0.961
		400	2,3	0.903	0.936	0.908	0.950	0.950
		1000	2,3	0.904	0.926	0.912	0.956	0.957
5	200,400,600	400	1,1	0.889	0.971	0.928	0.965	0.966
		1000	1,1	0.877	0.957	0.910	0.958	0.958
		400	2,3	0.888	0.947	0.906	0.953	0.952
		1000	2,3	0.898	0.935	0.909	0.959	0.960
10	400,400,400	800	1,1	0.939	0.964	0.914	0.959	0.960
		2000	1,1	0.938	0.954	0.912	0.957	0.957
		800	2,3	0.893	0.941	0.900	0.951	0.952
		2000	2,3	0.907	0.940	0.906	0.968	0.969
10	400,800,1200	800	1,1	0.773	0.965	0.921	0.961	0.961
		2000	1,1	0.772	0.965	0.920	0.970	0.972
		800	2,3	0.856	0.951	0.897	0.957	0.958
		2000	2,3	0.831	0.929	0.890	0.951	0.953
20	800,800,800	1600	1,1	0.935	0.965	0.912	0.962	0.962
		4000	1,1	0.940	0.952	0.910	0.960	0.961
		1600	2,3	0.896	0.937	0.909	0.960	0.959
		4000	2,3	0.885	0.924	0.897	0.954	0.954
20	800,1600,2400	1600	1,1	0.561	0.959	0.910	0.956	0.957
		4000	1,1	0.585	0.948	0.896	0.953	0.953
		1600	2,3	0.795	0.949	0.909	0.955	0.956
		4000	2,3	0.780	0.926	0.898	0.948	0.948

matrix $\boldsymbol{A} = (a_{ij})$ with $a_{ii} = 1$ and $a_{ij} = \psi$ where $0 \leq \psi < 1$ for $i \neq j$. Then the vector $\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{w}_i$ so that $Cov(\boldsymbol{u}_i) = \boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{A}^T = (\sigma_{ij})$ where the diagonal entries $\sigma_{ii} = [1 + (m-1)\psi^2]$ and the off diagonal entries $\sigma_{ij} = [2\psi + (m-2)\psi^2]$. Hence the correlations are $cor(x_i, x_j) = \rho = (2\psi + (m-2)\psi^2)/(1 + (m-1)\psi^2)$ for $i \neq j$ where x_i and x_j are nontrivial predictors. If $\psi = 1/\sqrt{cp}$, then $\rho \to 1/(c+1)$ as $p \to \infty$ where $c > 0$. As ψ gets close to 1, the predictor vectors cluster about the line in the direction of $(1, ..., 1)^T$. Let $Y_i = 1 + 1x_{i,2} + \cdots + 1x_{i,k+1} + e_i$ for $i = 1, ..., n$. Hence $\boldsymbol{\beta} = (1, .., 1, 0, ..., 0)^T$ with $k+1$ ones and $p-k-1$ zeros.

The zero mean iid errors $\tilde{e}_i = \epsilon_i$ were iid from five distributions: i) N(0,1), ii) t_3, iii) EXP(1) - 1, iv) uniform$(-1, 1)$, and v) 0.9 N(0,1) + 0.1 N(0,100). Only distribution iii) is not symmetric. Then wtype = 1 if $e_i = \epsilon_i$ (the WLS model is the OLS model), 2 if $e_i = |\boldsymbol{x}_i^T\boldsymbol{\beta} - 5|\epsilon_i$, 3 if $e_i = \sqrt{(1 + 0.5x_{i2}^2)}\epsilon_i$, 4 if $e_i = \exp[1 + \log(|x_{i2}|) + ... + \log(|x_{ip}|)]\epsilon_i$, 5 if $e_i = [1 + \log(|x_{i2}|) + ... + \log(|x_{ip}|)]\epsilon_i$, 6 if $e_i = [\exp([\log(|x_{i2}|) + ... + \log(|x_{ip}|)]/(p-1))]\epsilon_i$, 7 if $e_i = [[\log(|x_{i2}|) + ... + \log(|x_{ip}|)]/(p-1)]\epsilon_i$, The last four types were special cases of types suggested by Romano and Wolf (2017). For type 6, the weighting function is the geometric mean of $|x_{i2}|, ..., |x_{ip}|$.

When $\psi = 0$ and wtype = 1, the least squares confidence intervals for β_i should have length near $2t_{96,0.975}\sigma/\sqrt{n} \approx 2(1.96)\sigma/10 = 0.392\sigma$ when $n = 100$ and the iid zero mean errors have variance σ^2. The simulation computed the shorth(c) CI for each β_i and used bootstrap confidence regions to test $H_0 : \boldsymbol{\beta}_S = \boldsymbol{1}$ (whether first $k+1$ $\beta_i = 1$) and $H_0 : \boldsymbol{\beta}_E = \boldsymbol{0}$ (whether the last $p-k-1$ $\beta_i = 0$). The nominal coverage was 0.95 with $\delta = 0.05$. Observed coverage between 0.94 and 0.96 suggests coverage is close to the nominal value.

The tables have two rows for each model giving the observed confidence interval coverages and average lengths of the confidence intervals. The terms "npar", "wild", and "par" are for the nonparametric, wild, and parametric bootstrap using 7 slices. The last six columns give results for the tests. The terms pr, hyb, and br are for the prediction region method, hybrid, and Bickel and Ren regions. For terms such as pr0 or br1, the 0 indicates the test was $H_0 : \boldsymbol{\beta}_E = \boldsymbol{0}$, while the 1 indicates that the test was $H_0 : \boldsymbol{\beta}_S = \boldsymbol{1}$. The length and coverage = P(fail to reject H_0) for the interval $[0, D_{(U_B)}]$ or $[0, D_{(U_{BT})}]$ where $D_{(U_B)}$ or $D_{(U_{BT})}$ is the cutoff for the confidence region. The cutoff will often be near $\sqrt{\chi^2_{g,0.95}}$ if the statistic T is asymptotically normal. Note that $\sqrt{\chi^2_{2,0.95}} = 2.448$. Since $\boldsymbol{C}_n = n\boldsymbol{S}^*_T$, we need $B \geq 25p$. For the wild bootstrap and the parametric bootstrap, $\boldsymbol{C}_n = \text{Cov}(\hat{\boldsymbol{\beta}}^*)$ could be used with B near 100.

Pötscher and Preinerstorfer (2021) note that WLS tests tend to reject H_0 too often (liberal tests with undercoverage), and suggest that there is always a WLS model where the wild bootstrap is poor. These tests use a (scaled) χ^2 or F cutoff. Hence simulation results likely depend on the WLS models used.

Rajapaksha (2021) made 90 tables for WLS with $n = 100$ and $B = 200$. The wild bootstrap had the worst undercoverage for 62 tables, the parametric bootstrap for 13 tables, and there was little undercoverage for 15 tables. Coverage less than 0.87 was uncommon in the 90 tables. The mixture distribution sometimes had overcoverage, but was often the distribution where the parametric bootstrap had undercoverage worse than the nonparametric and wild bootstrap. Coverage was often poor for the shifted exponential distribution where n much larger than 100 is often needed. The nonparametric bootstrap gave good coverages for the confidence intervals for β_i while the wild and parametric bootstrap had occasional undercoverage. When the coverage ≥ 0.94, the wild CIs tended to be shorter than the nonparametric CIs which tended to be shorter than the parametric CIs. The nonparametric bootstrap also worked best for the tests for $\psi < 0.9$. When the nonparametric bootstrap had undercoverage for a test or CI, the wild bootstrap tended to have greater undercoverage. The parametric bootstrap performed well for $\psi = 0.9$. In Table 4, the parametric bootstrap was the worst. In Table 5, the nonparametric bootstrap worked best, with occasional overcoverage. The wild and parametric bootstrap had test undercoverage for $\psi = 0$.

6. Conclusions

The theory showing that the bootstrap BR and PR confidence regions give large sample tests is simple if $\sqrt{n}(T_n - \boldsymbol{\mu}) \xrightarrow{D} \boldsymbol{u}$, $\sqrt{n}(T^*_n - T_n) \xrightarrow{D} \boldsymbol{u}$, and $\boldsymbol{C}_n^{-1} \xrightarrow{P} \boldsymbol{C}^{-1}$. An interesting result is that the BR and PR confidence intervals do not depend on whether the wrong or

Table 4: Bootstrapping WLS, n=100, B=200, wtype = 6, p=4, k=p-2, etype= $N(0,1)$

ψ	β_1	β_2	β_{p-1}	β_p	pr0	hyb0	br0	pr1	hyb1	br1
npar,0	0.953	0.943	0.937	0.935	0.929	0.928	0.926	0.924	0.918	0.922
len	0.287	0.361	0.363	0.362	1.959	1.959	1.965	2.835	2.835	2.844
wild,0	0.950	0.936	0.931	0.927	0.918	0.925	0.925	0.908	0.907	0.909
len	0.288	0.356	0.356	0.355	1.930	1.930	1.936	2.732	2.732	2.737
par,0	0.954	0.906	0.906	0.880	0.872	0.874	0.875	0.885	0.884	0.883
len	0.295	0.324	0.325	0.303	1.959	1.959	1.965	2.826	2.826	2.834
npar,0.5	0.950	0.951	0.949	0.955	0.951	0.951	0.951	0.941	0.933	0.935
len	0.444	0.752	0.750	0.750	1.963	1.963	1.969	2.854	2.854	2.863
wild,0.5	0.949	0.943	0.943	0.940	0.932	0.932	0.933	0.907	0.910	0.910
len	0.443	0.728	0.726	0.723	1.925	1.925	1.930	2.726	2.726	2.731
par,0.5	0.948	0.954	0.965	0.950	0.945	0.947	0.946	0.941	0.941	0.942
len	0.454	0.786	0.785	0.761	1.959	1.959	1.964	2.826	2.826	2.832
npar,0.9	0.958	0.951	0.947	0.943	0.941	0.937	0.938	0.950	0.947	0.948
len	0.644	5.292	5.327	5.293	1.965	1.965	1.970	2.855	2.855	2.865
wild,0.9	0.951	0.944	0.946	0.938	0.931	0.932	0.935	0.924	0.926	0.924
len	0.643	5.132	5.143	5.119	1.926	1.926	1.932	2.729	2.729	2.735
par,0.9	0.956	0.955	0.955	0.957	0.950	0.949	0.950	0.938	0.941	0.943
len	0.658	5.447	5.477	5.436	1.957	1.957	1.963	2.823	2.823	2.831

Table 5: Bootstrapping WLS, n=100, B=200, wtype = 6, p=8, k=1, etype= t_3

ψ	β_1	β_2	β_{p-1}	β_p	pr0	hyb0	br0	pr1	hyb1	br1
npar,0	0.946	0.961	0.955	0.940	0.939	0.934	0.935	0.930	0.928	0.928
len	0.253	0.282	0.279	0.279	3.740	3.740	3.754	2.466	2.466	2.475
wild,0	0.934	0.954	0.945	0.939	0.868	0.871	0.873	0.913	0.911	0.910
len	0.250	0.270	0.269	0.269	3.459	3.459	3.468	2.389	2.389	2.394
par,0	0.941	0.954	0.935	0.919	0.875	0.879	0.878	0.919	0.922	0.922
len	0.259	0.283	0.264	0.262	3.627	3.627	3.635	2.457	2.457	2.464
npar,0.5	0.944	0.961	0.944	0.948	0.976	0.974	0.975	0.937	0.935	0.936
len	0.588	1.099	1.094	1.103	3.807	3.807	3.820	2.479	2.479	2.485
wild,0.5	0.935	0.951	0.937	0.946	0.908	0.911	0.908	0.914	0.916	0.918
len	0.576	1.030	1.031	1.036	3.443	3.443	3.450	2.379	2.379	2.383
par,0.5	0.940	0.965	0.948	0.955	0.946	0.946	0.948	0.940	0.940	0.941
len	0.600	1.184	1.129	1.130	3.623	3.623	3.632	2.461	2.461	2.467
npar,0.9	0.933	0.959	0.951	0.954	0.978	0.975	0.976	0.934	0.929	0.931
len	0.985	9.110	9.143	9.152	3.810	3.810	3.823	2.478	2.478	2.487
wild,0.9	0.929	0.949	0.937	0.947	0.919	0.921	0.920	0.913	0.915	0.915
len	0.963	8.600	8.614	8.621	3.448	3.448	3.455	2.378	2.378	2.384
par,0.9	0.932	0.955	0.953	0.957	0.955	0.955	0.957	0.929	0.928	0.931
len	0.999	9.339	9.360	9.383	3.629	3.629	3.639	2.456	2.456	2.464

consistent dispersion matrix was used.

Tests with the wrong dispersion matrix tend to be inferior to tests that use a consistent estimator of the correct covariance matrix if the sample sizes are large enough. Hence tests based on (11) and (13) are better than tests that make the common covariance matrix assumption if the $n_i \geq 20m$ are large enough. A useful diagnostic for tests that make the common covariance matrix assumption is to check whether the test cutoff is close to the bootstrap PR or BR cutoff when $\boldsymbol{C}_n = \boldsymbol{S}_p$. If the n_i are not large or if a test that uses a consistent estimator of the covariance matrix is not available, then the PR and BR tests can be useful, using, for example, $\boldsymbol{C}_n = \boldsymbol{I}$.

The Rupasinghe Arachchige Don and Olive (2019) bootstrap one-way MANOVA type tests needed $B \geq 50m(p-1)$, $n \geq (m+p)^2$, and $n_i \geq 40m$. Large B was needed so $\boldsymbol{S}_T^*$ would be a good estimator when the test statistic $\boldsymbol{T}$ is an $m(p-1) \times 1$ vector.

The new tests can use much smaller B if $\boldsymbol{C}_n^{-1}$ does not depend on the bootstrap sample. The wrong dispersion matrix $\boldsymbol{C}_n = \boldsymbol{I}$ appeared to be useful for one-way MANOVA type tests when the n_i were small or the data was highly skewed. If $(T_n - \boldsymbol{\mu}_0)^T \boldsymbol{I}_g (T_n - \boldsymbol{\mu}_0)$ is used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, we could use $(\boldsymbol{A}T_n - \boldsymbol{\theta}_0)^T \boldsymbol{I}_q (\boldsymbol{A}T_n - \boldsymbol{\theta}_0)$ to test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ if $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\mu}$ and $\boldsymbol{\theta}_0 = \boldsymbol{A}\boldsymbol{\mu}_0$ where $\boldsymbol{A}$ is a $q \times g$ matrix with full rank q, and $\boldsymbol{\mu}$ is a $g \times 1$ population location vector.

Large sample Wald type tests are fairly common, but need large sample sizes. See, for example, Zhang et al. (2016) for the two-way MANOVA model, Duchesne and Francq (2015), Konietschke et al. (2015), and Smaga (2017).

The *R* software was used in the simulations. See R Core Team (2019). Programs were added to the Olive (2017b) collection of *R* functions *mpack.txt* available from (http://parker.ad.siu.edu/Olive/mpack.txt). See Rajapaksha (2021) for more details and simulations.

one-way MANOVA: The function `manovasim` was used to simulate the tests of hypotheses, using the Bates and Maechler (2016) *R* library. See Tables 1-3.

weighted least squares: The function `wildboot` was used to bootstrap the nonparametric, wild, and parametric bootstrap. The function `wlsbootsim` was used for the simulation. See Tables 4 and 5. The function `wlsbootsim2` simulates the wild and parametric bootstrap using $\boldsymbol{C}_n = \text{Cov}(\hat{\boldsymbol{\beta}}^*)$ instead of $\boldsymbol{C}_n = n\boldsymbol{S}_T^*$.

Acknowledgments

The authors thank the referees for their work.

References

Bates, D., and M. Maechler. 2016. *matrix*: Sparse and dense matrix classes and methods. *R* package version 1.2-6. https://CRAN.R-project.org/package=Matrix.

Bickel, P. J., and D. A. Freedman. 1981. Some asymptotic theory for the bootstrap. *The Annals of Statistics* 9 (6):1196-1217. doi:10.1214/aos/1176345637.

Bickel, P. J., and J. J. Ren. 2001. The bootstrap in hypothesis testing. In *State of the art in probability and statistics: festschrift for William R. van Zwet*, ed. M. de Gunst, C. Klaassen, and A. van der Vaart, 91-112. Hayward, CA: The Institute of Mathematical Statistics.

Buja, A., L. Brown, R, Berk, E. George, E. Pitkin, M. Traskin, K. Zhang, and L. Zhao. 2019. Models as approximations I: consequences illustrated with linear regression. *Statistical Science* 34 (4):523-544. doi:10.1214/18-STS693.

Chen, S. X. 2016. Peter Hall's contributions to the bootstrap. *The Annals of Statistics* 44 (5):1821-1836. doi:10.1214/16-AOS1489.

Duchesne, P., and C. Francq. 2015. Multivariate hypothesis testing using generalized and {2}-inverses-with applications. *Statistics* 49 (3):475-496. doi:10.1080/02331888.2014.896917.

Efron, B. 1982. *The jackknife, the bootstrap and other resampling plans.* Philadelphia, PA: SIAM.

Efron, B. 2014. Estimation and accuracy after model selection (with discussion). *Journal of the American Statistical Association* 109 (507):991-1007. doi:10.1080/01621459.2013.823775.

Eicker, F. 1963. Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *The Annals of Mathematical Statistics* 34 (2):447-456. doi:10.1214/aoms/1177704156.

Eicker, F. 1967. Limit theorems for regressions with unequal and dependent errors. In

Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, ed. L. M. Le Cam and J. Neyman, 59-82. Berkeley, CA: University of California Press.

Flachaire, E. 2005. Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics* & *Data Analysis* 49 (2):361-376. doi:10.1016/j.csda.2004.05.018.

Freedman, D. A. 1981. Bootstrapping regression models. *The Annals of Statistics* 9 (6):1218-1228. doi:10.1214/aos/1176345638.

Frey, J. 2013. Data-driven nonparametric prediction intervals. *Journal of Statistical Planning and Inference* 143 (6):1039-1048. doi:10.1016/j.jspi.2013.01.004.

Fujikoshi, Y. (2002). Asymptotic expansions for the distributions of multivariate basic statistics and one-way MANOVA tests under nonnormality. *Journal of Statistical Planning and Inference* 108:263-282. doi:10.1016/S0378-3758(02)00313-0.

Hall, P. 1988. Theoretical comparisons of bootstrap confidence intervals (with discussion). *The Annals of Statistics* 16 (3):927-985. doi:10.1214/aos/1176350933.

Hall, P., and A. Rieck 2001. Improving coverage accuracy of nonparametric prediction intervals. *Journal of the Royal Statistical Society, B* 63 (4):717-725. doi:10.1111/1467-9868.00308.

Hesterberg, T. 2015. What teachers should know about the bootstrap: resampling in the undergraduate statistics curriculum. *The American Statistician* 69 (4):371-386. doi:10.1080/00031305.2015.1089789.

Hinkley, D. V. 1977. Jackknifing in unbalanced situations. *Technometrics* 19 (3):285-292. doi:10.1080/00401706.1977.10489550.

Huber, P. J. 1967. The behavior of maximum likelihood estimation under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, ed. L. M. Le Cam and J. Neyman, 221-223. Berkeley, CA: University of California Press.

Johnson, R. A., and D. W. Wichern 1988. *Applied multivariate statistical analysis.* 2nd ed., Englewood Cliffs, NJ: Prentice Hall.

Kakizawa, Y. 2009. Third-order power comparisons for a class of tests for multivariate linear hypothesis under general distributions. *Journal of Multivariate Analysis* 100 (3):473-496. doi:10.1016/j.jmva.2008.06.002.

Konietschke, F., A. C. Bathke, S. W. Harrar, and M. Pauly. 2015. Parametric and nonparametric bootstrap methods for general MANOVA. *Journal of Multivariate Analysis* 140:291-301. doi:10.1016/j.jmva.2015.05.001.

Long, J. S., and L. H. Ervin. 2000. Using heteroscedasticity consistent standard errors in the linear model. *The American Statistician* 54 (3): 217-224. doi:10.1080/00031305.2000.10474549.

Machado, J. A. F., and P. Parente. 2005. Bootstrap estimation of covariance matrices via the percentile method. *Econometrics Journal* 8 (1):70-78. doi:10.1111/j.1368-423X.2005.00152.x.

MacKinnon, J. G., and H. White. 1985. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29 (3):305-325. doi:10.1016/0304-4076(85)90158-7.

Mammen, E. 1992. Bootstrap, wild bootstrap, and asymptotic normality. *Probability Theory and Related Fields* 93:439-455. doi:10.1007/BF01192716.

Mammen, E. 1993. Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics* 21 (1):255-285. doi:10.1214/aos/1176349025.

Olive, D. J. 2014. *Statistical theory and inference.* New York, NY: Springer.

Olive, D. J. 2017a. *Linear regression.* New York, NY: Springer.

Olive, D. J. 2017b. *Robust multivariate analysis.* New York: Springer.

Olive, D. J. 2018. Applications of hyperellipsoidal prediction regions. *Statistical Papers* 59 (3):913-931. doi:10.1007/s00362-016-0796-1.

Pelawa Watagoda, L. C. R., and D. J. Olive. 2021. Bootstrapping multiple linear regression after variable selection. *Statistical Papers* 62 (2):681-700. doi:10.1007/s00362-019-01108-9.

Pötscher, B. M., and D. Preinerstorfer. 2021. How reliable are bootstrap-based heteroskedasticity robust tests? https://arxiv.org/pdf/2005.04089.pdf.

Rajapaksha, K. W. G. D. H. 2021. Wald Type Tests with the Wrong Dispersion Matrix. (Ph.D. Thesis), Southern Illinois University, USA, at (http://parker.ad.siu.edu/Olive/

skosman.pdf).

R Core Team. 2019. *R: A language and environment for statistical computing.* Vienna: R Foundation for Statistical Computing. www.R-project.org.

Romano, J. P., and M. Wolf. 2017. Resurrecting weighted least squares. *Journal of Econometrics* 197 (1):1-19. doi:10.1016/j.jeconom.2016.10.003.

Rupasinghe Arachchige Don, H. S., and D. J. Olive. 2019. Bootstrapping analogs of the one way MANOVA test. *Communications in Statistics: Theory and Methods* 48 (22):5546-5558. doi:10.1080/03610926.2018.1515363.

Rupasinghe Arachchige Don, H. S., and L. C. R. Pelawa Watagoda. 2018. Bootstrapping analogs of the two sample Hotelling's T^2 test. *Communications and Statistics: Theory and Methods* 47 (9):2172-2182. doi:10.1080/03610926.2017.1337146.

Smaga, L. 2017. Bootstrap methods for multivariate hypothesis testing. *Communications in Statistics: Simulation and Computation* 46 (10):7654-7667. doi:10.1080/03610918.2016.1248573.

Ueki, M., and K. Fueda. 2007. Adjusting estimative prediction limits. *Biometrika* 94 (2):509-511. doi:10.1093/biomet/asm032.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4):817-838. doi:10.2307/1912934.

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* 14 (4):1261-1350. doi:10.1214/aos/1176350142.

Zhang, J.-T., and X. Liu. 2013. A modified Bartlett test for heteroscedastic one-way MANOVA. *Metrika* 76 (1):135-152. doi:10.1007/s00184-011-0379-z.

Zhang, J.-T., B. Zhou, J. Guo, and X. Liu. 2016. A modified Bartlett test for heteroscedastic two-way MANOVA. *Journal of Advanced Statistics* 1 (2):94-108. doi:10.22606/JAS.2016.12006