Wald Type Tests With the Wrong Dispersion Matrix

Kosman W.G.D.H. Rajapaksha and David J. Olive
Southern Illinois University
January 11, 2021

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 \( g \) groups and it is assumed that the population covariance matrices from the \( g \) 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. Two bootstrap confidence regions are modified to obtain large sample Wald type tests with the wrong dispersion matrix.

KEY WORDS: ANOVA, bootstrap, MANOVA, regularized covariance matrix estimator.

1 INTRODUCTION

This section reviews Wald type tests and Wald type tests with the wrong dispersion matrix. Consider testing \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \) where a \( g \times 1 \) statistic \( T_n \) satisfies \( \sqrt{n}(T_n - \theta) \xrightarrow{D} u \sim N_g(0, \Sigma) \). If \( \hat{\Sigma}^{-1} \xrightarrow{P} \Sigma^{-1} \) and \( H_0 \) is true, then

\[
D_n^2 = D_n^2(\hat{\Sigma}) = D_{\hat{\theta}_0}^2(T_n, \hat{\Sigma}/n) = n(T_n - \theta_0)^T \hat{\Sigma}^{-1}(T_n - \theta_0) \xrightarrow{D} u^T \Sigma^{-1} u \sim \chi^2_g
\]

as \( n \to \infty \). Then a Wald type test rejects \( H_0 \) 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. Note that \( D_{\hat{\theta}_0}^2(T_n, \hat{\Sigma}/n) \) is a squared Mahalanobis distance.

It is common to implement a Wald type test using

\[
D_n^2 = D_n^2(C_n) = D_{\hat{\theta}_0}^2(T_n, C_n/n) = n(T_n - \theta_0)^T C_n^{-1}(T_n - \theta_0) \xrightarrow{D} u^T C^{-1} u
\]

as \( n \to \infty \) if \( H_0 \) is true, where the \( g \times g \) symmetric positive definite matrix \( C_n \xrightarrow{P} C \neq \Sigma \). Hence \( C_n \) is the wrong dispersion matrix, and \( u^T C^{-1} u \) does not have a \( \chi^2_g \) distribution.

*Kosman W.G.D.H. Rajapaksha is graduate student and David J. Olive is Professor, Department of Mathematics, Southern Illinois University, Carbondale, IL 62901, USA.
when $H_0$ is true. Often $C_n$ is a regularized estimator of $\Sigma$, or $C_n^{-1}$ is a regularized estimator of the precision matrix $\Sigma^{-1}$, such as $C_n = I_g$ or $C_n = \text{diag}(\Sigma)$. Another example is $C_n = S_p$ where $S_p$ is a pooled covariance matrix, and it is assumed that $k$ groups have the same covariance matrix $\Sigma$. When this assumption is violated, $C_n$ is not a consistent estimator of $\Sigma$. When the bootstrap is used, often $C_n = nS_T^*$ where $S_T^*$ is the sample covariance matrix of the bootstrap sample $T_1^*, ..., T_B^*$. The assumption that $nS_T^*$ is a consistent estimator of $\Sigma$ 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 $\Sigma$ of $\Sigma$. 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) often did better in the simulations.

If $\sqrt{n}(T_n - \theta) \overset{D}{\to} u$ and $\sqrt{n}(T_n^* - T_n) \overset{D}{\to} u$, then the percentiles of $n(T_n - \theta_0)^T C_n^{-1}(T_n - \theta_0)$ can be estimated with the sample percentiles of $n(T_n^* - T_n)^T C_n^{-1}(T_n^* - T_n)$. 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 gives some examples, 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 $C_n^{-1} \overset{P}{\to} C^{-1} \neq \Sigma^{-1}$. Consider testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ where $\theta_0$ is a known $g \times 1$ vector. Then a large sample $100(1 - \delta)%$ confidence region for $\theta$ is a set $A_n$ such that $P(\theta \in A_n)$ is eventually bounded below by $1 - \delta$ as the sample size $n \to \infty$. Then reject $H_0$ if $\theta_0$ is not in the confidence region $A_n$.

For a confidence region, let the $g \times 1$ vector $T_n$ be an estimator of the $g \times 1$ parameter vector $\theta$. Let $T_1^*, ..., T_B^*$ be the bootstrap sample for $T_n$. Let $A$ be a full rank $g \times q$ constant matrix where $g \leq q$, and consider testing $H_0 : A\mu = \theta_0$ versus $H_1 : A\mu \neq \theta_0$ with $\theta = A\mu$ where often $\theta_0 = 0$. Then let $T_n = A\hat{\mu}$ and let $T_i^* = A\hat{\mu}^*$ for $i = 1, ..., B$.

To bootstrap a 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 $C$ be a dispersion estimator. Then the $i$th squared sample Mahalanobis distance is the scalar

$$D_i^2 = D_i^2(T, C) = D^2_{z_i}(T, C) = (z_i - T)^T C^{-1}(z_i - T)$$  \hspace{1cm} (1)

for each observation $z_i$. Notice that the Euclidean distance of $z_i$ from the estimate of center $T$ is $D_i(T, I_g)$ where $I_g$ is the $g \times g$ identity matrix. The classical Mahalanobis distance $D_i$ uses $(T, C) = (\bar{z}, S)$, the sample mean and sample covariance matrix where

$$\bar{z} = \frac{1}{B} \sum_{i=1}^{B} z_i \text{ and } S = \frac{1}{B - 1} \sum_{i=1}^{B} (z_i - \bar{z})(z_i - \bar{z})^T.$$

(2)
Let $q_B = \min(1 - \delta + 0.05, 1 - \delta + b/B)$ for $\delta > 0.1$ and

$$q_B = \min(1 - \delta/2, 1 - \delta + 10\delta b/B), \quad \text{otherwise.}$$ (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 $\theta = A\mu$ and $\mu$ is a $q \times 1$ vector. Let $D_{(U_B)}$ be the 100$q_B$th sample quantile of the $D_i$. Equation (3) is often useful for getting good coverage when $B \geq 50b$. Undercoverage could occur without the correction factor. This result is useful because the bootstrap confidence regions can be slow to simulate. Hence we want to use small values of $B \geq 50b$.

For the following two confidence regions, let a statistic $T = T_n$ estimate $\theta$. Assume $\sqrt{n}(T_n - \theta) D \sim u$ and $\sqrt{n}(T_n - T_n) D \sim u$. Let the bootstrap sample be $T_1^*, ..., T_B^*$. Let $T^*$ and $S_{T^*}$ be the sample mean and sample covariance matrix of the bootstrap sample. The large sample 100$(1 - \delta)$% BR confidence region is

$$\{ w : n(w - T_n)^T C_n^{-1}(w - T_n) \leq D^2_{(U_B, T)} \} = \{ w : D^2_{w}(T_n, C_n/n) \leq D^2_{(U_B, T)} \}$$ (4)

where the cutoff $D^2_{(U_B, T)}$ is the 100$q_B$th sample quantile of the $D^2 = n(T_i^* - T_n)^T C_n^{-1}(T_i^* - T_n)$. Note that the corresponding test for $H_0 : \theta = \theta_0$ rejects $H_0$ if $n(T_n - \theta_0)^T C_n^{-1}(T_n - \theta_0) > D^2_{(U_B, T)}$. Note that $q_B$ is found from (3) with $z_i = T_i^*$.

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

$$\{ w : n(w - T^*)^T C_n^{-1}(w - T^*) \leq D^2_{(U_B)} \} = \{ w : D^2_{w}(T^*, C_n/n) \leq D^2_{(U_B)} \}$$ (5)

where $D^2_{(U_B)}$ is computed from $D_i^2 = n(T_i^* - T^*)^T C_n^{-1}(T_i^* - T^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \theta = \theta_0$ rejects $H_0$ if $n(T^* - \theta_0)^T C_n^{-1}(T^* - \theta_0) > D^2_{(U_B)}$.

If $nC_n^{-1} = [S_{T^*}]^{-1}$, 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 $\theta$. Under regularity conditions, Bickel and Ren (2001) and Olive (2017b, 2018) proved that (4) and (5) are large sample confidence regions when $nC_n^{-1} = [S_{T^*}]^{-1}$. Pelawa Watagoda and Olive (2019) gave simpler proofs.

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

$$\left( \frac{D_{(U_B)}}{D_{(U_B, T)}} \right)^g.$$ (6)

Hence region (5) has smaller volume than region (4) if $D_{(U_B)} < D_{(U_B, T)}$.

Pelawa Watagoda and Olive (2019) showed that under reasonable regularity conditions, i) $\sqrt{n}(T_n - \theta) D \sim u$, ii) $\sqrt{n}(T_i^* - T_n) D \sim u$, iii) $\sqrt{n}(T_i^* - T^*) D \sim u$, and iv) $\sqrt{n}(T_i^* - T^*) D \sim u$. Then

$$D_2^2 = D_{T_i^*}^2(T^*, C_n/n) = \sqrt{n}(T_i^* - T^*)^T C_n^{-1}\sqrt{n}(T_i^* - T^*),$$

$$D_3^2 = D_{\theta}^2(T_n, C_n/n) = \sqrt{n}(T_n - \theta)^T C_n^{-1}\sqrt{n}(T_n - \theta),$$

$$D_4^2 = D_{T_i^*}^2(T_n, C_n/n) = \sqrt{n}(T_i^* - T_n)^T C_n^{-1}\sqrt{n}(T_i^* - T_n),$$

and

$$D_4^2 = D_{T_i^*}^2(T_n, C_n/n) = \sqrt{n}(T_i^* - T_n)^T C_n^{-1}\sqrt{n}(T_i^* - T_n).$$
are well behaved. If $C_n^{-1} \overset{P}{\to} C^{-1}$, then $D_2^2 \overset{D}{\to} D^2 = u^T C^{-1} u$, and (4) and (5) are large sample confidence regions. If $C_n^{-1}$ is “not too ill conditioned” then $D_2^2 \approx u^T C_n^{-1} u$ for large $n$, and the confidence regions (4), and (5) will have coverage near $1 - \delta$.

The basic idea is to use sample percentiles of $D_2^2$ or $D_2^2$ from a bootstrap sample to get better cutoffs for Wald type tests that use the wrong dispersion matrix. The cutoffs can be used for confidence intervals if $g = 1$.

If $g = 1$, then confidence intervals are special cases of confidence regions. Suppose the parameter of interest is $\theta$, and there is a bootstrap sample $T_1^*, \ldots, T_B^*$ where the statistic $T = T_n$ is an estimator of $\theta$ based on a sample of size $n$. Let $T_{(1)}^*, \ldots, T_{(n)}^*$ be the order statistics of the bootstrap method. The bootstrap large sample $100(1 - \delta)\%$ percentile confidence interval (CI) for $\theta$ is an interval $[T_{(k_1)}^*, T_{(K_u)}^*]$ containing $U_B \approx [B(1 - \delta)]$ of the $T_i^*$. Let $k_1 = [B \delta/2]$ and $k_2 = [B(1 - \delta/2)]$. A common choice is

$$[T_{(k_1)}^*, T_{(k_2)}^*].$$

See Efron (1982) and Chen (2016).

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

$$[T_{(s)}^*, T_{(s+c-1)}^*]$$

uses the interval $[T_{(1)}^*, T_{(c)}^*], [T_{(3)}^*, T_{(c+2)}^*], \ldots, [T_{(B-c+1)}^*, T_{(B)}^*]$ of shortest length. Here

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

The shorth CI is obtained by applying the Frey (2013) prediction interval to the bootstrap sample. The shorth CI is the shortest percentile CI covering $c_n$ 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. 238, 2017b: p. 168, 2018) recommended using the shorth CI for the percentile CI.

If $a_i = |T_i^* - T^*|$, then the CI corresponding to (4) is $[T^* - a_{(U_B)}, T^* + a_{(V_B)}]$, which is a percentile CI centered at $T^*$ 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 $\hat{T}^*$ is asymptotically normal. The CI $[T_n - b_{(U_B,T)}, T_n + b_{(V_B,T)}]$ corresponding to (5) is a percentile CI centered at $T_n$ just long enough to cover “$U_B, T^*$” 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 $C_n$, $D_{(U_B)}$, or $D_{(U_B,T)}$. Hence these CIs correspond to the prediction region method CI and the modified Bickel and Ren CI. Suppose $\sqrt{n}(\hat{\mu} - \mu) \overset{D}{\to} N_0(0, \Sigma)$. Then confidence regions (4) and (5) do not depend on whether $C_n^{-1}$ or $d_n C_n^{-1}$ is used if the scalar $d_n > 0$. Let $\theta = a^T \mu$ and $T_n = a^T \hat{\mu}$. Then $a^T C_n^{-1} a = d_n a^T \Sigma^{-1} a$ where $d_n = a C_n^{-1} a / a^T \Sigma^{-1} a$. Hence the confidence intervals do not depend on whether the wrong dispersion matrix is used.

As noted by Pelawa Watagoda and Olive (2019), if $g = 1$, if $\sqrt{n}(T_n - \theta) \overset{D}{\to} U$, and if $\sqrt{n}(T_i^* - T_n) \overset{D}{\to} U$ 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 short confidence interval (8), and the “usual” percentile method confidence interval (7) are asymptotically equivalent (use the central proportion of the bootstrap sample, asymptotically).

For high dimensional data, \( n/g \) is not large, and in the simulations with \( n/g \) not large, the tests with \( C_n \equiv I \) tended to be conservative (if the nominal level for rejecting \( H_0 \) was \( \alpha \) when \( H_0 \) is true, then the simulated proportion \( \alpha_n \) of rejections tended to be near or less than \( \alpha \)). If \( \mu = E(y_i) \), then \( n||y - \mu||^2 D u^T u \) and \( n||y - \mu||^2 D u^T u \). Thus if \( n/g \) is large, \( ||y - \mu||^2 \) and \( ||y - \mu||^2 \) have approximately the same percentiles. If \( E(y) = \mu \) and \( \text{Cov}(y) = \Sigma/n \), then for the nonparametric bootstrap, \( E(y^*) = \bar{y} \) and \( \text{Cov}(y^*) = S_M/n \), where \( S_M = (n - 1)S/n \) is the method of moments estimator of the covariance matrix \( \Sigma = \Sigma_y \).

The following geometric argument is useful. Let \( T_1, ..., T_B \) be an iid sample of the statistic \( T = T_n \), and let \( T^*_1, ..., T^*_B \) be the bootstrap sample. Applying the PR region to the iid sample gives a prediction region: \( R_p = \{ w : D^2_w(T, C_n/n) \leq D^2_{(U_n)} \} \) centered at \( T \) contains a future value of the statistic \( T_f \) with probability \( 1 - \delta_B \) which is eventually bounded below by \( 1 - \delta \) as \( B \to \infty \). Hence the region \( R_c = \{ w : D^2_w(T, C_n/n) \leq D^2_{(U_n)} \} \) is a large sample \( 100(1 - \delta)\% \) confidence region for \( \theta \) where \( T_n \) is a randomly selected \( T_i \). This result holds since the region \( R_c \) contains \( T \) with probability \( 1 - \delta_B \). Suppose \( \sqrt{n}(T_n - \theta) D u \) with \( E(u) = 0 \) and \( \text{Cov}(u) = \Sigma_u \neq 0 \). Then \( T \) gets arbitrarily close to \( \theta \) compared to \( T_n \) as \( B \to \infty \). If \( \sqrt{n}(T_n - \theta) \) and \( \sqrt{n}(T^*_n - T) \) both converge in distribution to \( u \sim N_g(0, \Sigma_A) \), say, then the bootstrap sample data cloud of \( T^*_1, ..., T^*_B \) is like the data cloud of iid \( T_1, ..., T_B \) shifted to be centered at \( T_n \) when \( n/g \) is large.

Assume that the data is high dimensional, so \( n/g \) is not large. Let \( T_n = \bar{y} \) and \( n \geq 50 \) so \( (n - 1)/n \geq 0.98 \). Then \( E(S_M) = (n - 1)\Sigma/n \approx \Sigma \), but \( S_M \) is highly variable. If many bootstrap data clouds are generated, the average dispersion is about \( \Sigma \), but the bootstrap data clouds are much more variable than the iid data clouds. Let \( \delta_n \leq 0.1 \) and consider covering \( 100(1 - \delta_n)\% \) of the iid data and the bootstrap data with hyperspheres centered at \( T \) and \( T^* \), respectively. Heuristically, due to the variability of the bootstrap data clouds, we expect the PR bootstrap cutoff \( D^2(T, I/n) \) to tend to be larger than the iid cutoff \( D^2(T, I/n) \), resulting in conservative tests. Equivalently, we expect the upper sample percentiles of \( ||T^* - \bar{y}||^2 \) to tend to be larger than those of \( ||y - \mu||^2 \). Since the argument is heuristic, the high dimensional “tests” in this paper have not yet been proven to be conservative.

### 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 \( y_{ij} = (Y_{ij1}, ..., Y_{ijm})^T \) are measured for the \( i \)th treatment group and the \( j \)th case in the group. Hence \( i = 1, ..., p \) and \( j = 1, ..., n_i \). Assume the \( p \) treatments have possibly different population location vectors \( \mu_i \), such as \( E(y_{ij}) = \mu_i \). Coordinatewise population
medians and coordinatewise population trimmed means are also useful. Then a one-way MANOVA type test is used to test $H_0 : \mu_1 = \mu_2 = \cdots = \mu_p$ versus the alternative that not all of the $\mu_i$ are equal.

Large sample theory can be be used to derive Wald type tests. Let $\text{Cov}(y_{ij}) = \Sigma_i$ be the nonsingular population covariance matrix of the $i$th treatment group or population. To simplify the large sample theory, assume $n_i = \pi_in$ where $0 < \pi_i < 1$ and $\sum_{i=1}^p \pi_i = 1$. Let $T_i$ be a multivariate location estimator such that $\sqrt{n}(T_i - \mu_i) \sim N_m(0, \Sigma_i)$, and $\sqrt{n}(T_i - \mu_i) \sim N_m\left(0, \frac{\sum_i}{\pi_i}\right)$. Let $T = (T_1^T, T_2^T, \ldots, T_p^T)^T$, $\nu = (\nu_1, \nu_2, \ldots, \nu_p)^T$, and $A$ be a full rank $r \times mp$ matrix with rank $r$, then a large sample test of the form $H_0 : A\nu = \theta_0$ versus $H_1 : A\nu \neq \theta_0$ uses

$$A\sqrt{n}(T - \nu) \sim u \sim N_r\left(0, A \text{ diag}\left(\frac{\Sigma_1}{\pi_1}, \frac{\Sigma_2}{\pi_2}, \ldots, \frac{\Sigma_p}{\pi_p}\right) A^T\right).$$

Let the Wald type statistic

$$t_0 = [AT - \theta_0]^T A \text{ diag}\left(\frac{\Sigma_1}{n_1}, \frac{\Sigma_2}{n_2}, \ldots, \frac{\Sigma_p}{n_p}\right) A^T [AT - \theta_0]^{-1}.$$

These results prove the following theorem.

**Theorem 1.** Under the above conditions, $t_0 \sim \chi^2_r$ if $H_0$ is true.

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

This one-way MANOVA type test is due to Rupasinghe Arachchige Don and Olive (2019), and a special case was used by Zhang and Liu (2013) and Konietzcke et al. (2015) with $\bar{T}_i = \bar{y}_i$ and $\Sigma_i = S_i$, the sample covariance matrix corresponding to the $i$th treatment group. The $p = 2$ case gives analogs to the two sample Hotelling’s $T^2$ test. See Rupasinghe Arachchige Don and Pelawa Wattagoda (2018).

Several important tests use the the common covariance matrix assumption $\Sigma_i \equiv \Sigma$ for $i = 1, \ldots, p$. The test gives a Wald type test 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 $A$ be the $m(p - 1) \times mp$ block matrix

$$A = \begin{bmatrix} I & 0 & 0 & \cdots & -I \\ 0 & I & 0 & \cdots & -I \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & I & -I \end{bmatrix}.$$
Let $\mathbf{\mu}_i \equiv \mathbf{\mu}$, let $H_0 : \mathbf{\mu}_1 = \cdots = \mathbf{\mu}_p$ or, equivalently, $H_0 : \mathbf{A}\mathbf{\nu} = \mathbf{0}$, and let

$$w = \mathbf{A}T = \begin{bmatrix} T_1 - T_p \\ T_2 - T_p \\ \vdots \\ T_{p-2} - T_p \\ T_{p-1} - T_p \end{bmatrix}.$$  \hfill (12)

Then $\sqrt{n}w \overset{D}{\rightarrow} N_{m(p-1)}(0, \Sigma)$ if $H_0$ is true with $\Sigma = (\Sigma_{ij})$ where $\Sigma_{ij} = \frac{\Sigma_{p}}{\pi_{p}}$ for $i \neq j$, and $\Sigma_{ii} = \frac{\Sigma_{i}}{\pi_{i}} + \frac{\Sigma_{p}}{\pi_{p}}$ for $i = j$. Hence

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

as the $n_i \rightarrow \infty$ if $H_0$ is true. Here $\frac{\hat{\Sigma}_i}{n}$ is a block matrix where the off diagonal block entries equal $\hat{\Sigma}_{p}/n_p$ and the $i$th diagonal block entry is $\frac{\hat{\Sigma}_{i}}{n_i} + \frac{\hat{\Sigma}_{p}}{n_p}$ for $i = 1, \ldots, (p-1)$. Reject $H_0$ if

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

where $d_n = \min(n_1, \ldots, n_p)$. 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 $i = 1, \ldots, p$.

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

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

If $H_0$ is true but the $\Sigma_i$ are not equal, we get a bootstrap cutoff by using

$$t^*_{0i} = [\mathbf{A}T_i^* - \mathbf{A}T]^T \left[ \mathbf{A} \operatorname{diag} \left( \frac{\hat{\Sigma}}{n_1}, \frac{\hat{\Sigma}}{n_2}, \ldots, \frac{\hat{\Sigma}}{n_p} \right) \mathbf{A}^T \right]^{-1} \left[ \mathbf{A}T_i^* - \mathbf{A}T \right] \overset{D}{\rightarrow} \chi^2_{m(p-1)}.$$

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

For Theorem 2, Rajapaksha (2021) proved that $t_0 = (n-p)U$, and Fujikoshi (2002) also showed $(n-p)U \overset{D}{\rightarrow} \chi^2_{m(p-1)}$.

**Theorem 2.** For the one way MANOVA test using $\mathbf{\theta}_0 = \mathbf{0}$, $\mathbf{A}$ as defined above Equation (12), and $T_i = \mathbf{\bar{y}}_i$, $$(n-p)U = t_0 = [\mathbf{A}T]^T \left[ \mathbf{A} \operatorname{diag} \left( \frac{\hat{\Sigma}}{n_1}, \frac{\hat{\Sigma}}{n_2}, \ldots, \frac{\hat{\Sigma}}{n_p} \right) \mathbf{A}^T \right]^{-1} \left[ \mathbf{A}T \right]$$
where \( U \) is the Hotelling Lawley trace statistic. Hence if the \( \Sigma_i \equiv \Sigma \) and \( H_0 : \mu_1 = \cdots = \mu_p \) is true, then \((n - p)U = t_0 \xrightarrow{D} \chi^2_{m(p-1)}\).

### 4 SIMULATIONS FOR SOME EXAMPLES

For the bootstrap tests, the nonparametric bootstrap was used. Hence \( n_i \) samples with replacement were taken from the \( i \)th group of \( n_i \) cases. Let \([x]\) be the smallest integer \( \geq x \), e.g. \([7.7]\) = 8. Let \([x]\) be the greatest integer \( \leq x \), e.g. \([7.7]\) = 7. Variants on the denominator degrees of freedom are common in the literature. More details and simulations for examples of this section are in Rajapaksha (2021).

**Example 1.** We simulated six good alternatives to the poor pooled \( t \) CI, and the Welch CI performed the best in the simulations. The PR, BR, and shorth CIs were used for the bootstrap tests, the nonparametric bootstrap was used. Hence if \( \mu \) is the sample (1)\( \mu \) CI for \((X_{11} \cdots X_{1n})\) and Olive (2014, pp. 278-279, 290). The large sample \((1 - \alpha)100\% \) modified pooled \( t \) CI for \((\mu_1 - \mu_2)\) is

\[
\overline{X} - \overline{Y} \pm t_{n_1+n_2-4,1-\alpha/2} \hat{\tau} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}. \tag{14}
\]

The pooled \( t \) CI replaces \( \hat{\tau} \) by 1 and \( n_1 + n_2 - 4 \) by \( n_1 + n_2 - 2 \). See DasGupta (2008, pp. 402-404) and Olive (2014, pp. 278-279, 290). The large sample \((1 - \alpha)100\% \) Welch CI for \((\mu_1 - \mu_2)\) is

\[
\overline{X} - \overline{Y} \pm t_{d,1-\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \tag{15}
\]

where \( d = n_1 + n_2 - 4, d = [d_0], \) or \( d = \max(1, [d_0]) \) with

\[
d_0 = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{1}{n_1-1} (\frac{S_1^2}{n_1})^2 + \frac{1}{n_2-1} (\frac{S_2^2}{n_2})^2}.
\]

The Welch CI is a Wald type CI based on Equations (11) and (13) but with a modified cutoff. See Welch (1937) and Yuen (1974). See Rajapaksha (2021) for the seventh CI.

**Example 2.** It has been suggested that the one way ANOVA \( F \) test results will be approximately correct if \( \max(S_1, \ldots, S_p) \leq 2 \min(S_1, \ldots, S_p) \) or if the sample sizes \( n_i \) are all equal to \( k \) where \( pk = n \). See Moore (2007, p. 634) and Harwell et al. (1992). Confidence intervals have much less resistance to the constant variance assumption. Consider confidence intervals for \( \mu_i \) such as \( \overline{y}_{i0} \pm t_{n_i-1,1-\delta/2} \sqrt{\text{MSE}} / \sqrt{n_i} \). MSE is a weighted average of the \( S_i^2 \). Hence MSE overestimates small \( \sigma^2_i \) and underestimates large \( \sigma^2_i \) when the \( \sigma^2_i \) are not equal.

We recommend using the Theorem 1 test for \( H_0 : A\mu = \theta_0 \) where \( \theta = A\mu \) and \( \mu = (\mu_1, \ldots, \mu_p) \). For \( H_0 : \mu_1 = \cdots = \mu_p \), we used six alternatives to the one way ANOVA \( F \) test, including the next three tests. The Welch (1951) ANOVA \( F \) test uses test statistic

\[
F_W = \frac{\sum_{i=1}^{p} w_i (\overline{Y}_{i0} - \overline{Y}_{00})^2 / (p-1)}{1 + \frac{(p-2)}{p-1} \sum_{i=1}^{p} (1 - \frac{w_i}{w})^2 / (n_i - 1)}
\]
where \(w_i = n_i/S_i^2\), \(u = \sum_{i=1}^p w_i\) and \(\bar{Y}_{60} = \sum_{i=1}^p w_i \bar{y}_{i60}/u\). Then the test statistic is compared to an \(F_{p-1,d_W}\) distribution where \(d_W = [f]\) and

\[
1/f = \frac{3}{p^2 - 1} \sum_{i=1}^p \left(1 - \frac{w_i}{u}\right)^2/(n_i - 1).
\]

The PR bootstrap test is described above Theorem 2 with \(m = 1\). Then \(F_0 = t_0/(n-p)\) is equal to the one way ANOVA F statistic. Then \(F_{0,i}^* = (w_i^* - w)^T D^{-1}(w_i^* - w)/((p - 1)S_p^2)\) where \(S_p = MSE\) and

\[ D = C/MSE = \text{diag}\left(\frac{1}{n_1}, \ldots, \frac{1}{n_{p-1}}\right) + \frac{1}{n_p} 11^T. \]

Let \(F_{0,p(w,p)}\) be the 100\(U_B\)th percentile of the \(F_{0,i}^*\) using with \(b = p - 1\) in Equation (3). The large sample F test corresponding to (13) with \(m = 1\) uses \(F_L = t_0/(p-1) = w^T C^{-1} w/(p-1)\) where

\[ C = \text{diag}\left(\frac{S_1^2}{n_1}, \ldots, \frac{S_{p-1}^2}{n_{p-1}}\right) + \frac{S_p^2}{n_p} 11^T. \]

Then \(F(n-p, \min(n_i), 1-\delta)\) is the cutoff. In the simulations, \(F_W\) and \(F_L\) appeared to be equal, but the two tests used different denominator degrees of freedom in the F cutoff. The PR bootstrap test performed better than the ANOVA F test if the \(n_i\) were not too small, but the tests based on \(F_W\) and \(F_L\) were much better. Tests with the wrong dispersion matrix are outperformed by tests that use consistent estimators for the correct dispersion matrix if the sample sizes \(n_i\) are large enough.

**Example 3.** Next we consider one sample Hotelling’s \(T^2\) type test where \(n/m\) may not be large. Suppose there is a random samples \(y_1, \ldots, y_n\), and that it is desired to test \(H_0 : \mu = \mu_0\) versus \(H_1 : \mu \neq \mu_0\) where \(\mu\) is a \(m \times 1\) vector, and \(m > n\) is possible. We will use \(\mu = E(y_i)\) with test statistic \(T_n = \bar{y}\) and the bootstrapped test statistic \(T^* = \bar{y}^*\). We will also use \(T_n = \text{coordinatewise median}\) where \(\mu\) is the population coordinatewise median. We will use \(C_n = C_n^{-1} = I_m\). Let \(\theta = \mu_0 = 0\).

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

\[
\{w : (w - T_n)^T C_n^{-1} (w - T_n) \leq D^2_{(U_B,T)}\} = \{w : D^2_{w}(T_n, I) \leq D^2_{(U_B,T)}\} \tag{16}
\]

where the cutoff \(D^2_{(U_B,T)}\) is the 100\((1 - \alpha)\)th sample quantile of the squared Euclidean distance \(D_i^2 = (T_i^* - T_n)^T (T_i^* - T_n)\). Note that the corresponding test for \(H_0 : \theta = 0\) rejects \(H_0\) if \((T_n - 0)^T (T_n - 0) > D^2_{(U_B,T)}\).

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

\[
\{w : (w - \bar{T})^T C_n^{-1} (w - \bar{T}) \leq D^2_{(U_B)}\} = \{w : D^2_{w}(\bar{T}, I) \leq D^2_{(U_B)}\} \tag{17}
\]

where the cutoff \(D^2_{(U_B)}\) is the 100\((1 - \alpha)\)th sample quantile of the squared Euclidean distance \(D_i^2 = (T_i^* - \bar{T})^T (T_i^* - \bar{T})\) for \(i = 1, \ldots, B\). Note that the corresponding test for \(H_0 : \theta = 0\) rejects \(H_0\) if \((\bar{T}^* - 0)^T (\bar{T}^* - 0) > D^2_{(U_B)}\).
The test uses the result that \( \sqrt{n}(\mathbf{y} - \mathbf{u}) \xrightarrow{D} N_m(0, \Sigma \mathbf{y}) \) and \( \sqrt{n}(\mathbf{y'} - \mathbf{y}) \xrightarrow{D} N_m(0, \Sigma \mathbf{y}) \). Since \( \mathbf{I} \) is independent of the bootstrap sample, correction factors for the bootstrap cutoffs were not needed. Since the sample quantile is that of a random variable, \( B \) does not need to be large. If \( \Sigma \mathbf{y} = \mathbf{I} \), then

\[
(y - \mu)^T \mathbf{I}^{-1}(y - \mu) \approx \frac{1}{n} \chi^2_m
\]

since

\[
n(y - \mu)^T \mathbf{I}^{-1}(y - \mu) \xrightarrow{D} \chi^2_m
\]

as \( n \to \infty \). For high dimensional data with \( m \geq n \), the test has not yet been proven to be conservative, but \( E(\mathbf{y}) = \mu \), \( \text{Cov}(\mathbf{y}) = \Sigma \mathbf{y} / n \), \( E(\mathbf{y'}) = \mathbf{y} \), and \( \text{Cov}(\mathbf{y'}) = (n - 1) \mathbf{S} / n^2 \).

The simulations used \( \mathbf{y} = \mathbf{B} \mathbf{x} \) with \( \mathbf{x} \sim N_p(0, \mathbf{I}) \), \( \mathbf{x} \sim 0.6N_p(0, \mathbf{I}) + 0.4N_p(0, 25 \mathbf{I}) \), \( \mathbf{x} \) with a multivariate \( t_4 \), and a multivariate lognormal distribution shifted to have zero mean where \( \mathbf{x} = (x_1, \ldots, x_p) \) with \( w_i = \exp(Z) \) where \( Z \sim N(0, 1) \) and \( x_i = w_i - E(w_i) \) where \( E(w_i) = \exp(0.5) \). We used covariance matrix types 1 if \( \mathbf{B} = \mathbf{I} \), 2 if \( \mathbf{B} = \text{diag}(\sqrt{1}, \ldots, \sqrt{p}) \), and 3 if \( (Y_i, Y_j) = \rho \) where \( \rho = 0 \) if \( \psi = 0 \), \( \rho \to 1/(c+1) \) as \( p \to \infty \) if \( \psi = 1/\sqrt{cp} \) where \( c > 0 \), and \( \rho \to 1 \) as \( p \to \infty \) if \( \psi \in (0, 1) \) is a constant. \( E(\mathbf{x}) = \delta \mathbf{1} \) where \( \mathbf{1} \) is the \( p \times 1 \) vector of ones.

The first three distributions have mean \( \mu = E(\mathbf{y}) \) equal to the population coordinatewise median since the distributions are elliptically contoured distributions with center \( \mu \). For the lognormal distribution, if \( H_0 : \mu = 0 \) is true for \( \mu = E(\mathbf{y}) \), then \( H_0 \) is false if \( \mu \) is the population coordinatewise median. Then for this distribution the coverage is the power rather than the level, and power near 1 is good.

The simulation used 5000 runs, \( n = 100 \), and \( m = 10, 100, 200, 400 \). We used \( \psi = 1/\sqrt{p} \) for \( \mathbf{B} \) of type 3, and \( \delta = 0 \) or \( \delta = 1 \). The test appears to be conservative when \( n/m \) is not large, but we have not proven that the high dimensional test is conservative.

Table 1: HD One Sample Hotelling’s \( T \) Type Test, \( m=10, n=100, \delta = 0, \mathbf{B}=100 \)

<table>
<thead>
<tr>
<th>( \mathbf{x} )</th>
<th>covtyp</th>
<th>pr</th>
<th>br</th>
<th>pr med</th>
<th>br med</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1</td>
<td>0.1116</td>
<td>0.1004</td>
<td>0.0438</td>
<td>0.0486</td>
</tr>
<tr>
<td>Mix</td>
<td>1</td>
<td>t</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mix</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mix</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10
Example 4. The two sample Hotelling’s $T^2$ type test where $n/m$ may not be large is similar to Example 3. Suppose there are two independent random samples $y_{1,1}, \ldots, y_{n_1,1}$ and $y_{1,2}, \ldots, y_{n_2,2}$ from two populations or groups, and that it is desired to test $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ where $\mu_i$ are $m \times 1$ vectors. We will use $\mu_i = E(y_i)$, and $m > n_i$ is possible. Let the test statistic $T_n = \overline{y}_1 - \overline{y}_2$ and the bootstrapped test statistic $T^* = \overline{y}_1^* - \overline{y}_2^*$. We will also use the difference of coordinatewise medians. We will use $C_n = C_n^{-1} = I_m$. Let $\theta = \mu_1 - \mu_2$. Then the PR and BR confidence regions are as in Example 3.

If $\mu_1 = \mu_2$, $\Sigma y_i = I$, and $n_1 = n_2 = k$, then

$$(\overline{y}_1 - \overline{y}_2)^T I^{-1} (\overline{y}_1 - \overline{y}_2) \approx \frac{2}{k} \chi^2_m$$

since

$$(\overline{y}_1 - \overline{y}_2)^T (2I/k)^{-1} (\overline{y}_1 - \overline{y}_2) \overset{D}{\rightarrow} \chi^2_m$$

as $k \rightarrow \infty$.

Four types of data distributions $x_i$ from Example 3 were considered that were identical for $i = 1, 2$. Then $y_1 = A x_1 + \delta 1$ and $y_2 = \sigma B x_2$. We used $A = B = diag(1, \sqrt{2}, \ldots, \sqrt{m})$, $A = B = I$, and $A = I$ with $B = diag(1/\sqrt{2}, \ldots, \sqrt{m})$. Note that $\text{Cov}(y_2) = \sigma^2 \text{Cov}(y_1)$ when $A = B$, and $E(y_i) = E(x_i) = 0$ if $\delta = 0$.

CONCLUSIONS

The theory showing that the bootstrap BR and PR confidence regions give large sample tests is very simple. We need $\sqrt{n}(T_n - \mu) \overset{D}{\rightarrow} u$, $\sqrt{n}(T^*_n - T_n) \overset{D}{\rightarrow} u$, and $C_n^{-1} \overset{P}{\rightarrow} C^{-1}$. The results also hold if $G_n = C_n^{-1}$ and $G = C^{-1}$ where $G_n$ and $G$ are not necessarily nonsingular. An interesting result is that the BR and PR confidence intervals do not depend on whether the wrong or right 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. If the $n_i$ are not large or if a test that uses a consistent estimator of the correct covariance matrix is not available, then the PR and BR tests can be useful.

Tests using the wrong dispersion matrix $I$ are useful to illustrate the previous paragraph. If $\sqrt{n}(T_n - \theta) \overset{D}{\rightarrow} N_g(0, \Sigma)$, then $\{w : (w - T_n)^T \Sigma (w - T_n) \leq D^2\}$ is a hyperellipsoid. A hypersphere could be used to approximately cover the hyperellipsoid, but the power from the hypersphere test will be lower than that from the hyperellipsoid test if the sample sizes $n_i$ are large. For large sample level $\alpha$ tests, the hypersphere volume tends to be much larger than the hyperellipsoid volume unless $\Sigma \approx k I$ for some real $k > 0$.

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 $S_T^*$ would be a good estimator when the test statistic $T$ is an $m(p - 1) \times 1$ vector. The new tests can use much smaller $B$ if $C_n^{-1}$ does not depend on the bootstrap sample.
For high dimensional tests with data $\mathbf{x}_{ij} = (x_{ij1}, \ldots, x_{ijm})^T$ where $i = 1, \ldots, n$ and $j = 1, \ldots, g$, using $\mathbf{y}_{ij} = \mathbf{x}_{ij}$ may not work as well as using $\mathbf{y}_{ij} = \mathbf{z}_{ij} = (x_{ij1}/S_1(j)^2, \ldots, x_{ijm}/S_m(j)^2)^T$, where $S_i(j)^2 = S_{ii}(j)$ when the $m \times m$ sample covariance matrix $S_j = (S_{ik}(j))$ for the $j$th group. Other choices of $C_n$ than $C_n = I$ can be used as long as the computational complexity of $C_n^{-1}$ is not too high.

Some high dimensional one sample tests include Chen et al. (2011), Hyodo and Nishiyama (2017), Srivastava and Du (2008), and Wang, Peng, and Li (2015).

The $R$ software was used in the simulations. See R Core Team (2016). Programs were added to the Olive (2017b) collection of $R$ functions mpack.txt available from (http://parker.ad.siu.edu/Olive/mpack.txt).

**pooled $t$ test:** The function pcisim2 was used in the simulations.  
**one way ANOVA:** The function anovasim2 was used in the simulations.  
**high dimensional one sample Hotelling’s $T^2$ test:** The function hdhot1wsim was used for high dimensional data.  
**two sample Hotelling’s $T^2$ test:** The function hot2sim was used to simulate the tests of hypotheses, and predreg computes the confidence region given the bootstrap values from rhot2boot. The Curran (2013) $R$ package Hotelling was used to perform the classical 2 sample Hotelling’s $T^2$ test. The function hdhot2wsim was used for high dimensional data.  
**one way MANOVA:** The function manbtsim4 was used to simulate the tests of hypotheses, and predreg computes the confidence region given the bootstrap values.  

**References**


