A High Dimensional Omnibus Regression Test

Ahlam M. Abid, Paul A. Quaye, and David J. Olive*
Southern Illinois University

July 5, 2025

Abstract

Consider regression models where the response variable Y only depends on the $p \times 1$ vector of predictors $\boldsymbol{x} = (x_1, ..., x_p)^T$ through the sufficient predictor $SP = \alpha + \boldsymbol{x}^T \boldsymbol{\beta}$. Let the covariance vector $\text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Assume the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are independent and identically distributed random vectors for $i = 1, ..., n$. Then for many such regression models, $\boldsymbol{\beta} = \boldsymbol{0}$ if and only if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$ where $\boldsymbol{0}$ is the $p \times 1$ vector of zeroes.

The test of $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq \boldsymbol{0}$ is equivalent to the high dimensional one sample test $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{0}$ applied to $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$ where $\boldsymbol{u}_i = \boldsymbol{x}_i(Y_i - \mu_Y)$ and the expected value $E(Y) = \mu_Y$. Since μ_Y is unknown, the test of $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{0}$ is implemented by applying the one sample test to $\boldsymbol{z}_i = \boldsymbol{x}_i(Y_i - \overline{Y})$ for $i = 1, ..., n$.

KEY WORDS: Generalized Linear Models, Multiple Linear Regression, One Sample Test, Two Sample Test, U-Statistics.

1 Introduction

This section reviews regression models where the response variable Y depends on the $p \times 1$ vector of predictors $\boldsymbol{x} = (x_1, ..., x_p)^T$ only through the sufficient predictor $SP = \alpha + \boldsymbol{x}^T \boldsymbol{\beta}$. Then there are n cases $(Y_i, \boldsymbol{x}_i^T)^T$. For the regression models, the conditioning and subscripts, such as i, will often be suppressed. This paper gives a high dimensional test for $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{0}$ where $\boldsymbol{0} = (0, ..., 0)^T$ is the $p \times 1$ vector of zeroes.

A useful multiple linear regression model is $Y|\boldsymbol{x}^T\boldsymbol{\beta} = \alpha + \boldsymbol{x}^T\boldsymbol{\beta} + e$ or $Y_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ or

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{1}$$

for $i = 1, ..., n$. Assume that the e_i are independent and identically distributed (iid) with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma^2$. In matrix form, this model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}, \tag{2}$$

*David J. Olive is Professor, School of Mathematical & Statistical Sciences, Southern Illinois University, Carbondale, IL 62901, USA.

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times (p+1)$ matrix with ith row $(1, \boldsymbol{x}_i^T)$, $\boldsymbol{\phi} = (\alpha, \boldsymbol{\beta}^T)^T$ is a $(p+1) \times 1$ vector , and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. Also $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$ where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

For a multiple linear regression model with heterogeneity, assume model (1) holds with $E(\boldsymbol{e}) = \boldsymbol{0}$ 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. Under regularity conditions, the ordinary least squares (OLS) estimator $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$ can be shown to be a consistent estimator of $\boldsymbol{\beta}$.

For estimation with ordinary least squares, let the covariance matrix of $\boldsymbol{x}$ be $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma_x} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T] = E(\boldsymbol{x}\boldsymbol{x}^T) - E(\boldsymbol{x})E(\boldsymbol{x}^T)$ and $\boldsymbol{\eta} = \text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma_{x}}_Y = E[(\boldsymbol{x} - E(\boldsymbol{x})(Y - E(Y))] = E(\boldsymbol{x}Y) - E(\boldsymbol{x})E(Y) = E[(\boldsymbol{x} - E(\boldsymbol{x}))Y] = E[\boldsymbol{x}(Y - E(Y))]$. Let

$$\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\eta}}_n = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \boldsymbol{S}_{\boldsymbol{x}Y} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})$$

and

$$\tilde{\boldsymbol{\eta}} = \tilde{\boldsymbol{\eta}}_n = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y}).$$

Then the OLS estimators for model (1) are $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$, $\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}_{OLS}^T \overline{\boldsymbol{x}}$, and

$$\hat{\boldsymbol{\beta}}_{OLS} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \hat{\boldsymbol{\eta}}.$$

For a multiple linear regression model with iid cases, $\hat{\boldsymbol{\beta}}_{OLS}$ is a consistent estimator of $\boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ under mild regularity conditions, while $\hat{\alpha}_{OLS}$ is a consistent estimator of $E(Y) - \boldsymbol{\beta}_{OLS}^T E(\boldsymbol{x})$.

Cook, Helland, and Su (2013) showed that the one component partial least squares (OPLS) estimator $\hat{\boldsymbol{\beta}}_{OPLS} = \hat{\lambda} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ estimates $\lambda \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\beta}_{OPLS}$ where

$$\lambda = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}} \quad \text{and} \quad \hat{\lambda} = \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}}{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}} \tag{3}$$

for $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq \boldsymbol{0}$. If $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$, then $\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$. Also see Basa et al. (2024), Cook and Forzani (2024), and Wold (1975). Olive and Zhang (2025) derived the large sample theory for $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ and OPLS under milder regularity conditions than those in the previous literature, where $\boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Olive et al. (2025) showed that for iid cases $(\boldsymbol{x}_i, Y_i)$, these results still hold for multiple linear regression models with heterogeneity.

The marginal maximum likelihood estimator (MMLE or marginal least squares estimator) is due to Fan and Lv (2008) and Fan and Song (2010). This estimator computes the marginal regression of Y on x_i, such as Poisson regression, resulting in the estimator $(\hat{\alpha}_{i,M}, \hat{\beta}_{i,M})$ for $i = 1, ..., p$. Then $\hat{\boldsymbol{\beta}}_{MMLE} = (\hat{\beta}_{1,M}, ..., \hat{\beta}_{p,M})^T$.

For multiple linear regression, the marginal estimators are the simple linear regression (SLR) estimators, and $(\hat{\alpha}_{i,M}, \hat{\beta}_{i,M}) = (\hat{\alpha}_{i,SLR}, \hat{\beta}_{i,SLR})$. Hence

$$\hat{\boldsymbol{\beta}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}. \tag{4}$$

If the $\boldsymbol{t}_i$ are the predictors that are scaled or standardized to have unit sample variances, then

$$\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \boldsymbol{I}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{t}, Y) \tag{5}$$

where $(\boldsymbol{t}, Y)$ denotes that Y was regressed on $\boldsymbol{t}$, and $\boldsymbol{I}$ is the $p \times p$ identity matrix. Olive et al. (2025) derived large sample theory for the MMLE for the multiple linear regression models, including models with heterogeneity.

For Poisson regression and related models, the respnse variable Y is a nonnegative count variable. A useful *Poisson regression (PR) model* is $Y \sim \text{Poisson}\left(e^{SP}\right)$. This model has $E(Y|SP) = V(Y|SP) = \exp(SP)$. The *quasi-Poisson regression model* has $E(Y|SP) = \exp(SP)$ and $V(Y|SP) = \phi \exp(SP)$ where the dispersion parameter $\phi > 0$. Note that this model and the Poisson regression model have the same conditional mean function, and the conditional variance functions are the same if $\phi = 1$.

Some notation is needed for the negative binomial regression model. If Y has a (generalized) negative binomial distribution, $Y \sim NB(\mu, \kappa)$, then the probability mass function (pmf) of Y is

$$P(Y = y) = \frac{\Gamma(y+\kappa)}{\Gamma(\kappa)\Gamma(y+1)} \left(\frac{\kappa}{\mu+\kappa}\right)^{\kappa} \left(1 - \frac{\kappa}{\mu+\kappa}\right)^{y}$$

for $y = 0, 1, 2, \ldots$ where $\mu > 0$ and $\kappa > 0$. Then $E(Y) = \mu$ and $\text{V}(Y) = \mu + \mu^2/\kappa$.

The *negative binomial regression model* states that $Y_1, \ldots, Y_n$ are independent random variables with

$$Y|SP \sim \text{NB}(\exp(\text{SP}), \kappa).$$

This model has $E(Y|SP) = \exp(SP)$ and

$$V(Y|SP) = \exp(SP)\left(1 + \frac{\exp(SP)}{\kappa}\right) = \exp(SP) + \tau \exp(2\ SP).$$

Following Agresti (2002, p. 560), as $\tau \equiv 1/\kappa \to 0$, it can be shown that the negative binomial regression model converges to the Poisson regression model.

Let the log transformation $Z_i = \log(Y_i)$ if $Y_i > 0$ and $Z_i = \log(0.5)$ if $Y_i = 0$. This transformation often results in a linear model with heterogeneity:

$$Z_i = \alpha_Z + \boldsymbol{x}_i^T \boldsymbol{\beta}_Z + e_i \tag{6}$$

where the e_i are independent with expected value $E(Z_i) = 0$ and variance $V(Z_i) = \sigma_i^2$. For Poisson regression, the minimum chi-square estimator is the weighted least squares estimator from the regression of Z_i on $\boldsymbol{x}_i$ with weights $w_i = e^{Z_i}$. See Agresti (2002, pp. 611–612) and Olive (2013).

If the regression model for Y depends on $\boldsymbol{x}$ only through $\alpha + \boldsymbol{x}^T\boldsymbol{\beta}$, and if the predictors $\boldsymbol{x}_i$ are independent and identically distributed (iid) from a large class of elliptically contoured distributions, then Li and Duan (1989) and Chen and Li (1998) showed that, under regularity conditions, $\boldsymbol{\beta}_{OLS} = c\boldsymbol{\beta}$. Hence $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = c\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}$. Thus $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = d\boldsymbol{\beta}$ if $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \tau^2 \boldsymbol{I}_p$ for some constant $\tau^2 > 0$. If $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$ in this case, then $\beta_i = 0$ implies that $Cov(x_i, Y) = 0$. The constant c is typically nonzero unless m has a lot of symmetry about the distribution of $\alpha + \boldsymbol{x}^T\boldsymbol{\beta}$. Simulation with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ can be difficult if the population values of c and d are unknown. Results from Cameron and Trivedi (1998, p. 89) suggest that if a Poisson regression model is fit using OLS software for multiple linear regression, then a rough approximation is $\hat{\boldsymbol{\beta}}_{PR} \approx \hat{\boldsymbol{\beta}}_{OLS}/\overline{Y}$.

Zhao et al. (2024) have an interesting result for the multiple linear regression model (1). Assume that the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid with $E(Y) = \mu_Y$, $E(\boldsymbol{x}) = \boldsymbol{\mu_x}$ and nonsingular $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma_x}$. Let $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$. Then testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is equivalent to testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$ with $\boldsymbol{\mu} = E(\boldsymbol{w}_i) = \boldsymbol{\Sigma_x}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ where $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu_x})(Y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu_x})^T \boldsymbol{\beta}_0)$, and a one sample test can be applied to $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y} - (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{\beta}_0)$.

This paper modifies the above test for $\boldsymbol{\beta}_0 = \boldsymbol{0}$. The resulting test can be used for many regression models, not just multiple linear regression. Suppose $\boldsymbol{\beta}_D = \boldsymbol{D}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ where $\boldsymbol{D}$ is a $p \times p$ positive definite matrix. Then $\boldsymbol{\beta}_D = \boldsymbol{0}$ if and only if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$. Then $\boldsymbol{D}^{-1} = \lambda \boldsymbol{I}$ for OPLS, $\boldsymbol{D}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}$ for OLS, and $\boldsymbol{D}^{-1} = [diag(\boldsymbol{\Sigma_x})]^{-1}$ for the MMLE. The k-component partial least squares estimator can be found by regressing Y on a constant and on $W_i = \hat{\boldsymbol{\eta}}_i^T \boldsymbol{x}$ for $i = 1, ..., k$ where $\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{i-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ for $i = 1, ..., k$. See Helland (1990). Hence $\boldsymbol{\beta}_{kPLS} = \boldsymbol{0}$ if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$. Thus if the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid, then using $\boldsymbol{\beta}_0 = \boldsymbol{0}$ gives tests for $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$, $H_0 : \boldsymbol{\beta}_{MMLE} = \boldsymbol{0}$, $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$, $H_0 : \boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$, and $H_0 : \boldsymbol{\beta}_{kPLS} = \boldsymbol{0}$. For multiple linear regression with heterogeneity, $\hat{\boldsymbol{\beta}}_{OLS}$ is still a consistent estimator of $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Hence the test can be used when the constant variance assumption is violated.

For a generalized linear model and several other regression models that depend on the predictors $\boldsymbol{x}$ only through $SP = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}$, if $\boldsymbol{\beta} = \boldsymbol{0}$, then the Y_i are iid and do not depend on $\boldsymbol{x}$, and thus satisfy a multiple linear regression model with $\boldsymbol{\beta}_{OLS} = \boldsymbol{0}$. Typically, if $\boldsymbol{\beta} \neq \boldsymbol{0}$, then $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq 0$. An exception is when there is a lot of symmetry which rarely occurs with real data. For example, suppose $Y = m(SP) + e$ where the iid errors $e_i \sim N(0, \sigma_1^2)$ are independent of the predictors, $SP \sim N(0, \sigma_2^2)$, and the function m is symmetric about 0, e.g. $m(SP) = (SP)^2$. Then $\boldsymbol{\beta}_{OLS} = 0$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = 0$ even if $\boldsymbol{\beta} \neq \boldsymbol{0}$.

If $\boldsymbol{\beta}_0 = \boldsymbol{0}$, then $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu_x})(Y_i - \mu_Y)$, and $E(\boldsymbol{w}_i) = E(\boldsymbol{u}_i) = E[\boldsymbol{x}_i(Y_i - \mu_Y)] = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Hence we replace $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})$ by $\boldsymbol{z}_i = \boldsymbol{x}_i(Y_i - \overline{Y})$ and apply a high dimensional one sample test on the $\boldsymbol{z}_i$. Then $\boldsymbol{\mu_x}$ does not need to be estimated by $\overline{\boldsymbol{x}}$.

Section 2 reviews and derives some results for the one sample test that will be used. Section 3 reviews some two sample tests. Section 4 gives theory for the test given in the above paragraph.

2 A High Dimensional One Sample Test

This section reviews and derives some results for the one sample test that will be used. Suppose $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid random vectors with $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and covariance matrix $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$. Then the test $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$ is equivalent to the test $H_0 : \boldsymbol{\mu}^T\boldsymbol{\mu} = 0$ versus $H_1 : \boldsymbol{\mu}^T\boldsymbol{\mu} \neq 0$. Let $\boldsymbol{S} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$. A U-statistic for estimating $\boldsymbol{\mu}^T\boldsymbol{\mu}$ is

$$T_n = T_n(\boldsymbol{x}) = \frac{1}{n(n-1)} \sum_{i \neq j} \boldsymbol{x}_i^T \boldsymbol{x}_j = \frac{n\overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}} - tr(\boldsymbol{S})}{n} \tag{7}$$

where tr() is the trace function

To see that the last equality holds, note that

$$T_n = \frac{1}{n(n-1)}\left[\sum_i \sum_j \boldsymbol{x}_i^T \boldsymbol{x}_j - \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i\right] = \frac{n^2 \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i}{n(n-1)}.$$

Now

$$\boldsymbol{S} = \frac{1}{n-1}\sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T = \frac{1}{n-1}\left[\sum_i \boldsymbol{x}_i \boldsymbol{x}_i^T - n \overline{\boldsymbol{x}}\, \overline{\boldsymbol{x}}^T\right].$$

Thus

$$tr(\boldsymbol{S}) = \frac{1}{n-1}\left[\sum_i tr(\boldsymbol{x}_i \boldsymbol{x}_i^T) - n tr(\overline{\boldsymbol{x}}\, \overline{\boldsymbol{x}}^T)\right] = \frac{1}{n-1}\left[\sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i - n \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}}\right].$$

Thus

$$n \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - tr(\boldsymbol{S}) = n \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} + \frac{n}{n-1} \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - \frac{1}{n-1}\sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i = \frac{n^2 \overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - \sum_i \boldsymbol{x}_i^T \boldsymbol{x}_i}{n-1}.$$

Next we derive a simple test. Let the variance $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Let $m = \text{floor(n/2)} = \lfloor \text{n}/2 \rfloor$ be the integer part of $n/2$. So floor(100/2) = floor(101/2) = 50. Let the iid random variables $W_i = \boldsymbol{x}_{2i-1}^T \boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Hence $W_1, W_2, ..., W_m = \boldsymbol{x}_1^T \boldsymbol{x}_2, \boldsymbol{x}_3^T \boldsymbol{x}_4, ..., \boldsymbol{x}_{2m-1}^T \boldsymbol{x}_{2m}$. Note that $E(W_i) = \boldsymbol{\mu}^T \boldsymbol{\mu}$ and $V(W_i) = \sigma_W^2$. Let S_W^2 be the sample variance of the W_i:

$$S_W^2 = \frac{1}{m-1}\sum_{i=1}^m (W_i - \overline{W})^2.$$

If $\sigma_W^2 \propto \tau^2 p$ where $p > n$, then n may not be large enough for the normal approximation to hold. The following theorem follows from the univariate central limit theorem.

Theorem 1. Assume $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid, $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$, and the variance $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Let $W_1, ..., W_m$ be defined as above. Then
a) $\sqrt{m}(\overline{W} - \boldsymbol{\mu}^T \boldsymbol{\mu}) \xrightarrow{D} N(0, \sigma_W^2)$.

$$b) \ \frac{\sqrt{m}(\overline{W} - \boldsymbol{\mu}^T \boldsymbol{\mu})}{S_W} \xrightarrow{D} N(0,1)$$

as $n \to \infty$.

The following theorem derives the variance $V(T_n)$ under much simpler regularity conditions than those in the literature, and the proof of the theorem is also simpler.

Theorem 2. Assume $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid, $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$, and the variance $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Let $W_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$ for $i \neq j$. Let $\theta = Cov(W_{ij}, W_{id}) = \boldsymbol{\mu}^T \boldsymbol{\Sigma} \boldsymbol{\mu}$ where $j \neq d$, $i < j$, and $i < d$. Then

$$a) \ V(T_n) = \frac{2\sigma_W^2}{n(n-1)} + \frac{4(n-2)\theta}{n(n-1)}.$$

b) If $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ is true, then $\theta = 0$ and

$$V_0 = V(T_n) = \frac{2\sigma_W^2}{n(n-1)}.$$

Proof. a) To find the variance $V(T_n)$ with T_n from Equation (7), let $W_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j = W_{ji}$, and note that

$$T_n = \frac{2}{n(n-1)} H_n \quad \text{where} \quad \mathrm{H_n} = \sum_{\mathrm{i} <} \sum_{\mathrm{j}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{x}_{\mathrm{j}} = \sum_{\mathrm{i<j}} \boldsymbol{x}_{\mathrm{i}}^{\mathrm{T}} \boldsymbol{x}_{\mathrm{j}}.$$

Then $V(H_n) = Cov(H_n, H_n) =$

$$Cov\left(\sum_{i <} \sum_{j} W_{ij}, \sum_{k <} \sum_{d} W_{kd} \right) = \sum_{i <} \sum_{j} \sum_{k <} \sum_{d} Cov(W_{ij}, W_{kd}). \tag{8}$$

Let $V(W_{ij}) = \sigma_W^2$ for $i \neq j$. The covariances are of 3 types. First, if $(ij) = (kd)$ with $i < j$, then $Cov(W_{ij}, W_{kd}) = V(W_{ij}) = \sigma_W^2$. Second, if i, j, k, d are distinct with $i < j$ and $k < d$, then W_{ij} and W_{kd} are independent with $Cov(W_{ij}, W_{kd}) = 0$. Third, there are terms where exactly three of the four subscripts are distinct, which have $Cov(W_{ij}, W_{id}) = \theta$ where $j \neq d$, $i < j$, and $i < d$ or $Cov(W_{ij}, W_{kj}) = \theta$ where $i \neq k$, $i < j$, and $k < j$. These covariance terms are all equal to the same number θ since $W_{ij} = W_{ji}$. The number of ways to get three distinct subscripts is

$$a - b - c = \binom{n}{2}^2 - \binom{n}{2}\binom{n-2}{2} - \binom{n}{2} = n(n-1)(n-2)$$

since a is the number of terms on the right hand side of (3), b is the number of terms where i, j, k, d are distinct with $i < j$ and $k < d$, and c is the number of terms where $(ij) = (kd)$ with $i < j$. [Note that $n(n-1)$ terms have i and j distinct. Half of these terms have $i < j$ and half have $i > j$. Similarly, $n(n-1)(n-2)(n-3)$ terms have $ijkd$ distinct, and half of the $n(n-1)$ terms have $i < j$, while half of the $(n-2)(n-3)$ terms have $k < d$.] Thus

$$V(H_n) = 0.5 n(n-1)\sigma_W^2 + n(n-1)(n-2)\theta.$$

This calculation was adapted from Lehmann (1975, pp. 336-337). Thus

$$V(T_n) = \frac{4}{[n(n-1)]^2} V(H_n) = \frac{2\sigma_W^2}{n(n-1)} + \frac{4(n-2)\theta}{n(n-1)}.$$

b) Now $\theta = Cov(\boldsymbol{x}_i^T \boldsymbol{x}_j, \boldsymbol{x}_i^T \boldsymbol{x}_k)$ where $\boldsymbol{x}_i, \boldsymbol{x}_j$, and $\boldsymbol{x}_k$ are iid. Hence $\theta =$

$$Cov(\sum_d x_{id} x_{jd}, \sum_t x_{it} x_{kt}) = \sum_d \sum_t Cov(x_{id} x_{jd}, x_{it} x_{kt}) =$$

$$\sum_d \sum_t [E(x_{id} x_{jd} x_{it} x_{kt}) - E(x_{id} x_{jd}) E(x_{it} x_{kt})] =$$

$$\sum_d \sum_t [E(x_{id}x_{it})E(x_{jd})E(x_{kt}) - E(x_{id})E(x_{jd})E(x_{it})E(x_{kt})] =$$

$$\sum_d \sum_t [E(x_{jd})E(x_{kt})(E(x_{id}x_{it}) - E(x_{id})E(x_{it}))] =$$

$$\sum_d \sum_t [E(x_{jd})E(x_{kt})\ Cov(x_{id}, x_{it})] = \boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}.$$

Under H_0, $\boldsymbol{\mu} = \mathbf{0}$ and thus $\theta = 0$. $\square$

Note that T_n is the sample mean of the $0.5n(n-1)$ distinct, identically distributed $W_{ij} = \boldsymbol{x}_i^T\boldsymbol{x}_j$ for $i \neq j$. When $\boldsymbol{\mu} = \mathbf{0}$, Theorem 2 proves that the W_{ij} are uncorrelated. Hence when H_0 is true, $V(T_n)$ satisfies Theorem 2b). Chen and Qin (2010) obtained

$$V(T_n) = \frac{2}{n(n-1)} tr(\boldsymbol{\Sigma}^2) + \frac{4\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}}{n},$$

omitting $(n-2)/(n-1)$ from the second term. Since

$$V_0(T_n) = \frac{2}{n(n-1)} tr(\boldsymbol{\Sigma}^2),$$

$V(\boldsymbol{x}_i^T\boldsymbol{x}_j) = \sigma_W^2 = tr(\boldsymbol{\Sigma}^2)$.

For example, Li (2023) finds $V(T_n)$ when H_0 is true, using much stronger regularity conditions than in Theorem 2. In the simulations, we use a variant of the Li (2023) variance estimator $\hat{\sigma}_W^2$, and also use the estimator S_W^2 that is much easier to compute. Srivastava and Du (2008), Bai and Saranadasa (1996), Chen and Qin (2010), and others use $T_n/\sqrt{\hat{V}(T_n)} \xrightarrow{D} N(0,1)$, while Li (2023) uses $T_n/\sqrt{\hat{V}_0(T_n)} \xrightarrow{D} N(0,1)$. Theorem 2 and the following result show that the second statistic has more power. Adapting an argument from Lehmann (1999, pp. 367-368), let $Z(\boldsymbol{a}) = E(\boldsymbol{a}^T\boldsymbol{x}_j) = \boldsymbol{a}^T\boldsymbol{\mu}$. Then it can be shown that $\theta = V(Z(\boldsymbol{x}_i)) = V(\boldsymbol{x}_i^T\boldsymbol{\mu}) \geq 0$. Also, by Theorem 2, $\theta = \boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu} \geq 0$. Let $s_n^2 = \hat{V}$ be a consistent estimator of $V(T_n)$ and let

$$\hat{V}_0 = \frac{2\hat{\sigma}_W^2}{n(n-1)}.$$

The test statistics

$$t_1 = \frac{T_n}{\sqrt{\hat{V}_0}} \xrightarrow{D} N(0,1) \quad \text{and} \quad \mathrm{t}_2 = \frac{\mathrm{T_n}}{\sqrt{\hat{\mathrm{V}}}} \xrightarrow{\mathrm{D}} \mathrm{N}(0,1)$$

if $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is true. However, when H_0 is not true,

$$\hat{V} \approx \hat{V}_0 + \frac{4(n-2)\hat{\theta}}{n(n-1)}$$

where the second term is positive. If H_0 is not true and n and p are such that the second term dominates, then $|t_1|$ tends to be proportional to $\sqrt{n}|t_2|$, greatly increasing the power of the test that uses t_1.

For power, we expect $V_0(T_n) \to 0$ if $p/n^2 \to 0$ as $n \to \infty$. The high dimensional literature often gives very strong regularity conditions where $V(T_n) \to 0$ if $p = p_n = n^\gamma$ where γ is often much larger than 0.5 and $\boldsymbol{\mu} = \mathbf{0}$. Suppose $\boldsymbol{\mu} = \delta\mathbf{1}$ where the constant $\delta > 0$ and $\mathbf{1}$ is the $p \times 1$ vector of ones. Then $\boldsymbol{\mu}^T\boldsymbol{\mu} = \delta^2 p$, and the test using $\hat{V}_0(T_n)$ may have good power for $T_n/\sqrt{\hat{V}_0(T_n)} > 1.96 \approx 2$ or for

$$\frac{\delta^2 \ p}{\sqrt{\frac{2\sigma_W^2}{n(n-1)}}} > 2 \quad \text{or} \quad \delta^2 > \frac{2\sqrt{2}\ \sigma_{\text{W}}}{\text{n p}}.$$

The above theory can also be applied to the $\boldsymbol{z}_i = ss(\boldsymbol{x}_i)$ to test $H_0 : E(\boldsymbol{z}) = \mathbf{0}$. As noted near the end of Section 1, for elliptically contoured distributions, $E(\boldsymbol{z}) = \boldsymbol{\mu}_{\boldsymbol{z}} = \mathbf{0}$ if $E(\boldsymbol{x}) = \boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{x}} = \mathbf{0}$.

Let $V_0(T_n)$ be the variance of T_n when $H_0 : \boldsymbol{\mu} = \mathbf{0}$ is true. Let the variance $V(\boldsymbol{x}_i^T\boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. Abid and Olive (2025) give a straight forward proof that

$$V_0(T_n) = \frac{2\sigma_W^2}{n(n-1)}.$$

Chen and Qin (2010) proved that

$$V_0(T_n) = \frac{2}{n(n-1)} tr(\boldsymbol{\Sigma}^2)$$

where tr() is the trace function. Thus $V(\boldsymbol{x}_i^T\boldsymbol{x}_j) = \sigma_W^2 = tr(\boldsymbol{\Sigma}^2)$. Srivastava and Du (2008), Bai and Saranadasa (1996), Chen and Qin (2010), Li (2023) and others proved that under mild regularity conditions when H_0 is true, $T_n/\sqrt{\hat{V}_0(T_n)} \xrightarrow{D} N(0,1)$. Under regularity conditions when H_0 is true, Li (2023) proved that $T_n/\sqrt{\hat{V}_0(T_n)} \xrightarrow{D} t_k$ as $p \to \infty$ for fixed $n \geq 3$ where $k = 0.5n(n-1) - 1$.

Two estimators of σ_W^2 are simple to compute. Let $W_{ij} = \boldsymbol{x}_i^T\boldsymbol{x}_j$ for $i \neq j$. Let $s_n^2 = \hat{V}_0(T_n)$. An estimator nearly the same as the one used by Li (2023) is

$$n(n-1)s_n^2 = \hat{\sigma}_W^2 = \frac{1}{n(n-1)} \sum\sum_{i \neq j} (\boldsymbol{x}_i^T\boldsymbol{x}_j - T_n)^2 = \frac{1}{n(n-1)} \sum\sum_{i \neq j} (W_{ij} - T_n)^2.$$

Ahlam and Olive (2025) proposed the following estimator. Let $m = \text{floor}(n/2) = \lfloor n/2 \rfloor$ be the integer part of $n/2$. So floor(100/2) = floor(101/2) = 50. Let the iid random variables $W_i = \boldsymbol{x}_{2i-1}^T\boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Hence $W_1, W_2, ..., W_m = \boldsymbol{x}_1^T\boldsymbol{x}_2, \boldsymbol{x}_3^T\boldsymbol{x}_4, ..., \boldsymbol{x}_{2m-1}^T\boldsymbol{x}_{2m}$. Note that $E(W_i) = \boldsymbol{\mu}^T\boldsymbol{\mu}$ and $V(W_i) = \sigma_W^2$. Let $n(n-1)s_n^2 = S_W^2$ be the sample variance of the W_i.

Consider testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_A : \boldsymbol{\mu} \neq \mathbf{0}$ using independent and identically distributed (iid) $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ where the $\boldsymbol{x}_i$ are $p \times 1$ random vectors and p may be much larger than n. Assume the expected value $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and nonsingular covariance matrix $Cov(\boldsymbol{x}_i) = \boldsymbol{\Sigma}$. Replace $\boldsymbol{x}_i$ by $\boldsymbol{w}_i = \boldsymbol{x}_i - \boldsymbol{\mu}_0$ to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. This section reviews some tests while the following section gives simpler large sample theory for some of the tests, including a new test that has very simple large sample theory.

Suppose p is fixed, and 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.

Rajapaksha and Olive (2024) showed how to bootstrap Wald tests with the wrong dispersion matrix. When $\boldsymbol{C}_n = \boldsymbol{I}_g$, the bootstrap tests often became conservative as g increased to n. For some of these tests, the m out of n bootstrap, which draws a sample of size m without replacement from the n, works better than the nonparametric bootstrap.

When n is much larger than p, the one sample Hotelling (1931) T^2 test is often used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. The sample mean

$$\overline{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^n \boldsymbol{x}_i,$$

and the sample covariance matrix

$$\boldsymbol{S} = \frac{1}{n-1} \sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T = (S_{ij}).$$

That is, the ij entry of $\boldsymbol{S}$ is the sample covariance S_{ij}. If the $\boldsymbol{x}_i$ are iid with expected value $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and nonsingular covariance matrix $Cov(\boldsymbol{x}_i) = \boldsymbol{\Sigma}$, then by the multivariate central limit theorem

$$\sqrt{n}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}).$$

If H_0 is true, then

$$T_H^2 = n(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0)^T \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}_0) \xrightarrow{D} \chi^2_p.$$

The one sample Hotelling's T^2 test rejects H_0 if $T_H^2 > D^2_{1-\delta}$ where $D^2_{1-\delta} = \chi^2_{p,\delta}$ and $P(Y \leq \chi^2_{p,\delta}) = \delta$ if $Y \sim \chi^2_p$. Alternatively, use

$$D^2_{1-\delta} = \frac{(n-1)p}{n-p} F_{p,n-p,1-\delta}$$

where $P(Y \leq F_{p,d,\delta}) = \delta$ if $Y \sim F_{p,d}$. The scaled F cutoff can be used since $T_H^2 \xrightarrow{D} \chi^2_p$ if H_0 holds, and

$$\frac{(n-1)p}{n-p} F_{p,n-p,1-\delta} \to \chi^2_{p,1-\delta}$$

as $n \to \infty$.

The next two high dimensional tests are described in Srivastava and Du (2008). Also see Hu and Bai (2015). Let $tr(\boldsymbol{A})$ be the trace of square matrix $\boldsymbol{A}$. Let $\boldsymbol{R}$ be the sample correlation matrix. Consider testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{0}$. Let $\boldsymbol{D} = diag(\boldsymbol{S})$. Let

$$c_{p,n} = 1 + \frac{tr(\boldsymbol{R}^2)}{p^{3/2}}.$$

Let $n = O(p^\delta)$ where $0.5 < \delta \leq n$. Then under regularity conditions

$$Z_1 = \frac{n\overline{\boldsymbol{x}}^T \boldsymbol{D}^{-1} \overline{\boldsymbol{x}} - \frac{(n-1)p}{n-3}}{\sqrt{2\left(tr(\boldsymbol{R}^2) - \frac{p^2}{n-1}\right) c_{p,n}}} \xrightarrow{D} N(0,1)$$

as $n, p \to \infty$. The next test is attributed to Bai and Saranadasa (1996). Under regularity conditions,

$$Z_2 = \frac{n\overline{\boldsymbol{x}}^T \overline{\boldsymbol{x}} - tr(\boldsymbol{S})}{\left[\frac{2(n-1)n}{(n-2)(n+1)}\left(tr(\boldsymbol{S}^2) - \frac{1}{n}[tr(\boldsymbol{S})]^2\right)\right]^{1/2}} \xrightarrow{D} N(0,1)$$

as $n, p \to \infty$. Both of these test statistics used $p/n \to c > 0$ or $p/n^2 \to 0$.

Note that $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ holds if and only if $\|\boldsymbol{\mu}\|^2 = \boldsymbol{\mu}^T\boldsymbol{\mu} = 0$. The T_n in Equation (1) below can be viewed as a modification of $\|\overline{\boldsymbol{x}}\|^2 = \overline{\boldsymbol{x}}^T\overline{\boldsymbol{x}}$ that is a better estimator of $\boldsymbol{\mu}^T\boldsymbol{\mu}$ in high dimensions. Note that $E(\boldsymbol{x}_i^T\boldsymbol{x}_j) = \boldsymbol{\mu}^T\boldsymbol{\mu}$ if $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are iid with $E(\boldsymbol{x}_i) = \boldsymbol{\mu}$ and $i \neq j$. Let $V(T_n)$ be the variance of T_n and let $s_n^2 = \hat{V}(T_n)$ be a consistent estimator of $V(T_n)$.

The following test is due to Chen and Qin (2010). Also see Hu and Bai (2015). Let $\boldsymbol{a} = \sum_{i=1}^n \boldsymbol{x}_i$ and let $\boldsymbol{X} = (x_{ij})$ be the data matrix with ith row $= \boldsymbol{x}_i^T$ and ij element $= x_{ij}$. Let $vec(\boldsymbol{A})$ stack the columns of matrix $\boldsymbol{A}$ so that $\boldsymbol{c} = vec(\boldsymbol{X}^T) = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T, ..., \boldsymbol{x}_n^T]^T$. Then

$$\boldsymbol{c}^T\boldsymbol{c} = \sum_{i=1}^n \boldsymbol{x}_i^T\boldsymbol{x}_i = \sum_{i=1}^n \|\boldsymbol{x}_i\|^2 = \sum_{i=1}^n \sum_{j=1}^p (x_{ij})^2.$$

Let

$$T_n = \frac{1}{n(n-1)}[\boldsymbol{a}^T\boldsymbol{a} - \boldsymbol{c}^T\boldsymbol{c}] = \frac{1}{n(n-1)} \sum\sum_{i \neq j} \boldsymbol{x}_i^T\boldsymbol{x}_j = \frac{1}{n(n-1)} \sum_{i \neq j} \boldsymbol{x}_i^T\boldsymbol{x}_j. \tag{9}$$

The terms in $\boldsymbol{c}^T\boldsymbol{c} = \sum_{i=1}^n \boldsymbol{x}_i^T\boldsymbol{x}_i$ are the terms that cause the restriction on p for asymptotic normality for the previous two tests. Under $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ and additional regularity conditions,

$$\frac{T_n}{\sqrt{V(T_n)}} \xrightarrow{D} N(0,1) \quad \text{and} \quad \frac{\mathrm{T_n}}{\mathrm{s_n}} \xrightarrow{\mathrm{D}} \mathrm{N}(0,1) \tag{10}$$

where s_n is rather hard to compute. Here

$$s_n^2 = \frac{2}{n(n-1)} tr\left[\sum_{i \neq j} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_{(i,j)})\boldsymbol{x}_i^T(\boldsymbol{x}_j - \overline{\boldsymbol{x}}_{(i,j)})\boldsymbol{x}_j^T\right]$$

is a consistent estimator of $V(T_n)$ where $\overline{\boldsymbol{x}}_{(i,j)}$ is the sample mean computed without $\boldsymbol{x}_i$ or $\boldsymbol{x}_j$:

$$\overline{\boldsymbol{x}}_{(i,j)} = \frac{1}{n-2} \sum_{k \neq i,j} \boldsymbol{x}_k.$$

We will also consider replacing $\boldsymbol{x}_i$ by $\boldsymbol{z}_i = ss(\boldsymbol{x}_i)$ where the spatial sign function $ss(\boldsymbol{x}_i) = \boldsymbol{0}$ if $\boldsymbol{x}_i = \boldsymbol{0}$, and $ss(\boldsymbol{x}_i) = \boldsymbol{x}_i/\|\boldsymbol{x}_i\|$ otherwise. This function projects the nonzero $\boldsymbol{x}_i$ onto the unit p-dimensional hypersphere centered at $\boldsymbol{0}$. Let $T_n(\boldsymbol{w})$ denote the statistic T_n computed from an iid sample $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$. Since the $\boldsymbol{z}_i$ are iid if the $\boldsymbol{x}_i$ are iid, use $T_n(\boldsymbol{z})$ to test $H_0 : \boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{0}$ versus $H_A : \boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{0}$ where $\boldsymbol{\mu}_{\boldsymbol{z}} = E(\boldsymbol{z}_i)$. In general, $\boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{\mu} = \boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x}_i)$, but $\boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{\mu} = \boldsymbol{0}$ can occur if the $\boldsymbol{x}_i$ have a lot of symmetry about $\boldsymbol{0}$. In particular, $\boldsymbol{\mu}_{\boldsymbol{z}} = \boldsymbol{\mu} = \boldsymbol{0}$ if the $\boldsymbol{x}_i$ are iid from an elliptically contoured distribution with center $\boldsymbol{\mu} = \boldsymbol{0}$. The test based on the statistic $T_n(\boldsymbol{z})$ can be useful if the second moment of the $\boldsymbol{x}_i$ does not exist, for example if the $\boldsymbol{x}_i$ are iid from a multivariate Cauchy distribution. These results may be useful for understanding papers such as Wang, Peng, and Li (2015)

Section 2 considers two estimators s_n^2 of $V(T_n)$ that are easier to compute when H_0 is true, and gives a new test with very simple large sample theory. Section 3 considers two sample tests.

3 Estimating $V(T_n)$

The nonparametric bootstrap draws a bootstrap data set $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_n^*$ with replacement from the $\boldsymbol{x}_i$ and computes T_1^* by applying T_n on the bootstrap data set. This process is repeated B times to get a bootstrap sample $T_1^*, ..., T_B^*$. For the statistic T_n, the nonparametric bootstrap fails in high dimensions because terms like $\boldsymbol{x}_j^T \boldsymbol{x}_j$ need to be avoided, and the nonparametric bootstrap has replicates: the proportion of cases in the bootstrap sample that are not replicates is about $1 - e^1 \approx 2/3 \approx 7/11$. The m out of n bootstrap draws a sample of size m without replacement from the n cases. For $B = 1$, this is a data splitting estimator, and $T_m^* \approx N(0, s_m^2)$ for large enough m and p. Sampling without replacement is also known as subsampling and the delete d jackknife.

Theory for subsampling is given by Politis and Romano (1994) and Wu (1990). Subsampling tends to work well for a large variety of statistics if $m/n \to 0$ with $m \to \infty$. A linear statistic has the form

$$\frac{1}{n} \sum_{i=1}^{n} t(U_i)$$

where $\theta = E[t(U_i)]$ and the U_i are iid. For a linear statistic, subsampling tends to work well if $m/n \to \tau \in [0, 1)$ with $m \to \infty$. For the $W_i = U_i$ in Theorem 1, $t(U_i) = U_i = \boldsymbol{x}_{2i-1}^T \boldsymbol{x}_{2i}$. If different blocks were taken such that the W_i are still iid, then subsampling would still work, but the statistics from the different blocks are estimating the same quantiles. Hence subsampling from all of the data may also work well. That is, subsampling may work well for a U-statistic that is the analog of a linear statistic. Using $m = floor(2n/3)$ worked well in simulations.

Now let W_i be an indicator random variable with $W_i = 1$ if $\boldsymbol{x}_i^*$ is in the sample and $W_i = 0$, otherwise, for $i = 1, ..., n$. The W_i are binary and identically distributed, but not independent. Hence $P(W_i = 1) = m/n$. Let $W_{ij} = W_i W_j$ with $i \neq j$. Again, the W_{ij} are binary and identically distributed. $P(W_{ij} = 1) =$ P(ordered pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$) was selected in the sample. Hence $P(W_{ij} = 1) = m(m-1)/[n(n-1)]$ since $m(m-1)$ ordered pairs were selected out of $n(n-1)$ possible ordered pairs. Then

$$T_m^* = \frac{1}{m(m-1)} \sum \sum_{k \neq d} \boldsymbol{x}_{i_k}^T \boldsymbol{x}_{i_d} = \frac{1}{m(m-1)} \sum \sum_{i \neq j} W_i W_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

where the $\boldsymbol{x}_{i_1}, ..., \boldsymbol{x}_{i_m}$ are the m vectors $\boldsymbol{x}_i$ selected in the sample. The first double sum has $m(m-1)$ terms while the second double sum has $n(n-1)$ terms. Hence

$$E(T_m^*) = \frac{1}{m(m-1)} \sum \sum_{i \neq j} E[W_i W_j] \boldsymbol{x}_i^T \boldsymbol{x}_j = T_n.$$

See similar calculations in Buja and Stuetzle (2006). Note that $V(T_m^*) = E([T_m^*]^2) - [T_n]^2 = Cov(T_m^*, T_m^*)$.

4 High Dimensional Two Sample Tests

If $(\boldsymbol{x}_{1i}, \boldsymbol{x}_{2i})$ come in correlated pairs, a high dimensional analog of the paired t test applies the one sample test on $\boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$.

Now suppose there are two independent random samples $\boldsymbol{x}_{1,1}, ..., \boldsymbol{x}_{1,n_1}$ and $\boldsymbol{x}_{2,1}, ..., \boldsymbol{x}_{2,n_2}$ from two populations or groups, and that it is desired to test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ versus $H_1 : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$ where $E(\boldsymbol{x}_i) = \boldsymbol{\mu}_i$ are $p \times 1$ vectors. Let $n = n_1 + n_2$. Let $\boldsymbol{S}_i$ be the sample covariance matrix of $\boldsymbol{x}_i$ and let $Cov(\boldsymbol{x}_i) = \boldsymbol{\Sigma}_i$ for $i = 1, 2$.

The classical two sample Hotelling's T^2 test uses

$$T_C^2 = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \hat{\boldsymbol{\Sigma}}_{pool} \right]^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)$$

where

$$\hat{\boldsymbol{\Sigma}}_{pool} = \frac{(n_1 - 1)\boldsymbol{S}_1 + (n_2 - 1)\boldsymbol{S}_2}{n - 2}.$$

Then reject H_0 if $T_C^2 > m F_{m,n-2,1-\alpha}$.

The large sample test uses

$$T_L^2 = (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^T \left(\frac{\boldsymbol{S}_1}{n_1} + \frac{\boldsymbol{S}_2}{n_2} \right)^{-1} (\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2).$$

Let $d_n = \min(n_1 - p, n_2 - p)$. Then reject H_0 if $T_L^2 > m F_{m,d_n,1-\alpha}$.

Note that $T_C^2 \approx T_L^2$ if $n_1 \approx n_2 \geq 20p$ and the two tests are asymptotically equivalent if $n_i/n \to 0.5$ as $n_1, n_2 \to \infty$. If the n_i/n are not close to 0.5, then the test based on T_C^2 is useful if $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, a very strong assumption. Rajapaksha and Olive (2024) show how to get a bootstrap test based on T_C^2 where the assumption $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ is not needed.

There are test statistics T_n for testing $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ where p can be much larger than n with
$$\frac{T_n}{s_n} \xrightarrow{D} N(0,1)$$
where T_n is relatively simple to compute while s_n is much harder to compute. A simple test takes $m = \min(n_1, n_2)$ and $\boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Then apply the one sample test from Theorem 2 to the $\boldsymbol{z}_i$. This test might work well in high dimensions because of the superior power of the Theorem 2 test, but in low dimensions, it is known that there are better tests.

Let $\boldsymbol{x}_1$ be the $\boldsymbol{x}_i$ that has $n_1 \leq n_2$. Then let

$$\boldsymbol{y}_i = \boldsymbol{x}_{1i} - \sqrt{\frac{n_1}{n_2}}\boldsymbol{x}_{2i} + \frac{1}{\sqrt{n_1 n_2}} \sum_{j=1}^{n_1} \boldsymbol{x}_{2j} - \overline{\boldsymbol{x}}_2 = \boldsymbol{x}_{1i} - \sqrt{\frac{n_1}{n_2}}\boldsymbol{x}_{2i} + \boldsymbol{a}_{n_1,n_2} - \overline{\boldsymbol{x}}_2$$

for $i = 1, ..., n_1$. Note that $\boldsymbol{y}_i = \boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$ if $n_1 = n_2$. Anderson (1984, pp. 177-178) proved that $\overline{\boldsymbol{y}} = \overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2$, that $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ are uncorrelated for $i \neq j$, that $E(\boldsymbol{y}_i) = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$, and that $Cov(\boldsymbol{y}_i) = Cov(\boldsymbol{x}_1) + (n_1/n_2)Cov(\boldsymbol{x}_2)$ for $i = 1, ..., n_1$. Li (2023) showed that $T_n(\boldsymbol{y})/\sqrt{\hat{V}_0(\boldsymbol{y})} \xrightarrow{D} N(0,1)$ where the $\boldsymbol{y}$ denotes that the one sample test was computed using the $\boldsymbol{y}_i$.

Note that $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ holds if and only if $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = \boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T\boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_1^T\boldsymbol{\mu}_2 = 0$. These terms can be estimated by $T_n = T_n(\boldsymbol{x}, \boldsymbol{y}) = T_1 + T_2 - 2T_3$ where T_1 and T_2 are the one sample test statistic applied to samples 1 and 2 and $n_1 n_2 T_3 = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \boldsymbol{x}_{1i}^T \boldsymbol{x}_{2j}$. Let $\boldsymbol{a} = \sum_{i=1}^{n_1} \boldsymbol{x}_{1i}$ and let $\boldsymbol{X}_1 = (x_{1ij})$ be the data matrix with ith row $= \boldsymbol{x}_{1i}^T$ and ij element $= x_{1ij}$. Let $\boldsymbol{c} = vec(\boldsymbol{X}_1^T) = [\boldsymbol{x}_{11}^T, \boldsymbol{x}_{12}^T, ..., \boldsymbol{x}_{1n_1}^T]^T$. Then

$$\boldsymbol{c}^T\boldsymbol{c} = \sum_{i=1}^{n1} \boldsymbol{x}_{1i}^T \boldsymbol{x}_{1i} = \sum_{i=1}^{n_1} \|\boldsymbol{x}_{1i}\|^2 = \sum_{i=1}^{n_1} \sum_{j=1}^{p} (x_{1ij})^2.$$

Let $\boldsymbol{b} = \sum_{i=1}^{n_2} \boldsymbol{x}_{2i}$ and let $\boldsymbol{X}_2 = (x_{2ij})$ be the data matrix with ith row $= \boldsymbol{x}_{2i}^T$ and ij element $= x_{2ij}$. Let $\boldsymbol{d} = vec(\boldsymbol{X}_2^T) = [\boldsymbol{x}_{21}^T, \boldsymbol{x}_{22}^T, ..., \boldsymbol{x}_{2n_2}^T]^T$. Then

$$\boldsymbol{d}^T\boldsymbol{d} = \sum_{i=1}^{n_2} \boldsymbol{x}_{2i}^T \boldsymbol{x}_{2i} = \sum_{i=1}^{n_2} \|\boldsymbol{x}_{2i}\|^2 = \sum_{i=1}^{n_2} \sum_{j=1}^{p} (x_{2ij})^2.$$

Thus

$$T_n = T_1 + T_2 - 2T_3 = \frac{1}{n_1(n_1 - 1)}[\boldsymbol{a}^T\boldsymbol{a} - \boldsymbol{c}^T\boldsymbol{c}] \; + \; \frac{1}{n_2(n_2 - 1)}[\boldsymbol{b}^T\boldsymbol{b} - \boldsymbol{d}^T\boldsymbol{d}] \; - \; \frac{2\boldsymbol{a}^T\boldsymbol{b}}{n_1 n_2}.$$

The terms in $\boldsymbol{c}^T\boldsymbol{c}$ and $\boldsymbol{d}^T\boldsymbol{d}$ are the terms that cause the restriction on p for asymptotic normality. Under $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ and additional regularity conditions,

$$\frac{T_n}{s_n} \xrightarrow{D} N(0,1)$$

where s_n is rather hard to compute. See Hu and Bai (2015) and Chen and Qin (2010).

5 SIMULATIONS

Remark 1. Let $N = n_1 + n_2$ and assume $n_i/N \to \pi_i \in (0,1)$ for $i = 1, 2$. In Theorem 2, $V_0(T_n) \propto 1/n^2$ while $V(T_n) \propto 1/n$ if $\boldsymbol{\mu} \neq \mathbf{0}$, resulting in a large increase in power compared to tests that use $V(T_n)$. The Li (2023) two sample test has $V_0(T_n) \propto 1/N^2$. For the Chen and Qin (2010) two sample test with $T_n = T_n(\boldsymbol{x}, \boldsymbol{y})$, we conjecture that $V_0(T_n) \propto 1/N^2$ and $V(T_n) \propto 1/N$. See Abid (2025) for details. Programs based on the conjecture failed because often $\hat{V}_0(T_n) < 0$, perhaps because the plug in estimators had too much variability. The bootstrap test did work fairly well in the simulations.

5.1 One Sample Tests

In the simulations, we examined four one sample tests. The first "test" used the m out of n bootstrap to compute $T_1^*, ..., T_B^*$ with $B = 100$. We used the shorth bootstrap confidence interval described in Olive (2025, chapter 2) and Pelawa Watagoda and Olive (2021). This "test" has not been proven to have level α. The second test computed the usual t confidence interval

$$[\overline{W} - t_{1-\alpha/2,m-1} S_W/\sqrt{m}, \overline{W} + t_{1-\alpha/2,m-1} S_W/\sqrt{m}]$$

for $\boldsymbol{\mu}^T\boldsymbol{\mu}$ based on the W_i from Theorem 1. The third and fourth tests used Theorem 2 b) and Equation 2): $T_n/s_n \xrightarrow{D} N(0,1)$ if s_n^2 is a consistent estimator of $V(T_n)$ when H_0 is true. The third test used

$$n(n-1)s_n^2 = \hat{\sigma}_W^2 = \frac{1}{n(n-1)} \sum\sum_{i \neq j} (\boldsymbol{x}_i^T \boldsymbol{x}_j - T_n)^2 = \frac{1}{n(n-1)} \sum\sum_{i \neq j} (W_{ij} - T_n)^2.$$

If the denominator $n(n-1)$ was replaced by $n(n-1)-1$, this statistic would be the usual sample variance of the W_{ij}, which are not independent. This test is nearly the same as the Li (2023) test. The fourth test used $n(n-1)s_n^2 = S_W^2$ based on Theorem 1. These two tests computed intervals

$$[T_n - t_{1-\alpha/2,m-1}\sqrt{2s_n^2/[n(n-1)]}, T_n + t_{1-\alpha/2,m-1}\sqrt{2s_n^2/[n(n-1)]}].$$

The third test computed the usual t confidence interval

$$[\overline{W} - t_{1-\alpha/2,m-1} S_W/\sqrt{m}, \overline{W} + t_{1-\alpha/2,m-1} S_W/\sqrt{m}]$$

for $\boldsymbol{\mu}^T\boldsymbol{\mu}$ based on the W_i from Theorem 1. The tests 2–4 use the same cutoff $t_{1-\alpha/2,m-1}$ so that the average interval lengths are more comparable. The fifth test used the Theorem 2 test applied to the spatial sign vectors with S_W^2.

The estimator $\hat{\sigma}_W^2$ is easy to code in *R*. Let $\boldsymbol{X}$ be the $n \times p$ data matrix with ith row $\boldsymbol{x}_i^T$. Then the sum of squares and cross products matrix is $\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T = (c_{ij})$ with ijth element $c_{ij} = \boldsymbol{x}_i^T\boldsymbol{x}_j$. Let $\boldsymbol{A} = \boldsymbol{X}\boldsymbol{X}^T - T_n\mathbf{1}\mathbf{1}^T = (a_{ij})$ where $\mathbf{1}\mathbf{1}^T$ is the $n \times n$ matrix of ones. Let matrix $\boldsymbol{V} = (v_{ij})$ where $v_{ij} = a_{ij}^2 = (\boldsymbol{x}_i^T\boldsymbol{x}_j - T_n)^2$ is the ijth element of $\boldsymbol{V}$. Thus $n(n-1)\hat{\sigma}_W^2 = \sum_{i=1}^n \sum_{j=1}^n v_{ij} - \sum_{i=1}^n v_{ii}$.

```
k <- n*(n-1)
a <- apply(x,2,sum)   #a = n xbar and x is the data matrix
Thd <- (t(a)%*%a - sum(x^2))/k
Thd <- as.double(Thd)   #Thd = Tn
sscp <- x%*%t(x)
ss <- sscp - Thd
ss <- ss^2
vw1 <- (sum(ss) - sum(diag(ss)))/k    #\hat{\sigma}_W^2
```

The simulation used four distribution types where $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y} + \delta\boldsymbol{1}$ with $E(\boldsymbol{x}) = \delta\boldsymbol{1}$ where $\boldsymbol{1}$ is the $p \times 1$ vector of ones. Type 1 used $\boldsymbol{y} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, type 2 used a mixture distribution $\boldsymbol{y} \sim 0.6N_p(\boldsymbol{0}, \boldsymbol{I}) + 0.4N_p(\boldsymbol{0}, 25\boldsymbol{I})$, type 3 for a multivariate t_4 distribution, and type 4 for a multivariate lognormal distribution where $\boldsymbol{y} = (y_1, ..., y_p)$ with $w_i = \exp(Z)$ where $Z \sim N(0,1)$ and $y_i = w_i - E(w_i)$ where $E(w_i) = \exp(0.5)$. The covariance matrix type depended on the matrix $\boldsymbol{A}$. Type 1 used $\boldsymbol{A} = \boldsymbol{I}_p$, type 2 used $\boldsymbol{A} = diag(\sqrt{1}, ..., \sqrt{p})$, and type 3 used $\boldsymbol{A} = \psi\boldsymbol{1}\boldsymbol{1}^T + (1-\psi)\boldsymbol{I}_p$ giving $\text{cor}(x_{ij}, x_{ik}) = \rho$ for $j \neq k$ 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. We used $\delta = 0$ and $\delta > 0$ chosen so at least one test had good power. The simulation used 5000 runs, the 4 $\boldsymbol{x}$ distributions, and the 3 matrices $\boldsymbol{A}$. For the third $\boldsymbol{A}$, we used $\psi = 1/\sqrt{p}$.

Tables 1-3 summarize some simulation results. There are two lines for each simulation scenario. The first line gives the simulated power = proportion of times $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ was rejected. The second line gives the average length of the confidence interval where H_0 is rejected if 0 is not in the confidence interval. When $\delta = 0$, observed coverage between 0.04 and 0.06 suggests coverage = power = level is close to the nominal value 0.05. For larger δ, want the coverage near 1 for good power. See Abid (2025) for more simulations.

The bootstrap test corresponds to the boot column, the tests using $(\overline{w}, S_W)$, $(T_n, \hat{\sigma}_W)$, and (T_n, S_W) correspond to the next three columns. The last column corresponds to the spatial sign test. This test tends to have much shorter lengths because of the transformation of the data. The test using $(\overline{w}, S_W)$ has simple large sample theory, but low power compared to the other methods. This test's length is approximately $\sqrt{n-1}$ times the length of that corresponding to (T_n, S_W) where $\sqrt{99} \approx 10$ in the tables. The bootstrap test was sometimes conservative with observed coverage < 0.04 when delta=0. For xtype=4 and delta=0, H_0 was not true for the spatial test. Hence the coverage for the spatial test was sometimes higher than 0.06 for this scenario. For delta=0, the test with $(T_n, \hat{\sigma}_W)$ sometimes had coverage less than 0.04, while the test with (T_n, S_W) sometimes had coverage greater than 0.06. In the simulations, the spatial test often performed well, but typically $E(\boldsymbol{z}_i) = \boldsymbol{\mu}_{\boldsymbol{z}} \neq \boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x}_i)$, which makes the spatial test harder to use. For testing $H_0 : \boldsymbol{\mu}_{\boldsymbol{x}} = \boldsymbol{0}$, the test with $(T_n, \hat{\sigma}_W)$ appeared to perform better than the three competitors.

5.2 Two Sample Tests

In the simulations, we examined three sample tests. The first "test" used the m out of n bootstrap where $m_i = 2n_i/3$ to bootstrap the Chen and Qin (2010) test that estimates

Table 1: one sample tests, covtyp=1

n	p	psi/xtype	δ	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	100	0	0	0.0230	0.0580	0.0400	0.0452	0.0444
	len	1		0.6732	5.6520	0.5711	0.5681	0.0057
100	100	0	0.075	0.8160	0.0688	0.9216	0.9176	0.9166
	len	1		0.8081	5.7018	0.5741	0.5731	0.0057
100	100	0	0	0.0236	0.0436	0.0466	0.0776	0.0478
	len	2		7.0590	58.2593	6.0094	5.8553	0.0057
100	100	0	0.15	0.1938	0.0506	0.3128	0.3490	0.9988
	len	2		7.5830	58.1417	6.0204	5.8435	0.0057
100	100	0	0	0.0222	0.0466	0.0450	0.0680	0.0468
	len	3		1.3031	10.6946	1.1140	1.0749	0.0057
100	100	0	0.1	0.7536	0.0544	0.8720	0.8714	0.9956
	len	3		1.5563	10.8976	1.1260	1.0953	0.0057
100	100	0	0	0.0206	0.0556	0.0372	0.0656	0.0906
	len	4		3.1105	25.4558	2.6543	2.5584	0.0057
100	100	0	0.17	0.9024	0.0546	0.9622	0.9496	0.7668
	len	4		3.7816	25.5420	2.6708	2.5671	0.0057
100	1000	0	0	0.0236	0.0482	0.0448	0.0506	0.0506
	len	1		2.1403	17.8302	1.8059	1.7920	0.0018
100	1000	0	0.0415	0.872	0.068	0.9438	0.9398	0.9388
	len	1		2.2771	17.9004	1.8089	1.7991	0.0018
100	1000	0	0	0.0236	0.0448	0.0458	0.0712	0.0558
	len	2		22.4434	185.1105	19.0973	18.6043	0.0018
100	1000	0	0.075	0.142	0.0480	0.2222	0.2616	0.9978
	len	2		22.8203	182.6556	18.9772	18.3576	0.0018
100	1000	0	0	0.0214	0.0432	0.0436	0.0650	0.0450
	len	3		4.1649	34.1708	3.5444	3.4343	0.0018
100	1000	0	0.05	0.6458	0.0558	0.7642	0.7770	0.9908
	len	3		4.3708	34.0483	3.5586	3.4220	0.0018
100	1000	0	0	0.0192	0.0544	0.0378	0.0518	0.0484
	len	4		9.9417	82.3953	8.4267	8.2810	0.0018
100	1000	0	0.087	0.8430	0.0576	0.9282	0.9242	0.8774
	len	4		10.5664	82.8816	8.4523	8.3299	0.0018

Table 2: one sample tests, covtyp=2

n	p	psi/xtype	δ	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	100	0	0	0.0212	0.0498	0.0380	0.0430	0.0414
	len	1		38.9543	329.1668	33.2225	33.0825	0.0065
100	100	0	0.6	0.8966	0.0758	0.9560	0.9548	0.9556
	len	1		46.3236	330.7589	33.3672	33.2425	0.0065
100	100	0	0	0.0214	0.0502	0.0398	0.0726	0.0506
	len	2		410.1416	3394.75	350.1749	341.1852	0.0065
100	100	0	1.5	0.5062	0.0526	0.6492	0.6620	1
	len	2		455.0242	3396.337	350.6696	341.3447	0.0066
100	100	0	0	0.0230	0.0410	0.0454	0.0684	0.0474
	len	3		76.2693	629.0579	65.2686	63.2227	0.0065
100	100	0	0.75	0.7550	0.0600	0.8558	0.8608	0.997
	len	3		88.0646	634.0106	65.4900	63.7205	0.0065
100	100	0	0	0.0222	0.0608	0.0420	0.0738	0.1156
	len	4		178.6321	1470.551	153.3266	147.7959	0.0064
100	100	0	1.2	0.8532	0.0492	0.9320	0.9214	0.7410
	len	4		207.835	1459.873	154.4866	146.7227	0.0063
100	1000	0	0	0.0286	0.0476	0.0438	0.0482	0.0490
	len	1		1231.498	10344.15	1043.615	1039.626	0.0021
100	1000	0	0.975	0.8472	0.0648	0.9282	0.9204	0.9208
	len	1		1300.17	10379.01	1045.303	1043.129	0.0021
100	1000	0	0	0.0266	0.0386	0.0470	0.0784	0.0536
	len	2		12929.72	106330.2	11004.27	10686.59	0.0021
100	1000	0	1.5	0.078	0.0388	0.1286	0.1620	0.9474
	len	2		13095.03	106960.8	11016.42	10749.97	0.0021
100	1000	0	0	0.0222	0.0456	0.0446	0.0738	0.0454
	len	3		2387.572	19676.47	2033.522	1977.559	0.0021
100	1000	0	1.25	0.7222	0.0616	0.8276	0.8346	0.9986
	len	3		2514.451	19835.06	2051.272	1993.498	0.0021
100	1000	0	0	0.0268	0.0522	0.0462	0.0630	0.0546
	len	4		5747.818	47479.65	4864.88	4771.884	0.0020
100	1000	0	2.15	0.8958	0.0540	0.9544	0.9466	0.9198
	len	4		6064.615	47527.19	4876.035	4776.662	0.0021

Table 3: one sample tests, covtyp=3

n	p	psi/xtype	δ	boot	$(\overline{w}, S_W)$	$(T_n, \hat{\sigma}_W)$	(T_n, S_W)	spatial
100	1000	0	0	0.0282	0.0490	0.0516	0.056	0.0558
	len	1		2.1401	17.8831	1.8065	1.7973	0.0018
100	1000	0.0316	0	0.0066	0.0426	0.0512	0.0532	0.0500
	len	1		58.4898	591.9678	60.0672	59.495	0.0207
100	1000	0	0.04	0.8196	0.0610	0.9152	0.9124	0.9128
	len	1		2.2646	17.9067	1.8088	1.7997	0.0018
100	1000	0.0316	0.4	0.8342	0.1524	0.9732	0.9740	0.9572
	len	1		241.2136	672.2661	68.1736	67.5653	0.0218
100	1000	0	0	0.0272	0.0438	0.0484	0.0820	0.0522
	len	2		22.3855	182.4873	19.0115	18.3407	0.0018
100	1000	0.0316	0	0.0072	0.0306	0.0524	0.0636	0.0502
	len	2		617.6974	5850.04	625.8249	587.9512	0.0208
100	1000	0	0.1	0.3982	0.0460	0.5330	0.5552	1
	len	2		23.2021	184.5163	19.0359	18.5446	0.0018
100	1000	0.0316	0.7	0.2628	0.0522	0.5570	0.5734	0.9900
	len	2		1373.899	6128.062	652.5555	615.8934	0.0228
100	1000	0	0	0.0276	0.0464	0.0484	0.0742	0.0536
	len	3		4.1547	34.0314	3.5458	3.4203	0.0018
100	1000	0.0316	0	0.0082	0.0430	0.0504	0.0610	0.0482
	len	3		114.0482	1097.655	115.6618	110.3185	0.0207
100	1000	0	0.05	0.6502	0.0638	0.7662	0.7722	0.9924
	len	3		4.3608	34.1752	3.5518	3.4347	0.0018
100	1000	0.0316	0.5	0.7432	0.1282	0.9284	0.9294	0.9880
	len	3		419.3617	1241.326	129.0976	124.7579	0.0224
100	1000	0	0	0.0252	0.0486	0.0432	0.0568	0.0548
	len	4		9.9698	82.6130	8.4362	8.3029	0.0018
100	1000	0.0316	0	0.0068	0.0448	0.0500	0.0512	0.0800
	len	4		272.2052	2776.522	281.1257	279.051	0.0210
100	1000	0	0.09	0.8848	0.0614	0.9534	0.9484	0.9128
	len	4		10.5916	82.9419	8.4411	8.3360	0.0018
100	1000	0.0316	0.75	0.7026	0.0962	0.9192	0.9214	0.7900
	len	4		978.2186	3071.672	310.7018	308.7147	0.0216

Table 4: two sample tests, covtyp=3

(n_1, n_2, σ, p)	xtype	covtype	delta	boot	pair	Li
(100,100,1,100)	1	1	0	0.0246	0.0494	0.0494
len	1	1	0	1.3426	1.1389	1.1389
(100,100,1,100)	1	1	0.1	0.7224	0.8586	0.8586
len	1	1	0.1	1.5789	1.1417	1.1417
(100,200,1,100)	1	1	0	0.0256	0.0456	0.0462
len	1	1	0	1.0019	1.1360	0.8535
(100,200,1,100)	1	1	0.1	0.9166	0.8602	0.9612
len	1	1	0.1	1.2396	1.1432	0.8609

$\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2 = \boldsymbol{\mu}_1^T\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2^T\boldsymbol{\mu}_2 - 2\boldsymbol{\mu}_1^T\boldsymbol{\mu}_2$. The second test was the "paired test" with $m = \min(n_1, n_2)$ and $\boldsymbol{z}_i = \boldsymbol{x}_{1i} - \boldsymbol{x}_{2i}$ for $i = 1, ..., m$. Then apply the one sample test from Theorem 2 to the $\boldsymbol{z}_i$. The third test was the Li (2023) test. Both of these tests used S_W^2 applied to the $\boldsymbol{z}_i$ or the $\boldsymbol{y}_i$.

The simulation used four distribution types where $\boldsymbol{x}_1 = \boldsymbol{A}_1\boldsymbol{y}_1 + \delta\boldsymbol{1}$ and $\boldsymbol{x}_2 = \boldsymbol{A}_2\boldsymbol{y}_2$ where $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ had the same distribution, with $E(\boldsymbol{x}_1) = \delta\boldsymbol{1}$ and $E(\boldsymbol{x}_2) = \boldsymbol{0}$. Type 1 used $\boldsymbol{y} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, type 2 used a mixture distribution $\boldsymbol{y} \sim 0.6N_p(\boldsymbol{0}, \boldsymbol{I}) + 0.4N_p(\boldsymbol{0}, 25\boldsymbol{I})$, type 3 for a multivariate t_4 distribution, and type 4 for a multivariate lognormal distribution where $\boldsymbol{y} = (y_1, ..., y_p)$ with $w_i = \exp(Z)$ where $Z \sim N(0, 1)$ and $y_i = w_i - E(w_i)$ where $E(w_i) = \exp(0.5)$. The covariance matrix type depended on the matrix $\boldsymbol{A}$.

For the covariance types, $Cov(\boldsymbol{x}_1) = \boldsymbol{I}, Cov(x_2) = \sigma^2 Cov(\boldsymbol{x}_1)$ for covtyp=1. $Cov(\boldsymbol{x}_1) = diag(1, 2, ..., p), Cov(\boldsymbol{x}_2) = \sigma^2 Cov(\boldsymbol{x}_1)$ for covtyp=2. $Cov(\boldsymbol{x}_1) = \boldsymbol{I}, Cov(\boldsymbol{x}_2) = \sigma^2 diag(1, 2, ..., p)$ for covtyp=3. Table 4 shows some results. Two lines were used for each simulation scenario, with coverages on the first line and lengths on the second line. When $n_1 = n_2$, the paired test and Li test gave the same results. When n_1/n_2 was not near 1, the Li test had better power and shorter length. Increasing δ could greatly increase the length for the bootstrap test, but the coverage would be 1.

6 CONCLUSIONS

The one sample test statistic T_n estimates $\boldsymbol{\mu}^T\boldsymbol{\mu}$ and $V(T_n)$ is easy to estimate when $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ is true. Under regularity conditions when H_0 is true, Li (2023) proved that $T_n/\sqrt{V(T_n)} \xrightarrow{D} t_k$ as $p \to \infty$ for fixed $n \geq 3$ where $k = 0.5n(n-1) - 1$.

Zhao, Li, Li and Zhang (2024) have an interesting result for the multiple linear regression model

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{11}$$

for $i = 1, ..., n$. Assume that the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid with $E(Y) = \mu_Y$, $E(\boldsymbol{x}) = \boldsymbol{\mu}_{\boldsymbol{x}}$ and nonsingular $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$. Let $Cov(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Then testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is equivalent to testing $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$ with

$\boldsymbol{\mu} = E(\boldsymbol{z}_i) = \boldsymbol{\Sigma}_{\boldsymbol{x}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ where $\boldsymbol{z}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})^T \boldsymbol{\beta}_0)$, and the one sample test from Theorem 2 can be applied to $\boldsymbol{w}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y} - (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{\beta}_0)$. Since $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$, using $\boldsymbol{\beta}_0 = \boldsymbol{0}$ gives both a test for $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ and $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$. See Olive and Quaye (2025) for applications.

For classification with two groups, let $\boldsymbol{\Sigma}$ be the pooled covariance matrix. Then $\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = \boldsymbol{0}$ iff $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{0}$, which can be tested with a two sample test. For the importance of $\boldsymbol{\beta}$ in discriminant analysis, see, for example, Wang, Wu, and Wang (2025).

Let the "fail to reject region" be the compliment of the rejection region. Often the fail to reject region is a confidence region for the parameter or parameter vector of interest, where a confidence interval is a special case of a confidence region. For the one sample test, the fail to reject region using V_0 has much more power than using a confidence interval for $\boldsymbol{\mu}^T \boldsymbol{\mu}$. The two sample test statistic $T_N(\boldsymbol{x}, \boldsymbol{y})$ could be used to get a confidence interval for $\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\|^2$.

The literature for high dimensional one and two sample tests is rather large. Hu, Tong, and Genton (2024) have many references. Some high dimensional one sample tests include Chen et al. (2011), Feng and Sun (2016), Hyodo and Nishiyama (2017), Park and Ayyala (2013), Srivastava and Du (2008), Wang, Peng, and Li (2015), and Zhao (2017). Hu and Bai (2015) also describes some tests. Chakraborty and Chaudhuri (2017) suggest a method for obtaining a k-sample test of $\boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ from a one sample test statistic.

Some high dimensional two sample tests include Ahmad (2014), Chen, Li, and Zhong (2019), Feng and Sun (2015), Gregory et al. (2015), Jiang et al. (2022), Xue and Yao (2020), and Zhang et al. (2020). For more on the use of U-statistics for high dimensional methods, see, for example, Xu, Zhu, and Shao (2024).

Two sample tests that assume $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ may not work well since the assumption of equal covariance matrices rarely holds. This assumption is typically stronger than assuming that $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$. See, for example, Huang et al. (2022), Hu and Bai (2015), and Yang, Zheng, and Li (2024).

Simulations were done in *R*. See R Core Team (2024). The collection of Olive (2025) *R* functions *slpack*, available from (http://parker.ad.siu.edu/Olive/slpack.txt), has some useful functions for the inference. The function *hdhot1sim* was used to simulate the four tests, while the function *hdhot1sim2* simulates the first test, which is rather fast. The function *hdhot1sim3* added the test based on sample signs using the fast test. The function *hdhot2sim* simulates the two sample test which applies the fast one sample test on the $\boldsymbol{z}_i = \boldsymbol{x}_{i1} - \boldsymbol{x}_{i2}$ for $i = 1, ..., m$, the Li (2023) test, and the two sample test based on subsampling with $m_i = floor(2n_i/3)$ for i=1,2.

The spatial sign vectors have a some outlier resistance. If the predictor variables are all continuous, the *covmb2* and *ddplot5* functions are useful for detecting outliers in high dimensions. See Olive (2025, $\oint$ 1.4.3) and Olive (2017, pp. 120-123).

7 References

Abid, A.M. (2025), "Some Simple High Dimensional One and Two Sample Tests," Ph.D. Thesis, Southern Illinois University.

Agresti, A. (2002), *Categorical Data Analysis*, 2nd ed., Wiley, Hoboken, NJ.

Ahmad, M.R. (2014), "A U-Statistic Approach for a High-Dimensional Two-Sample Mean Testing Problem under Non-Normality and Behrens-Fisher Setting," *The Annals of the Institute of Statistical Mathematics*, 66, 33-61.

Anderson, T.W. (1984), *An Introduction to Multivariate Statistical Analysis,* 2nd ed., Wiley, New York, NY.

Bai, Z.D., and Saranadasa, H. (1996), "Effects of High Dimension: by an Example of a Two Sample Problem," *Statistica Sinica*, 6, 311-329.

Basa, J., Cook, R.D., Forzani, L., and Marcos, M. (2024), "Asymptotic Distribution of One-Component Partial Least Squares Regression Estimators in High Dimensions," *The Canadian Journal of Statistics*, 52, 118-130.

Buja, A., and Stuetzle, W. (2006), "Observations on Bagging," *Statistica Sinica*, 16, 323-352.

Cameron, A.C., and Trivedi, P.K. (1998), *Regression Analysis of Count Data*, Cambridge University Press, Cambridge, UK.

Chakraborty, A., and Chaudhuri, P. (2017), "Tests for High-Dimensional Data Based on Means, Spatial Signs, and Spatial Ranks," *The Annals of Statistics*, 45, 771-779.

Chen, C.H., and Li, K.C. (1998), "Can SIR be as Popular as Multiple Linear Regression?," *Statistica Sinica*, 8, 289-316.

Chen, L.S., Paul, D., Prentice, R.L., and Wang, P. (2011), "A Regularized Hotelling's T^2 Test for Pathway Analysis in Proteomic Studies," *Journal of the American Statistical Association*, 106, 1345-1360.

Chen, S., Li, J., and Zhong, P. (2019), "Two-Sample and ANOVA tests for High Dimensional Means," *The Annals of Statistics*, 47, 1443-1474.

Chen, S.X., and Qin, Y.L. (2010), "A Two Sample Test for High-dimensional Data with Applications to Gene-Set Testing," *The Annals of Statistics*, 38, 808-835.

Cook, R.D., and Forzani, L. (2024), *Partial Least Squares Regression: and Related Dimension Reduction Methods*, Chapman and Hall/CRC, Boca Raton, FL.

Cook, R.D., Helland, I.S., and Su, Z. (2013), "Envelopes and Partial Least Squares Regression," *Journal of the Royal Statistical Society, B*, 75, 851-877.

Fan, J., and Lv, J. (2008), "Sure Independence Screening for Ultrahigh Dimensional Feature Space," *Journal of the Royal Statistical Society, B*, 70, 849-911.

Fan, J., and Song, R. (2010), "Sure Independence Screening in Generalized Linear Models with np-Dimensionality," *The Annals of Statistics*, 38, 3217-3841.

Feng, L., and Sun, F. (2015), "A Note on High–Dimensional Two–Sample Test," *Statistics & Probability Letters,* 105, 29-36.

Feng, L., and Sun, F. (2016), "Spatial-Sign Based High-Dimensional Location Test," *Electronic Journal of Statistics*, 10, 2420-2434.

Gregory, K.B., Carroll, R.J., Baladandayuthapani, V., and Lahiri, S.N. (2015), "A Two-Sample Test for Equality of Means in High Dimension," *Journal of the American Statistical Association*, 110, 837-849.

Helland, I.S. (1990), "Partial Least Squares Regression and Statistical Models," *Scandanavian Journal of Statistics*, 17, 97-114.

Hotelling, H. (1931), "A Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2, 360-378.

Hu, J., and Bai, Z. (2015), "A Review of 20 Years of Naive Tests of Significance for High-Dimensional Mean Vectors and Covariance Matrices," *Science China Mathematics*, 55, online.

Hu, Z., Tong, T., and Genton, M.G. (2024), "A Pairwise Hotelling Method for Testing High-Dimensional Mean Vectors," *Statistica Sinica*, 34, 229-256.

Huang, Y., Li, C., Li, R., and Yang, S. (2022), "An Overview of Tests on High-Dimensional Means," *Journal of Multivariate Analysis*, 188, 104813

Hyodo, M., and Nishiyama, T. (2017), "A One-Sample Location Test Based on Weighted Averaging of Two Test Statistics When the Dimension and the Sample Size are Large," *Communications in Statistics: Theory and Methods*, 46, 3526-3541.

Jiang, Y.L., Wang, X.Q., Wen, C.H., Jiang, Y.K., and Zhang, H.P. (2024), "Nonparametric Two-Sample Tests of High Dimensional Mean Vectors via Random Integration," *Journal of the American Statistical Association,* 119, 701-714.

Lehmann, E.L. (1975), *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco.

Lehmann, E.L. (1999), *Elements of Large–Sample Theory*, Springer, New York, NY.

Li, J. (2023), "Finite Sample t-Tests for High-dimensional Means," *Journal of Multivariate Analysis*, 196, 105183.

Li, K.C., and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics,* 17, 1009-1052.

Olive, D.J. (2013), "Plots for Generalized Additive Models," *Communications in Statistics: Theory and Methods*, 42, 2610-2628.

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

Olive, D.J. (2025), *Prediction and Statistical Learning*, online course notes, see (http://parker.ad.siu.edu/Olive/slearnbk.htm).

Olive, D.J., Alshammari, A.A., Pathiranage, K.G., and Hettige, L.A.W. (2025), "Testing with the One Component Partial Least Squares and the Marginal Maximum Likelihood Estimators," *Communications in Statistics: Theory and Methods*, to appear.

Olive, D.J., and Zhang, L. (2025), "One Component Partial Least Squares, High Dimensional Regression, Data Splitting, and the Multitude of Models," *Communications in Statistics: Theory and Methods*, 54, 130-145.

Park, J., and Ayyala, D.N. (2013), "A Test for the Mean Vector in Large Dimension and Small Samples," *Journal of Statistical Planning and Inference*, 143, 929-943.

Pelawa Watagoda, L.C.R., and Olive, D.J. (2021), "Comparing Six Shrinkage Estimators with Large Sample Theory and Asymptotically Optimal Prediction Intervals," *Statistical Papers*, 62, 2407-2431.

Politis, D.N, and Romano, J.P. (1994), "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *The Annals of Statistics*, 22,2031-2050.

R Core Team (2024), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

Rajapaksha, K.W.G.D.H., and Olive, D.J. (2024), "Wald Type Tests with the Wrong Dispersion Matrix," *Communications in Statistics: Theory and Methods*, 53, 2236-2251.

Srivastava, M.S., and Du, M. (2008), "A Test for the Mean Vector with Fewer Observations Than the Dimension," *Journal of Multivariate Analysis*, 99, 386-402.

Wang, L., Peng, B., and Li, R. (2015), "A High-Dimensional Nonparametric Multivariate Test for Mean Vector," *Journal of the American Statistical Association*, 110, 1658-1669.

Wang, Y., Wu, Z., and Wang, C. (2025), "High Dimensional Discriminant Analysis under Weak Sparsity," *Communications in Statistics: Theory and Methods*, 54, 2657-2674.

Wold, H. (1975), "Soft Modelling by Latent Variables: the Non-Linear Partial Least Squares (NIPALS) Approach," *Journal of Applied Probability*, 12, 117-142.

Wu, C.F.J. (1990), "On the Asymptotic Properties of the Jackknife Histogram," *The Annals of Statistics*, 18, 1438-1452.

Xu, T., Zhu, R., and Shao, X. (2024), "On Variance Estimation of Random Forests with Infinite-Order U-Statistics," *Electronic Journal of Statistics*, 18, 2135-2207.

Xue, K., and Yao, F. (2020), "Distribution and Correlation-Free Two-Sample Test of High-Dimensional Mean," *The Annals of Statistics*, 48, 1304-1328.

Yang, S., Zheng, S., and Li, R. (2024), "A New Test for High-Dimensional Two-Sample Mean Problems with Consideration of Correlation Structure," *The Annals of Statistics*, 52, 2217-2240.

Zhang, J.T., Guo, J., Zhou, B., and Cheng, M.Y. (2020), "A Simple Two-Sample Test in High Dimensions Based on L^2-Norm," *Journal of the American Statistical Association*, 115(530):1011-1027.

Zhao, A., Li, C., Li, R., and Zhang, Z. (2024), "Testing High-Dimensional Regression Coefficients in Linear Models," *The Annals of Statistics*, 52, 2034-2058.

Zhao, J. (2017), "A New Test for the Mean Vector in Large Dimension and Small Samples," *Communications in Statistics: Simulation and Computation*, 46, 6115-6128.