High Dimensional Dimensional Reduction

David J. Olive *
Southern Illinois University

November 20, 2025

Abstract

Consider a regression or classification model with response variable Y that depends on the predictors $\boldsymbol{x} = (x_1, ..., x_p)^T$ through the sufficient predictor $SP = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}$. Let n be the number of cases. For a high dimensional model, n/p is small.

KEY WORDS: marginal maximum likelihood estimator, PCA, PLS.

1 INTRODUCTION

High dimensional statistics are used when $n < 5p$ where n is the sample size and p is the number of variables. Such a model is *overfitting*: the model does not have enough data to estimate p parameters accurately. Then n tends to not be large enough for the classical statistical method to be useful. An alternative (but less general) definition of high dimensional statistics is that p is large. Sometimes $p > Kn$ with $K \geq 10$ is called ultrahigh dimensional statistics.

In high dimensions, it is very difficult to estimate a $p \times 1$ vector $\boldsymbol{\theta}$. This result is a form of "the curse of dimensionality." If a $\sqrt{n}$ consistent estimator of $\boldsymbol{\theta}$ is available, then the squared norm

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 = \sum_{i=1}^{p} (\hat{\theta}_i - \theta_i)^2 \propto p/n. \tag{1}$$

When p is fixed, $p/n \to 0$ as $n \to \infty$ and $\hat{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}$. In high dimensions, often the estimator has not been shown to be consistent, except under very strong regularity conditions.

Some important statistical methods include regression, multivariate statistics, and classification. These methods are important for statistical learning $\approx$ machine learning, an important part of artificial intelligence. Let predictor variables for regression or multivariate statistics be $\boldsymbol{x} = (x_1, ..., x_p)^T$. Let Y be a response variable for regression or

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

classification. Important regression models include generalized linear models, nonlinear regression, nonparametric regression, and survival regression models. There are n cases $(Y_i, \boldsymbol{x}_i^T)^T$, and for some important models, Y depends on $\boldsymbol{x}$ through the sufficient predictor $SP = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}$. Some important classification models include binary regression, linear discriminant analysis, and quadratic discriminant analysis.

A binary regression model is $Y = Y|SP \sim \text{binomial}(1, \rho(\text{SP}))$ where $\rho(SP) = P(Y = 1|SP)$. There are many binary regression models, including binary logistic regression, binary probit regression, and support vector machines (SVMs) (with $Z_i = 2Y_i - 1$).

Let the multiple linear regression (MLR) model

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{2}$$

for $i = 1, ..., n$. In matrix form, this model is $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\delta} + \boldsymbol{e}$, where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times (p+1)$ matrix of predictors, $\boldsymbol{\delta} = (\alpha, \boldsymbol{\beta}^T)^T$ is a $(p+1) \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. Assume that the e_i are independent and identically distributed (iid) with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma^2$. A multiple linear regression model with heterogeneity has the zero mean e_i independent with $V(e_i) = \sigma_i^2$.

For estimation with ordinary least squares, let the covariance matrix of $\boldsymbol{x}$ be $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T]$ and the $p \times 1$ vector $\boldsymbol{\eta} = \text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = E[(\boldsymbol{x} - E(\boldsymbol{x})(Y - E(Y))] = (\text{Cov}(x_1, Y), ..., \text{Cov}(x_p, Y))^T$. Let the sample covariance matrix be

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T.$$

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 (2) 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 $E(Y) - \boldsymbol{\beta}_{OLS}^T E(\boldsymbol{x})$.

Let the population correlation $\rho_{ij} = \rho_{x_i,x_j} = \text{Cor}(x_i, x_j)$ and the sample correlation $r_{ij} = r_{x_i,x_j} = \text{cor}(x_i, x_j)$. Let the population correlation matrices $\text{Cor}(\boldsymbol{x}) = \boldsymbol{\rho}_{\boldsymbol{x}} = (\rho_{ij})$ and $\text{Cor}(\boldsymbol{x}, Y) = \boldsymbol{\rho}_{\boldsymbol{x}Y} = (\rho_{x_1,Y}, ..., \rho_{x_p,Y})^T$. Let the sample covariance matrices be $\boldsymbol{R}_{\boldsymbol{x}} = (r_{ij})$ and $\boldsymbol{r}_{\boldsymbol{x}Y} = (r_{x_1,Y}, ..., r_{x_p,Y})^T$. Then $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$ and $\boldsymbol{R}$ are dispersion estimators, and $(\overline{\boldsymbol{x}}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})$ is an estimator of multivariate location and dispersion.

Suppose the positive semidefinite dispersion matrix $\boldsymbol{\Sigma}$ has eigenvalue eigenvector pairs $(\lambda_1, \boldsymbol{d}_1), ..., (\lambda_p, \boldsymbol{d}_p)$ where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$. Let the eigenvalue eigenvector pairs of $\hat{\boldsymbol{\Sigma}}$

be $(\hat{\lambda}_1, \hat{\boldsymbol{d}}_1), ..., (\hat{\lambda}_p, \hat{\boldsymbol{d}}_p)$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$. These vectors are important quantities for principal component analysis (PCA).

Principal components regression (PCR), partial least squares (PLS), and several other dimension reduction models use p linear combinations $\boldsymbol{\gamma}_1^T\boldsymbol{x}, ..., \boldsymbol{\gamma}_p^T\boldsymbol{x}$. Estimating the $\boldsymbol{\gamma}_i$ and performing the ordinary least squares (OLS) regression of Y on $(\hat{\boldsymbol{\gamma}}_1^T\boldsymbol{x}, \hat{\boldsymbol{\gamma}}_2^T\boldsymbol{x}, ..., \hat{\boldsymbol{\gamma}}_k^T\boldsymbol{x})$ and a constant gives the k-component estimator, e.g. the k-component PLS estimator or the k-component PCR estimator, for $k = 1, ..., J$ where $J \leq p$ and the p-component estimator is the OLS estimator $\hat{\boldsymbol{\beta}}_{OLS}$. Let $\boldsymbol{\gamma}_i(PCR) = \boldsymbol{d}_i$ and $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i(PLS)$. The model selection estimator chooses one of the k-component estimators, e.g. using cross validation, and will be denoted by $\hat{\boldsymbol{\beta}}_{MSPLS}$ or $\hat{\boldsymbol{\beta}}_{MSPCR}$.

The k-component partial least squares estimator can be found by regressing Y on a constant and on $W_i = \hat{\boldsymbol{\gamma}}_i^T\boldsymbol{x}$ for $i = 1, ..., k$ where $\hat{\boldsymbol{\gamma}}_i = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{i-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ for $i = 1, ..., k$. See Helland (1990). Let $\boldsymbol{X} = [\boldsymbol{1}\ \boldsymbol{X}_1]$. Chun and Keleş (2010) noted that one way to formulate PLS is to solve an optimization problem by forming $\boldsymbol{b}_j = \hat{\boldsymbol{\gamma}}_j$ iteratively where

$$\boldsymbol{b}_k = \arg\max_{\boldsymbol{b}} \{[\mathrm{Cor}(\boldsymbol{Y}, \boldsymbol{X}_1\boldsymbol{b})]^2 V(\boldsymbol{X}_1\boldsymbol{b})\} \tag{3}$$

subject to $\boldsymbol{b}^T\boldsymbol{b} = 1$ and $\boldsymbol{b}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{b}_j = 0$ for $j = 1, ..., k-1$. Here V stands for the variance. So PLS is a model free way to get predictors $\hat{\boldsymbol{\gamma}}_i^T\boldsymbol{x}$ that are fairly highly correlated with the response, and the absolute correlations tend to decrease quickly. Brown (1993, pp. 71-72) shows that an equivalent way to compute the k-component PLS estimator is to maximize $\hat{\boldsymbol{\gamma}}^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ under some constraints. If the predictors are standardized to have unit sample variance, then this method becomes a correlation vector optimization problem.

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

2 What Are Some Dimension Reduction Estimators Estimating?

Several dimension reduction methods use p linear combinations $\hat{\boldsymbol{\gamma}}_1^T\boldsymbol{x}, ..., \hat{\boldsymbol{\gamma}}_p^T\boldsymbol{x}$. PCA and PLS are interesting since these two methods can be used with high dimensional data. Let $W_i = \hat{\boldsymbol{\gamma}}_i^T\boldsymbol{x}$ for $i = 1, ..., p$. For PLS, let $\hat{\boldsymbol{\gamma}}_i = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{j-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^0 = \boldsymbol{\Sigma}_{\boldsymbol{x}}^0 = \boldsymbol{I}_p$. For PCA, let $\hat{\boldsymbol{d}}_i$ be orthogonal eigenvectors of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$ where the $\boldsymbol{d}_i$ are orthogonal eigenvectors of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$. Then $\hat{\boldsymbol{\gamma}}_i = \hat{\boldsymbol{d}}_i$. In low dimensions, envelope methods have some optimality properties, and can be used with more than one response variable. See, for example, Cook and Forzani (2024).

In low dimensions with one response variable, canonical correlation analysis (CCA) also has some optimality properties. For CCA, if $(Y_i, \boldsymbol{x}_i^T)^T$ are iid with $V(Y) = \Sigma_Y$, then

$$M = \max_{\boldsymbol{\gamma} \neq \boldsymbol{0}} \mathrm{Cor}(\boldsymbol{\gamma}^T\boldsymbol{x}, Y) = \max_{\boldsymbol{\gamma} \neq \boldsymbol{0}} \frac{\boldsymbol{\gamma}^T\boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\sqrt{\Sigma_Y}\sqrt{\boldsymbol{\gamma}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\gamma}}}.$$

This optimization problem is equivalent to maximizing

$$\boldsymbol{\Sigma}_Y M^2 = \max_{\boldsymbol{\gamma} \neq \mathbf{0}} \frac{\boldsymbol{\gamma}^T \boldsymbol{\Sigma}_{\boldsymbol{x}Y} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\gamma}}{\boldsymbol{\gamma}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\gamma}}$$

which has a maximum at $\boldsymbol{\gamma} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\beta}_{OLS}$. Hence $\hat{\gamma}_1 = \hat{\boldsymbol{\beta}}_{OLS}$ for CCA of Y and $x_1, ..., x_p$. See Mardia, Kent, and Bibby (1979, pp. 168, 282). Hence PLS is a lot like CCA for $(Y_i, \boldsymbol{x}_i^T)^T$ but with more constraints, and PLS can be computed in high dimensions. From the dimension reduction literature, if Y depends on $\boldsymbol{x}$ only through $\alpha + \boldsymbol{\beta}^T \boldsymbol{x}$, then under the assumption of "linearly related predictors," $\hat{\boldsymbol{\beta}}_{OLS}$ estimates $\boldsymbol{\beta}_{OLS} = c\boldsymbol{\beta}$ for some constant c which is often nonzero. See, for example, Cook and Weisberg (1999, p. 432). Note that in high dimensions, $\hat{\gamma}_1 = \hat{\boldsymbol{\beta}}_{OLS}$ can be replaced by $\hat{\gamma}_1 = \hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is a high dimensional multiple linear regression estimator, such as lasso.

Instead of using the response variable Y and the predictors $X_1, ..., X_p$, the regression model or classification model can use Y and the predictors $W_1, ..., W_k$. Then the k-component estimator $(\hat{\alpha}_k, \hat{\boldsymbol{\beta}}_k)$ is obtained by fitting the working model

$$WSP = \alpha_k + \theta_1 W_1 + \cdots + \theta_k W_k = \alpha_k + \boldsymbol{\theta}_k^T \boldsymbol{w}$$

where $\boldsymbol{\theta}_k = (\theta_1, ..., \theta_k)^T$ and $\boldsymbol{w} = (W_1, ..., W_k)^T$. For the k-component estimator, assume the $k \times p$ matrix

$$\hat{\boldsymbol{A}}_{k,n} = \hat{\boldsymbol{A}}_k = \begin{pmatrix} \hat{\boldsymbol{\gamma}}_1^T \\ \vdots \\ \hat{\boldsymbol{\gamma}}_k^T \end{pmatrix} \xrightarrow{P} \boldsymbol{A}_k = \begin{pmatrix} \boldsymbol{\gamma}_1^T \\ \vdots \\ \boldsymbol{\gamma}_k^T \end{pmatrix}.$$

Then $\hat{\boldsymbol{A}}_k \boldsymbol{x} = \boldsymbol{w} = (W_1, ..., W_k)^T$, and $ESP(\boldsymbol{w}) = \hat{\alpha}_k + \hat{\boldsymbol{\theta}}_k^T \boldsymbol{w} = \hat{\alpha}_k + \hat{\boldsymbol{\theta}}_k^T \hat{\boldsymbol{A}}_k \boldsymbol{x} = \hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^T \boldsymbol{x} = ESP(\boldsymbol{x})$ with $\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{A}}_k^T \hat{\boldsymbol{\theta}}_k$. Assume $\hat{\boldsymbol{\theta}}_k \xrightarrow{P} \boldsymbol{\theta}_k$.

For example, fit a GLM, logistic regression, a support vector machine, multiple linear regression, et cetera. The θ_i depend on the method used to fit the working model. (Also, using θ_{ki} instead of θ_i is more accurate, but suppressing the subscript k is convenient.) Parts e), f), and g) of Theorem 1 and part b) of Theorem 2 are new.

Theorem 1. Consider the above notation.

a) $ESP = \hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^T \boldsymbol{x} = \hat{\alpha}_k + (\sum_{j=1}^k \hat{\theta}_j \hat{\boldsymbol{\gamma}}_j^T) \boldsymbol{x}$.

b) $SP = \alpha_k + \boldsymbol{\beta}_k^T \boldsymbol{x} = \alpha_k + (\sum_{j=1}^k \theta_j \boldsymbol{\gamma}_j^T) \boldsymbol{x}$.

c) $\hat{\boldsymbol{\beta}}_k = \sum_{j=1}^k \hat{\theta}_j \hat{\boldsymbol{\gamma}}_j = \hat{\boldsymbol{A}}_k^T \hat{\boldsymbol{\theta}}_k$.

d) $\boldsymbol{\beta}_k = \sum_{j=1}^k \theta_j \boldsymbol{\gamma}_j = \boldsymbol{A}_k^T \boldsymbol{\theta}_k$.

e) $\hat{\boldsymbol{\beta}}_{kPLS} = (\sum_{j=1}^k \hat{\theta}_j \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{j-1}) \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$

f) Under iid cases, $\boldsymbol{\beta}_{kPLS} = (\sum_{j=1}^k \theta_j \boldsymbol{\Sigma}_{\boldsymbol{x}}^{j-1}) \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$.

g) If $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \xrightarrow{P} \boldsymbol{V}_{\boldsymbol{x}}$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{P} \boldsymbol{V}_{\boldsymbol{x}Y}$, then $\hat{\boldsymbol{\beta}}_{kPLS} \xrightarrow{P} \boldsymbol{\beta}_{kPLS} = (\sum_{j=1}^k \theta_j \boldsymbol{V}_{\boldsymbol{x}}^{j-1}) \boldsymbol{V}_{\boldsymbol{x}Y}$.

Proof. Fit WSP to get the $ESP = \hat{\alpha}_k + \hat{\theta}_1 W_1 + \cdots + \hat{\theta}_k W_k = \hat{\alpha}_k + \hat{\theta}_1 \boldsymbol{\gamma}_1^T \boldsymbol{x} + \cdots + \hat{\theta}_k \boldsymbol{\gamma}_k^T \boldsymbol{x} = \hat{\alpha}_k + \hat{\boldsymbol{\beta}}_k^T \boldsymbol{x}$. Equating terms gives the result.

When the cases are not iid, $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ may be estimating $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS} \neq \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. When the errors e_i are iid, a common assumption for OLS MLR theory is

$$n(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \hat{\boldsymbol{V}} = \begin{pmatrix} \hat{\boldsymbol{V}}_{11} & \hat{\boldsymbol{V}}_{12} \\ \hat{\boldsymbol{V}}_{21} & \hat{\boldsymbol{V}}_{22} = n\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}/(n-1) \end{pmatrix} \xrightarrow{P} \boldsymbol{V} = \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{21} & \boldsymbol{V}_{22} \end{pmatrix}.$$

Thus $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \xrightarrow{P} \boldsymbol{V}_{22}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \xrightarrow{P} \boldsymbol{V}_{22}^{-1}$, and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{P} \boldsymbol{V}_{22}^{-1}\boldsymbol{\beta}$ since $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{P} \boldsymbol{\beta}$.

Remark 1. The following result is useful for several multiple linear regression estimators. Let $\boldsymbol{w}_i = \boldsymbol{A}_n\boldsymbol{x}_i$ for $i = 1, ..., n$ where $\boldsymbol{A}_n$ is a full rank $k \times p$ matrix with $1 \leq k \leq p$.

a) Let $\boldsymbol{\Sigma}^*$ be $\hat{\boldsymbol{\Sigma}}$ or $\tilde{\boldsymbol{\Sigma}}$. Then $\boldsymbol{\Sigma}_{\boldsymbol{w}}^* = \boldsymbol{A}_n\boldsymbol{\Sigma}_{\boldsymbol{x}}^*\boldsymbol{A}_n^T$ and $\boldsymbol{\Sigma}_{\boldsymbol{w}Y}^* = \boldsymbol{A}_n\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^*$.

b) If $\boldsymbol{A}_n$ is a constant matrix, then $\boldsymbol{\Sigma}_{\boldsymbol{w}} = \boldsymbol{A}_n\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{A}_n^T$ and $\boldsymbol{\Sigma}_{\boldsymbol{w}Y} = \boldsymbol{A}_n\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$.

The following result is known. Using the notation above Theorem 1, let $\hat{\boldsymbol{A}}_k\boldsymbol{x} = \boldsymbol{w}$. For multiple linear regression with OLS, let $Y = \alpha + \boldsymbol{x}^T\boldsymbol{\beta} + e$. Let the working model be $Y = \alpha_k + \boldsymbol{\theta}_k^T\boldsymbol{w} + \epsilon$. Then the OLS estimator $\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}Y}$. By Remark 1, the k-component estimator

$$\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{A}}_k^T\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{A}}_k^T(\hat{\boldsymbol{A}}_k\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{A}}_k^T)^{-1}\hat{\boldsymbol{A}}_k\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}.$$

Suppose $k = p$ and $\hat{\boldsymbol{A}}_p^{-1}$ exists. Then $\hat{\boldsymbol{\beta}}_p = \hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}$ since $\hat{\boldsymbol{\beta}}_p =$

$$\hat{\boldsymbol{A}}_p^T(\hat{\boldsymbol{A}}_p\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{A}}_p^T)^{-1}\hat{\boldsymbol{A}}_p\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y} = \hat{\boldsymbol{A}}_p^T(\hat{\boldsymbol{A}}_p^T)^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}(\hat{\boldsymbol{A}}_p)^{-1}\hat{\boldsymbol{A}}_p\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}.$$

The following theorem shows that if the p components W_i are plugged into a model that uses maximum likelihood estimation, such as a GLM, the p-component estimator $\hat{\boldsymbol{\beta}}_p = \hat{\boldsymbol{\beta}}_{\boldsymbol{x}}$, the MLE. Similar theory holds for other maximization or minimization problems, such as quasi-likelihood and partial likelihood. The profile likelihood function $L_p(\boldsymbol{\beta}_{\boldsymbol{x}}|\boldsymbol{x}) = L(\boldsymbol{\beta}_{\boldsymbol{x}}, \hat{\boldsymbol{\eta}}|\boldsymbol{x})$ where L is the likelihood function of all of the parameters $(\boldsymbol{\beta}_{\boldsymbol{x}}, \boldsymbol{\eta})$ and $\hat{\boldsymbol{\eta}}$ is the MLE of $\boldsymbol{\eta}$. As above, use $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ to denote the MLE with $\boldsymbol{w}$ instead of $\hat{\boldsymbol{\beta}}_{\boldsymbol{w}}$.

Theorem 2. Suppose the profile likelihood function $L_p(\boldsymbol{\beta}_{\boldsymbol{x}}|\boldsymbol{x}) = \prod_{i=1}^n f(\boldsymbol{x}_i|\boldsymbol{\beta}_{\boldsymbol{x}}) = \prod_{i=1}^n g(\boldsymbol{x}_i^T\boldsymbol{\beta}_{\boldsymbol{x}})$ depends on $\boldsymbol{x}$ and $\boldsymbol{\beta}_{\boldsymbol{x}}$ only through $\boldsymbol{x}^T\boldsymbol{\beta}_{\boldsymbol{x}}$. a) If the maximum likelihood estimator is computed using $\boldsymbol{w} = \hat{\boldsymbol{A}}_p\boldsymbol{x}$ instead of $\boldsymbol{x}$, then $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} = \hat{\boldsymbol{A}}_p^T\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}_p$ provided that $\hat{\boldsymbol{A}}_p$ is nonsingular. b) Thus $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} = \hat{\boldsymbol{\beta}}_{pPLS} = (\sum_{j=1}^p \hat{\theta}_j\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{j-1})\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ if the PLS components are used.

Proof. a)

$$L_p(\boldsymbol{\theta}|\boldsymbol{w}) = \prod_{i=1}^n g(\boldsymbol{w}_i^T\boldsymbol{\theta}) = \prod_{i=1}^n g(\boldsymbol{x}_i^T\hat{\boldsymbol{A}}^T\boldsymbol{\theta}) = \prod_{i=1}^n g(\boldsymbol{x}_i^T\boldsymbol{\beta}^*).$$

Since the second to last term is maximized by $\hat{\boldsymbol{A}}^T\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}_p$ and the last term is maximized by $\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}_{\boldsymbol{x}}$, it follows that $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} = \hat{\boldsymbol{A}}^T\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}_p$, and $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{A}}^T)^{-1}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}$. Nonsingularity was used so that $\boldsymbol{\beta}^*$ varies through $\mathbb{R}^p$ as $\boldsymbol{\beta}_{\boldsymbol{x}}$ varies through $\mathbb{R}^p$.

b) Plug in $\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} = \hat{\boldsymbol{\beta}}_p = \hat{\boldsymbol{\beta}}_{pPLS}$ from Theorem 1 e). $\square$

A useful high dimensional technique is to use PCA for dimension reduction. Let $U_1, ..., U_p$ be the PCA linear combinations ($U_i = \hat{\boldsymbol{\gamma}}_i^T \boldsymbol{x}$) ordered with respect to the largest eigenvalues. Then use $U_1, ..., U_k$ in the regression or classification model where k is chosen in some manner. This method can be used for models with m response variables $Y_1, ..., Y_m$. See, for example, Artigue and Smith (2019), Cook (2007, 2018), and Zhang and Chen (2020).

Consider a low or high dimensional regression or classification method with a univariate response variable Y. Let $W_1, ..., W_p$ be the linear combinations ordered with respect to the highest squared correlations $r_1^2 \geq r_2^2 \geq \cdots \geq r_p^2$ where the sample correlation $r_{i,Y} = \text{cor}(x_i, Y)$. As noted by Olive (2025), from a model selection viewpoint, using $W_1, ..., W_k$ should work much better than using $U_1, ..., U_k$. Also, the PLS components W_i should be used instead of the PCA W_i, since the PLS components are chosen to be fairly highly correlated with Y. Cook and Forzani (2021) used the PLS components as predictors for nonlinear regression.

3 The OPLS Estimator

The OPLS estimator is the PLS estimator from Section 2 with $k = 1$. Then the ESP $= \hat{\alpha}_1 + \hat{\theta}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T \boldsymbol{x} = \hat{\alpha}_{OPLS} + \hat{\boldsymbol{\beta}}_{OPLS}^T \boldsymbol{x}$ where $\hat{\boldsymbol{\beta}}_{OPLS} = \hat{\theta}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. Let $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. Testing $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} \neq \boldsymbol{0}$ is equivalent to testing $H_0 : \boldsymbol{A}\boldsymbol{\eta} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A}\boldsymbol{\eta} \neq \boldsymbol{0}$ where $\boldsymbol{A}$ is a $k \times p$ constant matrix and $\boldsymbol{\eta} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$.

For multiple linear regression, Cook, Helland, and Su (2013) and Basa et al. (2024) showed that $\hat{\boldsymbol{\beta}}_{OPLS} = \hat{\theta}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ estimates $\theta\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\beta}_{OPLS}$ where

$$\theta = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}} \text{ and } \hat{\theta} = \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}}{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}} \tag{4}$$

for $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq \boldsymbol{0}$. If $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$, then $\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$.

Next, some large sample theory is reviewed for $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ and OPLS for the multiple linear regression model, including some high dimensional tests for low dimensional quantities such as $H_0 : \beta_i = 0$ or $H_0 : \beta_i - \beta_j = 0$. These tests depended on iid cases, but not on linearity or the constant variance assumption. Hence the tests are useful for multiple linear regression with heterogeneity.

The following Olive and Zhang (2025) theorem gives the large sample theory for $\hat{\boldsymbol{\eta}} = \widehat{\text{Cov}}(\boldsymbol{x}, Y)$. Olive et al. (2025) gave alternative proofs. This theory needs $\boldsymbol{\eta} = \boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$ to exist for $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}$ to be a consistent estimator of $\boldsymbol{\eta}$. Let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^T$ and let $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ be defined below where

$$\text{Cov}(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{w}} = E[(\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})^T (Y_i - \boldsymbol{\mu}_Y)^2)] - \boldsymbol{\Sigma}_{\boldsymbol{x}Y} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T.$$

Then the low order moments are needed for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}$ to be a consistent estimator of $\boldsymbol{\Sigma}_{\boldsymbol{w}}$.

Theorem 3. Assume the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid. Assume $E(x_{ij}^k \, Y_i^m)$ exist for $j = 1, \ldots, p$ and $k, m = 0, 1, 2$. Let $\boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x})$ and $\mu_Y = E(Y)$. Let $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \boldsymbol{\mu}_Y)$

with sample mean $\overline{\boldsymbol{w}}_n$. Let $\boldsymbol{\eta} = \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$. Then (a)

$$\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}), \ \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}), \tag{5}$$

$$and \ \ \sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}).$$

(b) Let $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_n)(Y_i - \overline{Y}_n)$. Then $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}} + O_P(n^{-1/2})$. Hence $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}} + O_P(n^{-1/2})$.

(c) Let $\boldsymbol{A}$ be a $k \times p$ full rank constant matrix with $k \leq p$, assume $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$ is true, and assume $\hat{\theta} \xrightarrow{P} \theta \neq 0$. Then

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) \xrightarrow{D} N_k(\boldsymbol{0}, \theta^2 \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{A}^T). \tag{6}$$

For the following theorem, consider a subset of k distinct elements from $\tilde{\boldsymbol{\Sigma}}$ or from $\hat{\boldsymbol{\Sigma}}$. Stack the elements into a vector, and let each vector have the same ordering. For example, the largest subset of distinct elements corresponds to

$$vech(\tilde{\boldsymbol{\Sigma}}) = (\tilde{\sigma}_{11}, \ldots, \tilde{\sigma}_{1p}, \tilde{\sigma}_{22}, \ldots, \tilde{\sigma}_{2p}, \ldots, \tilde{\sigma}_{p-1,p-1}, \tilde{\sigma}_{p-1,p}, \tilde{\sigma}_{pp})^T = [\tilde{\sigma}_{jk}].$$

For random variables $x_1, \ldots, x_p$, use notation such as $\overline{x}_j$ = the sample mean of the x_j, $\mu_j = E(x_j)$, and $\sigma_{jk} = Cov(x_j, x_k)$. Let

$$n \ vech(\tilde{\boldsymbol{\Sigma}}) = [n \ \tilde{\sigma}_{jk}] = \sum_{i=1}^{n} [(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)].$$

For general vectors of elements, the ordering of the vectors will all be the same and be denoted by vectors such as $\hat{\boldsymbol{c}} = [\hat{\sigma}_{jk}]$, $\tilde{\boldsymbol{c}} = [\tilde{\sigma}_{jk}]$, $\boldsymbol{c} = [\sigma_{jk}]$, $\boldsymbol{v}_i = [(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)]$, and $\boldsymbol{w}_i = [(x_{ij} - \mu_j)(x_{ik} - \mu_k)]$. Let $\overline{\boldsymbol{w}}_n = \sum_{i=1}^{n} \boldsymbol{w}_i/n$ be the sample mean of the $\boldsymbol{w}_i$. Assuming that $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{w}}$ exists, then $E(\boldsymbol{w}_i) = E(\overline{\boldsymbol{w}}_n) = \boldsymbol{c}$.

The following Olive et al. (2025) theorem provides large sample theory for $\hat{\boldsymbol{c}}$ and $\tilde{\boldsymbol{c}}$. We use $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{d}}$ to avoid confusion with the $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ used in Theorem 3. Note that $\boldsymbol{x}_i$ are dummy variables and could be replaced by $\boldsymbol{u}_i = (Y_{i1}, \ldots, Y_{im}, x_{i1}, \ldots, x_{ip})^T$ to get information about m response variables $Y_1, \ldots, Y_m$.

Theorem 4. Assume the cases $\boldsymbol{x}_i$ are iid and that $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{d}}$ exists. Using the above notation with $\boldsymbol{c}$ a $k \times 1$ vector,

(i) $\sqrt{n}(\tilde{\boldsymbol{c}} - \boldsymbol{c}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{d}})$.

(ii) $\sqrt{n}(\hat{\boldsymbol{c}} - \boldsymbol{c}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{d}})$.

(iii) $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{d}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{v}} + O_P(n^{-1/2})$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{d}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}} + O_P(n^{-1/2})$.

4 Large Sample Theory and Testing

Suppose the classification or regression model has a response variable Y that depends on the predictors $\boldsymbol{x}$ through $SP = \alpha + \boldsymbol{\beta}^T\boldsymbol{x}$. In low dimensions, important tests include a) $H_0 : \beta_i = 0$ (the Wald tests for MLR), b) $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ (the Anova F test for MLR), and c) $H_0 : (\beta_{i_1}, \ldots, \beta_{i_k})^T = \boldsymbol{0}$ (the partial F test for MLR).

This section will derive some high dimensional analogs of the above tests.

4.1 Testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$

An Omnibus or Universal Test

This subsection follows Abid, Quaye, and Olive (2025) closely. Consider classification and 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 $\mathrm{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Assume the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid random vectors for $i = 1, ..., n$. Then for many such regression models, $\boldsymbol{\beta} = \mathbf{0}$ if and only if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$ where $\mathbf{0} = (0, ..., 0)^T$ is the $p \times 1$ vector of zeroes.

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

Zhao et al. (2024) have an interesting result for the multiple linear regression model (2). 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 $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$. Then testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is equivalent to testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ with $\boldsymbol{\mu} = E(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{x}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ where $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})^T\boldsymbol{\beta}_0)$, and a one sample test can be applied to $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y} - (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T\boldsymbol{\beta}_0)$.

Abid, Quaye, and Olive (2025) used the above test for $\boldsymbol{\beta}_0 = \mathbf{0}$. 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 $p \times p$ nonsingular matrix. Then $\boldsymbol{\beta}_D = \mathbf{0}$ if and only if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$. Then $\boldsymbol{D}^{-1} = \theta\boldsymbol{I}$ for OPLS, $\boldsymbol{D}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}$ for OLS, and $\boldsymbol{D}^{-1} = [diag(\boldsymbol{\Sigma}_{\boldsymbol{x}})]^{-1}$ for the MMLE for multiple linear regression (MLR). By Theorem 1e), $\boldsymbol{\beta}_{kPLS} = \mathbf{0}$ if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$. Thus if the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid, then using $\boldsymbol{\beta}_0 = \mathbf{0}$ gives tests for $H_0 : \boldsymbol{\beta} = \mathbf{0}$, $H_0 : \boldsymbol{\beta}_{MMLE} = \mathbf{0}$ (for MLR), $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$, $H_0 : \boldsymbol{\beta}_{OPLS} = \mathbf{0}$, and $H_0 : \boldsymbol{\beta}_{kPLS} = \mathbf{0}$. 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.

Assume the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid. 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 $\boldsymbol{\beta} = \mathbf{0}$, then the Y_i are iid and do not depend on $\boldsymbol{x}$, and thus satisfy a multiple linear regression model with $\boldsymbol{\beta}_{OLS} = \mathbf{0}$. Typically, if $\boldsymbol{\beta} \neq \mathbf{0}$, then $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq 0$. Also see Theorem 2 b). 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 $\boldsymbol{\beta} \neq \mathbf{0}$.

If $\boldsymbol{\beta}_0 = \mathbf{0}$, then $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \mu_Y)$, and $E(\boldsymbol{w}_i) = E(\boldsymbol{u}_i) = E[\boldsymbol{x}_i(Y_i - \mu_Y)] = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Then apply a high dimensional one sample test on the $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})$. Note that the sample mean $\overline{\boldsymbol{v}} = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$.

Suppose $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ are iid random vectors with $E(\boldsymbol{x}) = \boldsymbol{\mu}$ and covariance matrix $\mathrm{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}$. Then the test $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ is equivalent to the test

$H_0 : \boldsymbol{\mu}^T\boldsymbol{\mu} = 0$ versus $H_1 : \boldsymbol{\mu}^T\boldsymbol{\mu} \neq 0$. Let $\boldsymbol{S} = \hat{\boldsymbol{\Sigma}}$. 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. See, for example, Abid, Quaye, and Olive (2025).

Let the variance $V(W) = V(W_{ij}) = V(\boldsymbol{x}_i^T\boldsymbol{x}_j) = \sigma_W^2$ for $i \neq j$. 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 S_W^2 be the sample variance of the W_i:

$$S_W^2 = \frac{1}{m-1} \sum_{i=1}^{m} (W_i - \overline{W})^2. \tag{8}$$

Zhao et al. (2024, p. 2024) showed that $\sigma_W^2 = tr(\boldsymbol{\Sigma}^2) + 2\boldsymbol{\mu}^T\boldsymbol{\Sigma}\boldsymbol{\mu}$.

The following Abid, Quaye, and Olive (2025) theorem derived the variance $V(T_n)$ under simpler regularity conditions than those in the literature. The second formula in Theorem 5a) was obtained by Chen and Qin (2010).

Theorem 5. 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

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

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

$$V_0 = V(T_n) = \frac{2\sigma_W^2}{n(n-1)} = \frac{2tr(\boldsymbol{\Sigma}^2)}{n(n-1)} = \frac{2\sigma_W^2 - 4\theta}{n(n-1)}.$$

Let $\hat{V}(T_n)$ and $\hat{V}_0(T_n)$ be consistent estimators of $V(T_n)$ and $V_0(T_n)$, respectively. Then 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,

$$T_n / \sqrt{\hat{V}(T_n)} = T_n / \sqrt{\hat{V}_0(T_n)} \xrightarrow{D} N(0, 1).$$

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 $k = 0.5n(n-1) - 1$.

A consistent estimator of $V_0(T_n)$ needs a consistent estimator of $\sigma_W^2 = 0.5n(n-1)$ $V_0(T_n)$. Let $s_n^2 = \hat{V}_0(T_n)$. Then one estimator is $0.5n(n-1)s_n^2 = S_W^2$ from Equation (8). An estimator nearly the same as the one used by Li (2023) is

$$0.5n(n-1)s_n^2 = \hat{\sigma}_W^2 = \frac{1}{n(n-1)} \sum \sum_{i \neq j} (\boldsymbol{x}_i^T\boldsymbol{x}_j - T_n)^2 = \frac{1}{n(n-1)} \sum \sum_{i \neq j} (W_{ij} - T_n)^2.$$

A New Competing Test

If the parametric distribution D is known, then the iid cases assumption can be changed to independent cases. Assume $Y_i|\boldsymbol{x}_i^T\boldsymbol{\beta} \sim D(\tau(\alpha + \boldsymbol{x}_i^T\boldsymbol{\beta}), \boldsymbol{\theta})$. If $\boldsymbol{\beta} = \mathbf{0}$, then the iid $Y_i \sim D(\tau(\alpha), \boldsymbol{\theta})$. Hence testing $H_0 : \boldsymbol{\beta} = \mathbf{0}$ vs. $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$ is equivalent to testing whether the Y_i are a random sample from the $D(\tau(\alpha), \boldsymbol{\theta})$ distribution. Such a test can be done with the Kolmogorov-Smirnov test, the chi-square test, the Anderson-Darling test, the Cramér-von Mises test, et cetera. For specific distributions, there are often tests. For example, the Lilliefors test can be used to test if the Y_i are iid from a $N(\mu, \sigma^2)$ distribution where μ and σ^2 are unknown. See, for example, Kellison and London (2011, pp. 455-465), Conover (1971, pp. 295-308), Zheng, Lai, and Gould (2023), and Zheng et al. (2025).

This test has great level and extreme dimension reduction since the test does not depend on the predictors $\boldsymbol{x}$. The power can be sometimes be very poor if the cases are iid. a) If the $(Y_i, \boldsymbol{x}_i^T)^T$ are iid from a multivariate normal distribution, then the Y_i are iid $N(\mu_Y, \sigma_Y^2)$ regardless of whether $\boldsymbol{\beta} = \mathbf{0}$ or $\boldsymbol{\beta} \neq \mathbf{0}$ for the multiple linear regression model $Y|(\alpha + \boldsymbol{x}^T\boldsymbol{\beta}) \sim N(\alpha + \boldsymbol{x}^T\boldsymbol{\beta}, \sigma^2)$. b) If the $(Y_i, \boldsymbol{x}_i^T)^T$ are iid from some distribution where the $Y_i \in \{0, 1\}$ are binary, then the Y_i are iid $bin(n = 1, \rho_Y)$ regardless of whether $\boldsymbol{\beta} = \mathbf{0}$ or $\boldsymbol{\beta} \neq \mathbf{0}$ for the binary regression model $Y|(\alpha + \boldsymbol{x}^T\boldsymbol{\beta}) \sim bin(n = 1, \rho(\alpha + \boldsymbol{x}^T\boldsymbol{\beta}))$.

The test does not depend on $\boldsymbol{x}$, and can thus be done after variable selection. Also, all of the predictors can have outliers and missing values.

A Test for Binary Regression or Classification

Olive (2017, pp. 396-397) gave the result for a binary response variable $Y \in \{0, 1\}$.

Theorem 6. Let $\pi_j = P(Y = j)$ for $j = 0, 1$. Let $\boldsymbol{\mu}_j = E(\boldsymbol{x}|Y = j)$ for $j = 0, 1$. Then a) $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \hat{\pi}_1\hat{\pi}_0(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_0)$, and b) $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \pi_1\pi_0(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$.

Proof. Let N_i be the number of Ys that are equal to i for $i = 0, 1$ with $n = N_1 + N_2$. Then

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{N_i} \sum_{j:Y_j=i} \boldsymbol{x}_j$$

for $i = 0, 1$ while $\hat{\pi}_i = N_i/n$ and $\hat{\pi}_1 = 1 - \hat{\pi}_0$. Hence $\hat{\boldsymbol{\mu}}_i = \overline{\boldsymbol{x}}_i$ is the sample mean of the $\boldsymbol{x}_k$ corresponding to $Y_k = j$ for $j = 0, 1$. Then

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

$$\text{Thus } \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}} = \frac{1}{\mathrm{n}} \left[\sum_{\mathrm{j:Y_j=1}} \boldsymbol{x}_{\mathrm{j}}(1) + \sum_{\mathrm{j:Y_j=0}} \boldsymbol{x}_{\mathrm{j}}(0) \right] - \overline{\boldsymbol{x}}\ \hat{\pi}_1 =$$

$$\frac{1}{n}(N_1\hat{\boldsymbol{\mu}}_1) - \frac{1}{n}(N_1\hat{\boldsymbol{\mu}}_1 + N_0\hat{\boldsymbol{\mu}}_0)\hat{\pi}_1 = \hat{\pi}_1\hat{\boldsymbol{\mu}}_1 - \hat{\pi}_1^2\hat{\boldsymbol{\mu}}_1 - \hat{\pi}_1\hat{\pi}_0\hat{\boldsymbol{\mu}}_0 =$$

$$\hat{\pi}_1(1 - \hat{\pi}_1)\hat{\boldsymbol{\mu}}_1 - \hat{\pi}_1\hat{\pi}_0\hat{\boldsymbol{\mu}}_0 = \hat{\pi}_1\hat{\pi}_0(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_0).$$

Thus $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \pi_1\pi_0(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$. □

This result means $\boldsymbol{\eta} = \boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \pi_1\pi_0(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$ and $\boldsymbol{\phi} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$ are quantities of interest for binary regression. Note that $\boldsymbol{x} = (w_1, ..., w_k, w_1w_2, ..., w_1w_k, ..., w_{k-1}w_k)^T$ could be used to include pairwise interactions of the w_i.

Theorem 2b) suggests that typically the binary regression $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{C}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. If the cases $(Y_i, \boldsymbol{x}_i^T)^T$ are iid, then $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ can be tested with the omnibus test for $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. If the cases within each group are iid, if the two groups are independent, and if $N_1/(N_1+N_2) \to \pi_1$, then $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \pi_1\pi_0(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$ by Theorem 6b). Thus $H_0 : \boldsymbol{\beta} = \boldsymbol{0}$ can be tested with a high dimensional two sample test for $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_0$.

4.2 Testing $H_0 : \beta_i = 0$

4.3 Testing $H_0 : \boldsymbol{\beta}_I = (\beta_{i1}, ..., \beta_{ik})^T = \boldsymbol{0}$

High Dimensional Tests

Some tests when n/p is not large are simple. Testing $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{BR} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A}\boldsymbol{\beta}_{BR} \neq \boldsymbol{0}$ is equivalent to testing $H_0 : \boldsymbol{A}\boldsymbol{\eta} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A}\boldsymbol{\eta} \neq \boldsymbol{0}$ where $\boldsymbol{A}$ is a $k \times p$ constant matrix. Let $\mathrm{Cov}(\hat{\boldsymbol{\eta}}) = \boldsymbol{\Sigma}_{\boldsymbol{W}}$ be the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}}$. In high dimensions where $n < 5p$, we can't get a good nonsingular estimator of $\mathrm{Cov}(\hat{\boldsymbol{\eta}})$, but we can get good nonsingular estimators of $\mathrm{Cov}((\hat{\eta}_{i1}, ..., \hat{\eta}_{ik})^T)$ with $\boldsymbol{u} = (x_{i1}, ..., x_{ik})^T$ where $n \geq Jk$ with $J \geq 10$. (Values of J much larger than 10 may be needed if some of the k predictors are skewed or if a π_i in near 0 or 1.) Simply use the sample covariance matrix with $\boldsymbol{u}$ replacing $\boldsymbol{x}$. Hence we can test hypotheses like $H_0 : \beta_i - \beta_j = 0$. In particular, testing $H_0 : \beta_i = 0$ is equivalent to testing $H_0 : \eta_i = 0$.

Data splitting uses model selection (variable selection is a special case) to reduce the high dimensional problem to a low dimensional problem. The above procedure also reduces the high dimensional problem to a low dimensional problem.

5 CONCLUSIONS

Binary regression is closely related to two sample tests. Note that $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2$ can use other multivariate location estimators than sample means. For example, sample coordinatewise medians, sample coordinatewise trimmed means, and the Olive (2017b) T_{RMVN} estimator have large sample theory given by Rupasinghe Arachchige Don and Olive (2019) and Rupasinghe Arachchige Don and Pelawa Watagoda (2018).

Some papers on binary regression include Cai, Guo, and Ma (2023), Candès and Sur (2020), Mukherjee, Pillai, and Lin (2015), Sur and Candès (2019), Sur, Chen, and Candès (2019), and Tang and Ye (2020). Empirically, often $\boldsymbol{\beta}_{LR} \approx d\ \boldsymbol{\beta}_{OLS}$. Haggstrom (1983) suggests that d is not far from 1/MSE for logistic regression.

These binary regression estimators also give new ways to compare multivariate location estimators from two groups. The tests using k predictors can be performed. High dimensional tests for means from two groups can also be used. The tests that make very strong assumptions, such as multivariate normality or equal covariance matrices for the two groups, should be avoided. See Feng and Sun (2015), Gregory et al. (2015), Hu and Bai (2015), Rajapaksha and Olive (2024), and Xue and Yao (2020).

Software

The *R* software was used in the simulations. See R Core Team (2024). Programs will be added to the Olive (2025) collections of *R* functions *slpack.txt*, available from

(http://parker.ad.siu.edu/Olive/slpack.txt).

References

Abid, A.M., Quaye, P.A., and Olive, D.J. (2025), "A High Dimensional Omnibus Regression Test," *Stats*, 8, 107.

Artigue, H., and Smith, G. (2019), "The Principal Problem with Principal Components Regression," *Cogent Mathematics & Statistics*, 6, 1622190.

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.

Brown, P.J. (1993), *Measurement, Regression, and Calibration*, Oxford University Press, New York, NY.

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.

Conover, W.J. (1971), *Practical Nonparametric Statistics*, Wiley, New York, NY.

Cook, R.D. (2007), "Fisher Lecture: Dimension Reduction in Regression," *Statistical Science*, (with discussion), 22, 1-26.

Cook, R.D. (2018), "Principal Components, Sufficient Dimension Reduction, and Envelopes," *Annual Review of Statistics and Its Application*, 5, 533-559.

Cook, R.D., and Forzani, L. (2021), "PLS Regression Algorithms in the Presence of Nonlinearity," *Chemometrics and Intelligent Laboratory Systems*, 213, 104307.

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.

Cook, R.D., and Weisberg, S. (1999), *Applied Regression Including Computing and Graphics,* Wiley, New York, NY.

Chun, H., and Keleş, S. (2010), "Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Predictor Selection," *Journal of the Royal Statistical Society, B,* 72, 3-25.

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.

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

Kellison, S.G. and London, R.L. (2011), *Risk Models and Their Estimation*, ACTEX Publications, Winsted, CT.

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

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979), *Multivariate Analysis,* Academic Press, London, UK.

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

Olive, D.J. (2025), "Some Useful Techniques for High Dimensional Statistics," *Stats,* 8, 60.

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. https://doi.org/10.1080/03610926.2025.2527340

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.

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

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.

Zhang, J., and Chen, X. (2020), "Principal Envelope Model," *Journal of Statistical Planning and Inference*, 206, 249-262.

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.

Zheng, W., Lai, D., and Gould, K. L. (2023), "A Simulation Study of a Class of Nonparametric Test Statistics: a Close Look of Empirical Distribution Function-Based Tests," *Communications in Statistics - Simulation and Computation*, 52, 1132-1148.

Zheng, W., Zhu, H., Lance Gould, K., and Lai, D. (2025), "Comparing Heart PET Scans: an Adjustment of Kolmogorov-Smirnov Test under Spatial Autocorrelation," *Journal of Applied Statistics*, 52, 253-269.