Testing Multivariate Linear Regression with Univariate OPLS Estimators

David J. Olive*
Southern Illinois University

August 21, 2024

Abstract

A useful *multivariate linear regression model* is $\boldsymbol{y}_i = \boldsymbol{B}^T \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ for $i = 1, ..., n$. The model has $m \geq 2$ response variables $Y_1, ..., Y_m$ and p predictor variables $x_1, x_2, ..., x_p$. One technique is to fit the m univariate multiple linear regressions of Y_j on the predictors $\boldsymbol{x}$ to get $\hat{\boldsymbol{B}}_U = [\hat{\boldsymbol{\beta}}_1 \ \hat{\boldsymbol{\beta}}_2 \ \cdots \ \hat{\boldsymbol{\beta}}_m]$. Testing is considered for the estimators $\hat{\boldsymbol{B}}_U$ that use the the one component partial least squares estimators and marginal maximum likelihood estimators, including some high dimensional tests.

KEY WORDS: Lasso, MMLE, OLS, OPLS, Ridge Regression.

1 INTRODUCTION

A useful *multivariate linear regression model* is $\boldsymbol{y}_i = \boldsymbol{B}^T \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ for $i = 1, ..., n$. The model has $m \geq 2$ response variables $Y_1, ..., Y_m$ and p predictor variables $x_1, x_2, ..., x_p$. The $\boldsymbol{\epsilon}_i$ are assumed to be independent and identically distributed (iid). The ith case is $(\boldsymbol{x}_i^T, \boldsymbol{y}_i^T) = (x_{i1}, x_{i2}, ..., x_{ip}, Y_{i1}, ..., Y_{im})$, where the constant $x_{i1} = 1$. The model is written in matrix form as $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}$ where the matrices are defined below. The model has $E(\boldsymbol{\epsilon}_k) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_k) = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = (\sigma_{ij})$ for $k = 1, ..., n$. Also $E(\boldsymbol{e}_i) = \boldsymbol{0}$ while $\text{Cov}(\boldsymbol{e}_i, \boldsymbol{e}_j) = \sigma_{ij} \boldsymbol{I}_n$ for $i, j = 1, ..., m$ where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix and $\boldsymbol{e}_i$ is defined below. Then the $p \times m$ coefficient matrix $\boldsymbol{B} = [\boldsymbol{\beta}_1 \ \boldsymbol{\beta}_2 \ \cdots \ \boldsymbol{\beta}_m]$ and the $m \times m$ covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ are to be estimated, and $E(\boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{B}$ while $E(Y_{ij}) = \boldsymbol{x}_i^T \boldsymbol{\beta}_j$.

The $n \times m$ matrix of response variables and $n \times m$ matrix of errors are

$$\boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 & \ldots & \boldsymbol{Y}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{y}_1^T \\ \vdots \\ \boldsymbol{y}_n^T \end{bmatrix} \quad \text{and} \quad \boldsymbol{E} = \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_m \end{bmatrix} = \begin{bmatrix} \boldsymbol{\epsilon}_1^T \\ \vdots \\ \boldsymbol{\epsilon}_n^T \end{bmatrix},$$

while the $n \times p$ design matrix of predictor variables is $\boldsymbol{X}$.

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

Least squares is the classical method for fitting the multivariate linear model. The *least squares estimators* are $\hat{\boldsymbol{B}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Z} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 & \hat{\boldsymbol{\beta}}_2 & \dots & \hat{\boldsymbol{\beta}}_m \end{bmatrix}$. The matrix of *predicted values* or *fitted values* $\hat{\boldsymbol{Z}} = \boldsymbol{X}\hat{\boldsymbol{B}} = \begin{bmatrix} \hat{\boldsymbol{Y}}_1 & \hat{\boldsymbol{Y}}_2 & \dots & \hat{\boldsymbol{Y}}_m \end{bmatrix}$. The matrix of *residuals* $\hat{\boldsymbol{E}} = \boldsymbol{Z} - \hat{\boldsymbol{Z}} = \boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}} = \begin{bmatrix} \boldsymbol{r}_1 & \boldsymbol{r}_2 & \dots & \boldsymbol{r}_m \end{bmatrix}$. These quantities can be found from the m multiple linear regressions of Y_j on the predictors: $\widehat{\boldsymbol{\beta}}_j = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}_j$, $\widehat{\boldsymbol{Y}}_j = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j$ and $\boldsymbol{r}_j = \boldsymbol{Y}_j - \hat{\boldsymbol{Y}}_j$ for $j = 1, ..., m$. Hence $\hat{\epsilon}_{i,j} = Y_{i,j} - \hat{Y}_{i,j}$ where $\hat{\boldsymbol{Y}}_j = (\hat{Y}_{1,j}, ..., \hat{Y}_{n,j})^T$. Finally,

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \frac{(\boldsymbol{Z} - \hat{\boldsymbol{Z}})^T(\boldsymbol{Z} - \hat{\boldsymbol{Z}})}{n-p} = \frac{(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}})^T(\boldsymbol{Z} - \boldsymbol{X}\hat{\boldsymbol{B}})}{n-p} = \frac{\hat{\boldsymbol{E}}^T\hat{\boldsymbol{E}}}{n-p} = \frac{1}{n-p}\sum_{i=1}^n \hat{\boldsymbol{\epsilon}}_i\hat{\boldsymbol{\epsilon}}_i^T.$$

There are many other estimators. One technique is to fit the m univariate multiple linear regressions of Y_j on the predictors $\boldsymbol{x}$ to get $\hat{\boldsymbol{B}}_U = [\hat{\boldsymbol{\beta}}_1 \ \hat{\boldsymbol{\beta}}_2 \ \cdots \ \hat{\boldsymbol{\beta}}_m]$. A second technique is to let $W_{ij} = \hat{\boldsymbol{\eta}}_{ij}^T\boldsymbol{x}$ for $i = 1, ..., m$ and $j = 1, ..., k$. This results in mk predictors. Perform the OLS multivariate linear regression of $Y_1, ..., Y_m$ on $W_{11}, ..., W_{mk}$ to get $\hat{\boldsymbol{B}}_M = [\hat{\boldsymbol{\beta}}_1 \ \hat{\boldsymbol{\beta}}_2 \ \cdots \ \hat{\boldsymbol{\beta}}_m]$.

1.1 Multiple Linear Regression Estimators

One multiple linear regression model is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{1}$$

for $i = 1, ..., n$. Here n is the sample size and the random variable e_i is the ith error. 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 notation, these n equations become $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors.

Let the second multiple linear regression model be $Y|\boldsymbol{x}^T\boldsymbol{\beta} = \alpha + \boldsymbol{x}^T\boldsymbol{\beta} + e$ or $Y_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ or

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

for $i = 1, ..., n$. Let the e_i be as for model (1). In matrix form, this model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}, \tag{3}$$

$\boldsymbol{X}$ 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 estimation with ordinary least squares, let the covariance matrix of $\boldsymbol{x}$ be $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T] = E(\boldsymbol{x}\boldsymbol{x}^T) - E(\boldsymbol{x})E(\boldsymbol{x}^T)$ and $\boldsymbol{\eta} = \text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = E[(\boldsymbol{x} - E(\boldsymbol{x})(Y - E(Y))] = E(\boldsymbol{x}Y) - E(\boldsymbol{x})E(Y) = E[(\boldsymbol{x} - E(\boldsymbol{x}))Y] = E[\boldsymbol{x}(Y - E(Y))]$. 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 (3) 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 independent, identically distributed (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

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

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, Cook, Forzani, and Marcos (2024), Cook and Forzani (2024), and Wold (1975). Olive and Zhang (2024) 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 $\boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. The OPLS estimator is computed from the OLS simple linear regression of Y on $W = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T\boldsymbol{x}$, giving $\hat{Y} = \hat{\alpha}_{OPLS} + \hat{\lambda}W = \hat{\alpha}_{OPLS} + \hat{\boldsymbol{\beta}}_{OPLS}^T\boldsymbol{x}$.

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 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

$$\hat{\boldsymbol{\beta}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}. \tag{5}$$

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} = \boldsymbol{I}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{t}, Y) \tag{6}$$

where $(\boldsymbol{t}, Y)$ denotes that Y was regressed on $\boldsymbol{t}$, and $\boldsymbol{I}$ is the $p \times p$ identity matrix.

High dimensional regression has n/p small. A fitted or population regression model is sparse if a of the predictors are active (have nonzero $\hat{\beta}_i$ or β_i) where $n \geq Ja$ with $J \geq 10$. Otherwise the model is nonsparse. A high dimensional population regression model is abundant or dense if the regression information is spread out among the p predictors (nearly all of the predictors are active). Hence an abundant model is a nonsparse model.

The Tibshirani (1996) lasso estimator and Hoerl and Kennard (1970) ridge regression estimator are also interesting. The k component partial least squares (PLS) estimator $\hat{\boldsymbol{\beta}}_{kPLS}$ can be found from the OLS regression of Y on $W_1, ..., W_k$ where $W_i = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{i-1}\boldsymbol{x} = \hat{\boldsymbol{\eta}}_i^T\boldsymbol{x}$ with $\hat{\boldsymbol{\eta}}_i = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{i-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T$.

Olive and Zhang (2024) proved that there are often many valid population models for multiple linear regression, gave theory for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ and OPLS, and gave some theory for the MMLE for multiple linear regression under the constant variance assumption. Olive et al. (2024) gave more theory for the MMLE.

1.2 More Multivariate Estimators

Let the first multivariate linear regression model be

$$\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}. \tag{7}$$

Let the second multivariate linear regression model be

$$\boldsymbol{Z} = \boldsymbol{\alpha} + \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E} \tag{8}$$

where $\boldsymbol{\alpha}$ is the $n \times m$ matrix of constants with ith row equal to the vector of constants $(\alpha_1, ..., \alpha_m)^T$. Use the second model for the following three estimators.

Then

$$\hat{\boldsymbol{B}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},\boldsymbol{y}}.$$

The MMLE tends to estimate $p\boldsymbol{Y}_i$ for $i = 1, ..., m$.

The multivariate estimator obtained from the univariate OPLS regressions is

$$\hat{\boldsymbol{B}}_{UOPLS} = [\hat{\boldsymbol{\beta}}_1 \ \hat{\boldsymbol{\beta}}_2 \ \cdots \ \hat{\boldsymbol{\beta}}_m] = [\hat{\lambda}_1 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y_1}, \cdots, \hat{\lambda}_m \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y_m}]. \tag{9}$$

There are PLS estimators $\hat{\boldsymbol{B}}_{PLS}$ for multivariate linear regression. Obtain the mk predictors $W_{ij} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{i-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y_i}^T \boldsymbol{x}$ for $i = 1, ..., m$ and $j = 1, ..., k$. Then perform the OLS multivariate linear regression of $Y_1, ..., Y_m$ on the W_{ij} to get $\hat{\boldsymbol{B}}_{MkPLS} = [\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_m]$ which is not the PLS estimator.

2 Large Sample Theory

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

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

For random variables $x_1, ..., x_p$, 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 vectors such as $\tilde{\boldsymbol{c}} = [\tilde{\sigma}_{jk}]$, $\boldsymbol{c} = [\sigma_{jk}]$, $\boldsymbol{z}_i = [(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)]$, 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. (2024) theorem proves that sample covariance matrices are asymptotically normal. We use $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{d}}$ to avoid confusion with the $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ used in OPLS theory.

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

3 Testing

Consider model (8) with iid cases where $\boldsymbol{d}$ is a vector of distinct elements of $\boldsymbol{B} = \boldsymbol{B}_{MMLE}$ or $\boldsymbol{B} = \boldsymbol{B}_{UOPLS}$. Then $H_0 : \boldsymbol{d} = \boldsymbol{0}$ is true iff $H_0 : \boldsymbol{c} = \boldsymbol{0}$ is true where the c_{ij} are the covariances corresponding to the d_{ij}. As an illustration, some of important tests are whether a subset of rows of $\boldsymbol{B}$ are equal to $\boldsymbol{0}$. See Olive, Pelawa Watagoda, and Rupasinghe Arachchige Don (2015). Let the ith row of $\boldsymbol{B}$ in model (8) be $\boldsymbol{b}_i = (\beta_{i1}, ..., \beta_{im})$. Consider $\hat{\boldsymbol{B}}_{UOPLS}$ where $\lambda_i \neq 0$ for $i = 1, ..., m$. Testing $H_0 : \boldsymbol{b}_i^T = \boldsymbol{0}$ is equivalent to testing $H_0 : (\text{Cov}(x_i Y_1), ..., \text{Cov}(x_i Y_m))^T = \boldsymbol{\Sigma}_{x_i \boldsymbol{y}} = \boldsymbol{0}$. Under iid cases, this test is similar to testing $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$ with the $\boldsymbol{x}$ replaced by $\boldsymbol{y}$ and Y replaced by x_i. Tests for rows $i_1, i_2, ..., i_k$ use $H_0 : (\boldsymbol{b}_{i_1}, ..., \boldsymbol{b}_{i_K})^T = \boldsymbol{0}$, and are similar. If m and K are small, high dimensional tests can be done.

If m is small, rows of $\boldsymbol{B}_{UOPLS}$ can be tested. If p is small, columns of $\boldsymbol{B}_{UOPLS}$ can be tested. If both p and m are large, $\boldsymbol{B}_{ij}$ can be tested.

4 EXAMPLES AND SIMULATIONS

5 CONCLUSIONS

There are many multivariate linear regression estimators, including envelope estimators and partial least squares. See, for example, Cook (2018), Cook and Forzani (2024), Cook and Su (2013), Cook, Helland, and Su (2013), and Su and Cook (2012). Univariate methods like ridge regression and lasso can also be extended to multivariate linear regression. See, for example, Obozinski, Wainwright, and Jordan (2011).

6 References

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.

Cook, R.D. (2018), *An Introduction to Envelopes: Dimension Reduction for Efficient Estimation in Multivariate Statistics,* Wiley, Hoboken, NJ.

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 Su, Z. (2013), "Scaled Envelopes: Scale-Invariant and Efficient Estimation in Multivariate Linear Regression," *Biometrika*, 100, 929-954.

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.

Hoerl, A.E., and Kennard, R. (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12, 55-67.

Obozinski, G., Wainwright, M.J., and Jordan, M.I. (2011), "Support Union Recovery in High-Dimensional Multivariate Regression," *The Annals of Statistics*, 39, 1-47.

Olive, D.J., Alshammari, A.A., Pathiranage, K.G., and Hettige, L.A.W. (2024), "Testing with the One Component Partial Least Squares and the Marginal Maximum Likelihood Estimators," unpublished manuscript at (http://parker.ad.siu.edu/Olive/pphdwls.pdf).

Olive, D.J., Pelawa Watagoda, L.C.R., and Rupasinghe Arachchige Don, H.S. (2015), "Visualizing and Testing the Multivariate Linear Regression Model," *International Journal of Statistics and Probability*, 4, 126-137.

Olive, D.J., and Zhang, L. (2024), "One Component Partial Least Squares, High Dimensional Regression, Data Splitting, and the Multitude of Models," *Communications in Statistics: Theory and Methods*, to appear.

Su, Z., and Cook, R.D. (2012), "Inner Envelopes: Efficient Estimation in Multivariate Linear Regression," *Biometrika*, 99, 687-702.

Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, B*, 58, 267-288.

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