Summer 8-2024

Testing Multiple Linear Regression with the One Component Partial Least Squares Estimator

Kasun Halagoda Vidana Pathiranage
kasun.pathiranage@siu.edu

TESTING MULTIPLE LINEAR REGRESSION WITH THE ONE COMPONENT PARTIAL LEAST SQUARES ESTIMATOR

by

Kasun G. Pathiranage

B.S., University of Kelaniya, Sri Lanka 2021

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
August, 2024

RESEARCH PAPER APPROVAL

TESTING MULTIPLE LINEAR REGRESSION WITH THE ONE COMPONENT PARTIAL LEAST SQUARES ESTIMATOR

by

Kasun G. Pathiranage

A Research Paper Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Mathematics

Approved by:

David J. Olive

S. Yaser Samadi

Michael Sullivan

Graduate School
Southern Illinois University Carbondale
June 26, 2024

AN ABSTRACT OF THE RESEARCH PAPER OF

Kasun G. Pathiranage, for the Master of Science degree in MATHEMATICS, presented on June 26, 2024, at Southern Illinois University Carbondale.

TITLE: TESTING MULTIPLE LINEAR REGRESSION WITH THE ONE COMPONENT PARTIAL LEAST SQUARES ESTIMATOR

MAJOR PROFESSOR: Dr. David J. Olive

We consider testing the multiple linear regression model with the one component partial least squares (OPLS) estimator and the marginal maximum likelihood estimator (MMLE) where the sample covariance vector $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$, including the case where the predictors have been standardized to have unit variance. Some of the tests can be done in high dimensions.

KEY WORDS: Dimension reduction, high dimensional data, lasso, marginal maximum likelihood estimator.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my research advisor, Dr. David Olive for overseeing my Master's project and Dr. S. Yaser Samadi and Dr. Michael Sullivan for sitting on my committee. I would also like to thank the professors of Southern Illinois University for their instruction and care over the past two years. Finally, I want to say thank you to my wife and my family for all of your support and encouragement. You have pushed me to succeed and I could not have done it without you!

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This section reviews multiple linear regression models, including variable selection and data splitting, and follows Olive and Zhang (2024) and Olive, Alshammari, Pathiranage, and Hettige (2024) closely. Consider a multiple linear regression model with response variable Y and predictors $\boldsymbol{x} = (x_1, ..., x_p)$. Then there are n cases $(Y_i, \boldsymbol{x}_i^T)^T$, and the sufficient predictor $SP = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}$. For these regression models, the conditioning and subscripts, such as i, will often be suppressed. Ordinary least squares (OLS) is often used for the multiple linear regression (MLR) model.

Let the first multiple linear regression model be

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{1.1}$$

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

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{1.2}$$

for $i = 1, ..., n$. Let the e_i be as for model (1.1). In matrix form, this model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}, \tag{1.3}$$

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 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 (1.3) are $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, $\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}^T_{OLS}\overline{\boldsymbol{x}}$, and

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

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

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

$$\lambda = \frac{\boldsymbol{\Sigma}^T_{\boldsymbol{x}Y}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}^T_{\boldsymbol{x}Y}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}} \quad \text{and} \quad \hat{\lambda} = \frac{\hat{\boldsymbol{\Sigma}}^{\mathrm{T}}_{\boldsymbol{x}\mathrm{Y}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}}{\hat{\boldsymbol{\Sigma}}^{\mathrm{T}}_{\boldsymbol{x}\mathrm{Y}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}} \tag{1.4}$$

for $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq \mathbf{0}$. If $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$, then $\boldsymbol{\beta}_{OPLS} = \mathbf{0}$. Also see Basa, Cook, Forzani, and Marcos (2022) 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. The OPLS estimator is computed from the OLS simple linear regression (SLR) of Y on $W = \hat{\boldsymbol{\Sigma}}^T_{\boldsymbol{x}Y}\boldsymbol{x}$, giving $\hat{Y} = \hat{\alpha}_{OPLS} + \hat{\lambda}W = \hat{\alpha}_{OPLS} + \hat{\boldsymbol{\beta}}^T_{OPLS}\boldsymbol{x}$.

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 estimators, and $(\hat{\alpha}_{i,M}, \hat{\beta}_{i,M}) = (\hat{\alpha}_{i,SLR}, \hat{\beta}_{i,SLR})$. Hence

$$\hat{\boldsymbol{\beta}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}. \tag{1.5}$$

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} = \boldsymbol{I}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{t}, Y) \tag{1.6}$$

where $(\boldsymbol{t}, Y)$ denotes that Y was regressed on $\boldsymbol{t}$, and $\boldsymbol{I}$ is the $p \times p$ identity matrix. Olive, Alshammari, Pathiranage, and Hettige (2024) gave some large sample theory for the MMLE.

Sparse regression methods can be used for variable selection even if n/p is not large: the OLS submodel uses the predictors that had nonzero sparse regression estimated coefficients. These methods include least angle regression, lasso, relaxed lasso, elastic net, and sparse regression by projection. See Efron et al. (2004, p. 421), Meinshausen (2007, p. 376), Qi et al. (2015), Tay, Narasimhan, and Hastie (2023), Rathnayake and Olive (2023), Tibshirani (1996), and Zou and Hastie (2005).

Data splitting divides the training data set of n cases into two sets: H and the validation set V where H has n_H of the cases and V has the remaining $n_V = n - n_H$ cases $i_1, ..., i_{n_V}$. An application of data splitting is to use a variable selection method, such as forward selection or lasso, on H to get submodel I_{min} with a predictors, then fit the selected model to the cases in the validation set V using standard inference. See, for example, Rinaldo et al. (2019).

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.

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, gave theory for data splitting estimators, and gave some theory for the MMLE for multiple linear regression under the constant variance assumption.

Chapter 2 gives some large sample theory, while Chapter 3 considers tests of hypotheses.

CHAPTER 2

LARGE SAMPLE THEORY

Olive and Zhang (2024) derived the large sample theory for $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ and OPLS, including some high dimensional tests for low dimensional quantities such as $H_O : \beta_i = 0$ 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. Data splitting uses model selection (variable selection is a special case) to reduce the high dimensional problem to a low dimensional problem.

The following Olive and Zhang (2024) theorem gives the large sample theory for $\hat{\boldsymbol{\eta}} = \widehat{\text{Cov}}(\boldsymbol{x}, Y)$. 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 $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T$ and let $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ be defined below where

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

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

Theorem 1. Assume the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid. Assume $E(x_{ij}^k Y_i^m)$ exist for $j = 1, ..., p$ and $k, m = 0, 1, 2$. Let $\boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x})$ and $\mu_Y = E(Y)$. Let $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \mu_Y)$ with sample mean $\overline{\boldsymbol{w}}_n$. Let $\boldsymbol{\eta} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Then a)

$$\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}), \ \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}), \tag{2.1}$$

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

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

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 $\hat{\lambda} \xrightarrow{P} \lambda \neq 0$. Then

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) \xrightarrow{D} N_k(\boldsymbol{0}, \lambda^2 \boldsymbol{A} \boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{A}^T). \tag{2.2}$$

We will give a sketch of the proofs of a) and c). Also see Olive, Alshammari, Pathiranage, and Hettige (2024). For a), note that $\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$ by the multivariate central limit theorem since the $\boldsymbol{w}_i$ are iid with $E(\boldsymbol{w}_i) = \boldsymbol{\eta} = \text{Cov}(\boldsymbol{x}, Y)$ and $\text{Cov}(\boldsymbol{w}) = \boldsymbol{\Sigma}_{\boldsymbol{w}}$. Then it can be shown that $n\tilde{\boldsymbol{\eta}}_n =$ $\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}} + \boldsymbol{\mu}_{\boldsymbol{x}} - \overline{\boldsymbol{x}})(Y_i - \mu_Y + \mu_Y - \overline{Y}) = \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \mu_Y) = \sum_i \boldsymbol{w}_i - n\boldsymbol{a}_n =$ $\sum_i \boldsymbol{w}_i - n(\boldsymbol{\mu}_{\boldsymbol{x}} - \overline{\boldsymbol{x}})(\mu_Y - \overline{Y})$.

$$\text{Hence } \sqrt{\text{n}}(\tilde{\boldsymbol{\eta}}_\text{n} - \boldsymbol{\eta}) = \sqrt{\text{n}}(\overline{\boldsymbol{w}}_\text{n} - \boldsymbol{\eta}) + \text{o}_\text{P}(1).$$

$$\text{Thus } \sqrt{\text{n}}(\tilde{\boldsymbol{\eta}}_\text{n} - \boldsymbol{\eta}) \xrightarrow{\text{D}} \text{N}_\text{p}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$$

by Slutsky's theorem. c) If H_0 is true, then $\boldsymbol{A\eta} = \boldsymbol{0}$, and

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) = \sqrt{n}\boldsymbol{A}(\hat{\lambda}\hat{\boldsymbol{\eta}} - \hat{\lambda}\boldsymbol{\eta} + \hat{\lambda}\boldsymbol{\eta} - \boldsymbol{\beta}_{OPLS}) =$$

$$\hat{\lambda}\boldsymbol{A}\sqrt{n}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) + \boldsymbol{A}\sqrt{n}(\hat{\lambda} - \lambda)\boldsymbol{\eta} = \boldsymbol{Z}_n + \boldsymbol{b}_n \xrightarrow{D} N_k(\boldsymbol{0}, \lambda^2 \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{A}^T)$$

since $\boldsymbol{b}_n = \boldsymbol{0}$ when H_0 is true.

For iid cases, $\boldsymbol{\beta}_{MMLE} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_{OLS}$ where $\boldsymbol{V} = diag(\sigma_1^2, ..., \sigma_p^2) = diag(\boldsymbol{\Sigma}_{\boldsymbol{x}})$. For standardized predictors, let s_j and σ_j be the sample and population standard deviations of x_j. Let $\boldsymbol{t}_i = \hat{\boldsymbol{D}}\boldsymbol{x}_i = diag(1/s_1, ..., 1/s_p)\boldsymbol{x}_i$ and $\boldsymbol{u}_i = \boldsymbol{D}\boldsymbol{x}_i = diag(1/\sigma_1, ..., 1/\sigma_p)\boldsymbol{x}_i$. Note that $\hat{\boldsymbol{V}}^{-1} = \hat{\boldsymbol{D}}^2$ and $\boldsymbol{V}^{-1} = \boldsymbol{D}^2$. Olive and Zhang (2024) proved that $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y}$ is a $\sqrt{n}$ consistent estimator of $\boldsymbol{\Sigma}_{\boldsymbol{u}Y}$. For iid cases, $\boldsymbol{\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{t}Y} = \boldsymbol{\eta}_{OPLS}(\boldsymbol{t}, Y)$.

Olive, Alshammari, Pathiranage, and Hettige (2024) show that

$$\sqrt{n}\left[\begin{pmatrix} s_1^2 \\ \vdots \\ s_p^2 \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \end{pmatrix} - \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_p^2 \\ \boldsymbol{\Sigma}_{\boldsymbol{x}Y} \end{pmatrix}\right] = \sqrt{n}(\hat{\boldsymbol{c}} - \boldsymbol{c}) \xrightarrow{D} N_{2p}\left(\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{v}} & \boldsymbol{\Sigma}_{\boldsymbol{v},\boldsymbol{w}} \\ \boldsymbol{\Sigma}_{\boldsymbol{w},\boldsymbol{v}} & \boldsymbol{\Sigma}_{\boldsymbol{w}} \end{pmatrix}\right). \tag{2.3}$$

Let

$$\boldsymbol{g}(\boldsymbol{c}) = \boldsymbol{\beta}_{MMLE} = \begin{pmatrix} g_1(\boldsymbol{c}) \\ \vdots \\ g_p(\boldsymbol{c}) \end{pmatrix} = \begin{pmatrix} \sigma_{1Y}/\sigma_1^2 \\ \vdots \\ \sigma_{pY}/\sigma_p^2 \end{pmatrix}.$$

Let $\boldsymbol{D_g} = (\boldsymbol{D}_1, \boldsymbol{D}_2)$ where $\boldsymbol{D}_1 = diag(-\sigma_{1Y}/\sigma_1^4, -\sigma_{2Y}/\sigma_2^4, ..., -\sigma_{pY}/\sigma_p^4)$ and $\boldsymbol{D}_2 = \boldsymbol{D}^2 = diag(1/\sigma_1^2, 1/\sigma_2^2, ..., 1/\sigma_p^2)$. Typically $\hat{\boldsymbol{\Sigma}}_{x_{i_j}Y} = O_P(1)$, but if $\boldsymbol{\Sigma}_{x_{i_j}Y} = 0$, then $\hat{\boldsymbol{\Sigma}}_{x_{i_j}Y} = O_P(n^{-1/2})$.

Theorem 2. Let the cases $(\boldsymbol{x}_i^T, Y_i)^T$ be iid such that Equation (2.3) holds. Then a)

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{MMLE} - \boldsymbol{\beta}_{MMLE}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{MMLE}) \sim N_p\left(\boldsymbol{0}, \boldsymbol{D_g} \begin{pmatrix} \boldsymbol{\Sigma_v} & \boldsymbol{\Sigma_{v,w}} \\ \boldsymbol{\Sigma_{w,v}} & \boldsymbol{\Sigma_w} \end{pmatrix} \boldsymbol{D_g}^T \right).$$

Let $\boldsymbol{A}$ be a full rank $k \times p$ constant matrix such that $\boldsymbol{A\beta} = (\beta_{i_1}, ..., \beta_{i_k})^T$ with $i_1, i_2, ..., i_k$ distinct. Hence the jth row of $\boldsymbol{A}$ has a 1 in the i_jth position and zeroes elsewhere. Assume $H_0 : \boldsymbol{A\beta}_{MMLE} = \boldsymbol{0}$. Then b)

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{MMLE} - \boldsymbol{\beta}_{MMLE}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{AD}^2\boldsymbol{\Sigma_w}\boldsymbol{D}^2\boldsymbol{A}^T).$$

c) For standardized predictors, assume $H_0 : \boldsymbol{A\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{A\Sigma_{t}}_Y = \boldsymbol{0}$. Then

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) - \boldsymbol{\beta}_{MMLE}(\boldsymbol{t}, Y)) = \sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \boldsymbol{\Sigma}_{\boldsymbol{u}Y}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{AD\Sigma_w DA}^T).$$

Proof. Theorem 2a) holds by the multivariate delta method.

b) Note that $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{MMLE} - \boldsymbol{\beta}_{MMLE}) = \sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{D}}^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{D}^2\boldsymbol{\Sigma}_{\boldsymbol{x}Y}) =$
$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{D}}^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{D}^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} + \boldsymbol{D}^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{D}^2\boldsymbol{\Sigma}_{\boldsymbol{x}Y}) =$

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{D}}^2 - \boldsymbol{D}^2)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} + \sqrt{n}\boldsymbol{AD}^2(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{\Sigma}_{\boldsymbol{x}Y})$$

where by Theorem 1,

$$\sqrt{n}\boldsymbol{AD}^2(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{\Sigma}_{\boldsymbol{x}Y}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{AD}^2\boldsymbol{\Sigma_w}\boldsymbol{D}^2\boldsymbol{A}^T).$$

Now $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{D}}^2 - \boldsymbol{D}^2)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} =$

$$\boldsymbol{A}\begin{pmatrix} \sqrt{n}\left(\frac{1}{s_1^2} - \frac{1}{\sigma_1^2}\right)\hat{\boldsymbol{\Sigma}}_{x_1 Y} \\ \vdots \\ \sqrt{n}\left(\frac{1}{s_p^2} - \frac{1}{\sigma_p^2}\right)\hat{\boldsymbol{\Sigma}}_{x_p Y} \end{pmatrix} = \begin{pmatrix} \sqrt{n}\left(\frac{1}{s_{i_1}^2} - \frac{1}{\sigma_{i_1}^2}\right)\hat{\boldsymbol{\Sigma}}_{x_{i_1} Y} \\ \vdots \\ \sqrt{n}\left(\frac{1}{s_{i_k}^2} - \frac{1}{\sigma_{i_k}^2}\right)\hat{\boldsymbol{\Sigma}}_{x_{i_k} Y} \end{pmatrix} = o_P(1)$$

if $(\Sigma_{x_{i_1}Y}, ..., \Sigma_{x_{i_k}Y})^T = \boldsymbol{0}$. Hence the result follows if H_0 is true.

c) Note that $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \boldsymbol{\Sigma}_{\boldsymbol{u}Y}) = \sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y} + \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y} - \boldsymbol{\Sigma}_{\boldsymbol{u}Y}) =$ $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}) + \sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y} - \boldsymbol{\Sigma}_{\boldsymbol{u}Y})$ where by Theorem 1,

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y} - \boldsymbol{\Sigma}_{\boldsymbol{u}Y}) = \sqrt{n}\boldsymbol{A}\boldsymbol{D}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{\Sigma}_{\boldsymbol{x}Y}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{D}\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{D}\boldsymbol{A}^T).$$

Now $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}) = \sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{D}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} - \boldsymbol{D}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}) = \sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{D}} - \boldsymbol{D})\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} =$

$$\boldsymbol{A}\begin{pmatrix} \sqrt{n}\left(\frac{1}{s_1} - \frac{1}{\sigma_1}\right)\hat{\boldsymbol{\Sigma}}_{x_1 Y} \\ \vdots \\ \sqrt{n}\left(\frac{1}{s_p} - \frac{1}{\sigma_p}\right)\hat{\boldsymbol{\Sigma}}_{x_p Y} \end{pmatrix} = \begin{pmatrix} \sqrt{n}\left(\frac{1}{s_{i_1}} - \frac{1}{\sigma_{i_1}}\right)\hat{\boldsymbol{\Sigma}}_{x_{i_1} Y} \\ \vdots \\ \sqrt{n}\left(\frac{1}{s_{i_k}} - \frac{1}{\sigma_{i_k}}\right)\hat{\boldsymbol{\Sigma}}_{x_{i_k} Y} \end{pmatrix},$$

and $\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}) = o_p(1)$ if $(\Sigma_{x_{i_1}Y}, ..., \Sigma_{x_{i_k}Y})^T = \boldsymbol{0}$. Hence if H_0 is true, then

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} - \boldsymbol{\Sigma}_{\boldsymbol{u}Y}) \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{D}\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{D}\boldsymbol{A}^T). \quad \square$$

It can be shown that if $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}} = (c_{ij})$, then $\hat{\boldsymbol{D}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}\hat{\boldsymbol{D}} = (b_{ij})$ where $b_{ij} = c_{ij}/(s_i s_j)$.

Olive, Alshammari, Pathiranage, and Hettige (2024) considered testing using Theorem 1a), estimating $\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{w}}\boldsymbol{A}^T$ with $\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}\boldsymbol{A}^T$.

The following simple testing method reduces a possibly high dimensional problem to a low dimensional problem. 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. Let $\text{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}) = \text{Cov}(\hat{\boldsymbol{\eta}}) = \boldsymbol{\Sigma}_{\boldsymbol{w}}$ be the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. In high dimensions where $n < 5p$, we can't get a good nonsingular estimator of $\text{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y})$, but we can get good nonsingular estimators of $\text{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}) = \text{Cov}((\hat{\eta}_{i1}, ..., \hat{\eta}_{ik})^T)$ with $\boldsymbol{u} = \boldsymbol{x}_I = (x_{i1}, ..., x_{ik})^T$ where $n \geq Jk$ with $J \geq 10$. (Values of J much larger than 10 may be needed if some of the k predictors and/or Y are skewed.) Simply apply Theorem 1 to the predictors $\boldsymbol{u}$ used in the hypothesis test, and thus use the sample covariance matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}_I}$ of the vectors $\boldsymbol{u}_i(Y_i - \overline{Y})$. 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 $H_0 : \eta_i = \sigma_{x_i,Y} = 0$ where $\sigma_{x_i,Y} = \text{Cov}(x_i, Y)$.

The tests with $\hat{\boldsymbol{\beta}}_{OPLS} = \hat{\lambda}\hat{\boldsymbol{\eta}}$ and k predictor variables may not be as good as the tests with $\hat{\boldsymbol{\eta}}$ since $\hat{\lambda}$ needs to be a good estimator of λ. Note that $\hat{\lambda}$ can be a good estimator if $\hat{\boldsymbol{\eta}}^T\boldsymbol{x}$ is a good estimator of $\boldsymbol{\eta}^T\boldsymbol{x}$.

Note that the tests with $\hat{\boldsymbol{\eta}}$ using k predictors x_{ij} do not depend on other predictors, including important predictors that were left out of the model (underfitting). Hence the tests can have considerable resistance to underfitting and overfitting. The tests also have some resistance to measurement error: assume that $(\boldsymbol{x}_i^T, \boldsymbol{u}_i^T, v_i, Y_i)^T$ are iid but $\boldsymbol{w}_i = \boldsymbol{x}_i + \boldsymbol{u}_i$ and $Z_i = Y_i + v_i$ are observed instead of $(\boldsymbol{x}_i, Y_i)$. Then $\hat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{w}, Z)$ estimates $\boldsymbol{\Sigma}_{\boldsymbol{w}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{w}Z}$, while $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}Z}$ estimates $\text{Cov}(\boldsymbol{x}, Y)$ if $\text{Cov}(\boldsymbol{x}, v) + \text{Cov}(\boldsymbol{u}, Y) + \text{Cov}(\boldsymbol{u}, v) = \boldsymbol{0}$, which occurs, for example, if $\boldsymbol{x} \perp\!\!\!\perp v$, $\boldsymbol{u} \perp\!\!\!\perp Y$, and $\boldsymbol{u} \perp\!\!\!\perp v$.

CHAPTER 3

REGRESSION WITH HETEROGENEITY

A multiple linear regression model with heterogeneity is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i \tag{3.1}$$

for $i = 1, ..., n$ where the e_i are independent with $E(e_i) = 0$ and $V(e_i) = \sigma_i^2$. In matrix form, this model is

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

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. Also $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma_e} = diag(\sigma_i^2) = diag(\sigma_1^2, ..., \sigma_n^2)$ is an $n \times n$ positive definite matrix. In Chapter 2, the constant variance assumption was used: $\sigma_i^2 = \sigma^2$ for all i. Hence heterogeneity means that the constant variance assumption does not hold. A common assumption is that the $e_i = \sigma_i \epsilon_i$ where the ϵ_i are independent and identically distributed (iid) with $V(\epsilon_i) = 1$. See, for example, Zhou, Cook, and Zou (2023).

Weighted least squares (WLS) would be useful if the σ_i^2 were known. Since the σ_i^2 are not known, ordinary least squares (OLS) is often used. The OLS theory for MLR with heterogeneity often assume iid cases.

CHAPTER 4

EXAMPLE AND SIMULATIONS

Example. The Hebbler (1847) data was collected from $n = 26$ districts in Prussia in 1843. Let $Y =$ the *number of women married to civilians* in the district with a constant and predictors $x_1 =$ the *population of the district in 1843*, $x_2 =$ the *number of married civilian men* in the district, $x_3 =$ the *number of married men in the military* in the district, and $x_4 =$ the *number of women married to husbands in the military* in the district. Sometimes the person conducting the survey would not count a spouse if the spouse was not at home. Hence Y and x_2 are highly correlated but not equal. Similarly, x_3 and x_4 are highly correlated but not equal. Then $\hat{\boldsymbol{\beta}}_{OLS} = (0.00035, 0.9995, -0.2328, 0.1531)^T$, forward selection with OLS and the C_p criterion used $\hat{\boldsymbol{\beta}}_{I,0} = (0, 1.0010, 0, 0)^T$, lasso had $\hat{\boldsymbol{\beta}}_L = (0.0015, 0.9605, 0, 0)^T$, lasso variable selection $\hat{\boldsymbol{\beta}}_{LVS} = (0.00007, 1.006, 0, 0)^T$, $\hat{\boldsymbol{\beta}}_{MMLE} = (0.1782, 1.0010, 48.5630, 51.5513)^T$, and $\hat{\boldsymbol{\beta}}_{OPLS} = (0.1727, 0.0311, 0.00018, 0.00018)^T$. With scaled predictors, $\hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t}Y} = (40678.97, 40937.98, 21877.44, 22308.46)^T$. The fitted values from the MMLE estimator tend not to estimate Y. Let $W = \boldsymbol{x}^T \hat{\boldsymbol{\beta}}_{MMLE}$ and perform the simple linear regression of Y on W to get the reweighted or scaled estimators $\hat{\alpha}_R$ and b. Then $\hat{\boldsymbol{\beta}}_R = b\hat{\boldsymbol{\beta}}_{MMLE}$. Then the fitted values $\hat{Y}_i = \hat{\alpha}_R + \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}_R$ can be used for prediction. If the scaled predictors $\boldsymbol{u}$ have unit sample variances, then $\hat{\boldsymbol{\beta}}_{OPLS}(\boldsymbol{u}, Y) = \hat{\boldsymbol{\beta}}_R(\boldsymbol{u}, Y)$.

Next, we describe a small WLS simulation study somewhat similar to that done by Rajapaksha and Olive (2024). The simulation used $\psi = 0$ and $1/\sqrt{p}$; and $k = 1$ and $p - 1$ where k and ψ are defined in the following paragraph.

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

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

When $\psi = 0$ and wtype = 1, the OLS confidence intervals for β_i should have length near $2t_{96,0.975}\sigma/\sqrt{n} \approx 2(1.96)\sigma/10 = 0.392\sigma$ when $n = 100$ and the iid zero mean errors have variance σ^2.

The simulation computed $\boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = (\eta_1, ..., \eta_{p-1})^T = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_{OLS}$ where $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{A}^T$ is a $(p-1) \times (p-1)$ matrix. Storage problems can occur if $p > 10000$. Then the Theorem 1 large sample $100(1-\delta)\%$ CI is $\hat{\eta}_i \pm t_{n-1,1-\delta/2}SE(\hat{\eta}_i)$ could be computed for each η_i. If 0 is not in the confidence interval, then $H_0 : \eta_i = 0$ and $H_0 : \beta_{iE} = 0$ are both rejected for estimators E = OPLS and MMLE. In the simulations with $n = 50$, $p = 4$, and $\psi > 0$, the maximum observed undercoverage was about $0.05 = 5\%$. Hence the program has the option to replace the cutoff $t_{n-1,1-\delta/2}$ by $t_{n-1,up}$ where $up = min(1-\delta/2 + 0.05, 1-\delta/2 + 2.5/n)$ if $\delta/2 > 0.1$,

$$up = min(1-\delta/4, 1-\delta/2 + 12.5\delta/n)$$

if $\delta/2 \leq 0.1$. If $up < 1-\delta/2 + 0.001$, then use $up = 1-\delta/2$. This correction factor was used in the simulations for the nominal 95% CIs, where the correction factor uses a cutoff that is between $t_{n-1,0.975}$ and the cutoff $t_{n-1,0.9875}$ that would be used for a 97.5% CI. The nominal coverage was 0.95 with $\delta = 0.05$. Observed coverage between 0.94 and 0.96 suggests coverage is close to the nominal value. Pötscher and Preinerstorfer (2023) noted that WLS tests tend to reject H_0 too often (liberal tests with undercoverage).

The simulation computed $p-1$ confidence intervals $[L_{in}, U_{in}]$ for $\eta_i = Cov(x_i, Y) = \sigma_{iY}$ for $i = 1, ..., p-1 = 99$. Let $\sigma_i^2 = Var(x_i)$, the variance of the ith predictor x_i. Let $\boldsymbol{\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{\eta}(\boldsymbol{t})$ and $\beta_{i,MMLE(\boldsymbol{t},Y)} = Cov(t_i, Y) = \eta_i(\boldsymbol{t})$. Then the program checked whether $\beta_{i,MMLE(\boldsymbol{t},Y)} = Cov(t_i, Y) = \eta_i(\boldsymbol{t}) = \sigma_{iY}/\sigma_i$ was in the interval $(1/s_i)[L_{in}, U_{in}]$. 5000 intervals were generated for each $\eta_i(\boldsymbol{t})$, and the coverage was the proportion of times $\eta_i(\boldsymbol{t})$ was in its interval. Hence if $\eta_1(\boldsymbol{t})$ was in its interval $4750/5000 = 0.95$, then the observed coverage was 0.95. This procedure corresponds to a large sample test for $H_0 : \eta_i(\boldsymbol{t}) = 0$ only if $\eta_i(\boldsymbol{t}) = 0$. This occurred when $\psi = 0$ for $i = 2, ..., p-1 = 99$, but not for $i = 1$ or $\psi = 0.1$. The correction factor was used.

To summarize the $p-1$ intervals, the average length of the $p-1$ intervals over 5000 runs was computed. Then the minimum, mean, and maximum of the average lengths was computed. The proportion of times each interval contained its population parameter was computed. These proportions were the observed coverages of the $p-1$ intervals. Then the minimum observed coverage was found. The percentage of the observed coverages that were $\geq$ 0.9, 0.92, 0.93, 0.94, and 0.96 were also recorded. The coverage of the test $H_0 : \boldsymbol{\beta}_{I,MMLE(\boldsymbol{t},Y)} = \boldsymbol{\eta}_I(\boldsymbol{t}) = \boldsymbol{0}$ was recorded and a correction factor was not used. Here $I = \{98, 99\}$.

Suppose $\boldsymbol{A}\boldsymbol{\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{A}\boldsymbol{\eta}(\boldsymbol{t}) = (\eta_{i_1}(\boldsymbol{t}), ..., \eta_{i_k}(\boldsymbol{t}))^T = \boldsymbol{\beta}_{I,MMLE}(\boldsymbol{t}, Y) = \boldsymbol{\eta}_I(\boldsymbol{t})$ where $I = \{i_1, ..., i_k\}$. Let $\hat{D}_I = diag(1/s_{i_1}, ..., 1/s_{i_k})$. Let $\boldsymbol{u} = \boldsymbol{x}_I = (x_{i1}, ..., x_{ik})^T$. The test statistic for the test $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{\beta}_{I,MMLE}(\boldsymbol{t}, Y) = \boldsymbol{A}\boldsymbol{\eta}(\boldsymbol{t}) = \boldsymbol{\eta}_I(\boldsymbol{t}) = \boldsymbol{0}$ is $T_n = n\hat{\boldsymbol{\eta}}(\boldsymbol{t})\boldsymbol{A}^T(\boldsymbol{A}\hat{\boldsymbol{D}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}}\hat{\boldsymbol{D}}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\boldsymbol{\eta}}(\boldsymbol{t}) = n\hat{\boldsymbol{\eta}}_I^T(\boldsymbol{t})(\hat{\boldsymbol{D}}_I\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}_I}\hat{\boldsymbol{D}}_I)^{-1}\hat{\boldsymbol{\eta}}_I(\boldsymbol{t}) \xrightarrow{D} \chi^2_k$ when H_0 is true. The simulation used $I = \{98, 99\}$ and tested $H_0 : \boldsymbol{A}\boldsymbol{\eta}(\boldsymbol{t}) = (\eta_{98}(\boldsymbol{t}), \eta_{99}(\boldsymbol{t}))^T = \boldsymbol{0}$. In the simulation H_0 was true for $k = 1$ and $\psi = 0$, but false for either $\psi = 0.1$ or $k = 99$.

In the simulation if the model is linear, $\boldsymbol{\beta}_{OLS} = (1, 0, ..., 0)^T$ for $k = 1$, and $\boldsymbol{\beta}_{OLS} = \mathbf{1}$ for $k = 99$. If $\psi = 0$ and the model is linear, then $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{I}_p$, $\lambda = 1$, and $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Then $\hat{\lambda}$ was often less than 0.5 for $n = 100$ and $p = 100$. If $\psi = 0.1$, $k = 99$, and the model is linear, then $\lambda = 1/116.64 = 0.008573$, $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OPLS} = \mathbf{1}$, and $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = 116.64\ \mathbf{1}$. Now $\hat{\lambda}$ tended to be close to λ. The models appeared to be linear except for wtype=4 with $\psi = 0.1$. (This model appeared to generate massive outliers with entries of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ often larger than 10^{50} for $n = 100$ and $p = 100$.)

```
source("http://parker.ad.siu.edu/Olive/slpack.txt")
args(mmlesim2)
function (n = 100, p = 4, k = 1, nruns = 100, eps = 0.1, shift = 9,
 etype = 1, wtype = 1, psi = 0, cfac = "T", indices = c(1,2), alph = 0.05)

mmlesim2(n=100,p=100,k=1,nruns=5000,etype=1,wtype=1,psi=0,indices = c(98,99))
$lens
[1] 0.5879743 0.5912557 0.7146509
$covprop
[1] 0.9494000 1.0000000 1.0000000 1.0000000 1.0000000 0.7676768
$testcov
[1] 0.9416



#change etype and psi to get the rest of Table 4.1.
#then repeat to get Tables 4.2-4.7 corresponding to wtype =2,...,7
#do not use psi=0.1 for wtype=4

#then repeat with k=99 to get Tables 4.8-4.14
```

```
#so the first two line of table 4.8 use the following R command

mmlesim2(n=100,p=100,k=99,nruns=5000,etype=1,wtype=1,psi=0,indices = c(98,99))
```

Table 4.1. Cov(t,Y), wtype=1, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9494	1.0	1.0	1.0	1.0	0.7677	0.9416
	len	1	0.5880	0.5913	0.7147				
100	100	0.1	0.6362	0.0	0.0	0.0	0.0	0.0	0.0006
	len	1	0.7908	0.7938	0.9146				
100	100	0	0.9488	1.0	1.0	1.0	1.0	0.9596	0.9466
	len	2	0.7975	0.8042	0.8969				
100	100	0.1	0.7350	0.0	0.0	0.0	0.0	0.0	0.0284
	len	2	0.9592	0.9648	1.0710				
100	100	0	0.9514	1.0	1.0	1.0	1.0	0.8586	0.9456
	len	3	0.5853	0.5887	0.7118				
100	100	0.1	0.6366	0.0	0.0	0.0	0.0	0.0	0.0008
	len	3	0.7909	0.7943	0.9186				
100	100	0	0.9472	1.0	1.0	1.0	1.0	0.8081	0.9464
	len	4	0.4793	0.4823	0.6277				
100	100	0.1	0.5740	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	0.7141	0.7175	0.8511				
100	100	0	0.9662	1.0	1.0	1.0	1.0	1.0	0.9600
	len	5	1.3556	1.3676	1.4303				
100	100	0.1	0.8494	0.9899	0.8990	0.0202	0.0	0.0	0.2676
	len	5	1.4664	1.4773	1.5483				

Table 4.2. Cov(t,Y), wtype=2, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9524	1.0	1.0	1.0	1.0	0.8485	0.9498
	len	1	1.7597	1.7666	1.8901				
100	100	0.1	0.9206	1.0	1.0	0.9899	0.9899	0.0	0.5762
	len	1	1.9283	1.9363	2.0839				
100	100	0	0.9590	1.0	1.0	1.0	1.0	0.9697	0.9502
	len	2	2.8021	2.8259	2.9580				
100	100	0.1	0.9368	1.0	1.0	1.0	0.9899	0.1616	0.7614
	len	2	2.9824	3.0173	3.2102				
100	100	0	0.9584	1.0	1.0	1.0	1.0	0.9495	0.9432
	len	3	1.7316	1.7416	1.8727				
100	100	0.1	0.8536	0.9899	0.4646	0.0202	0.0	0.0	0.4856
	len	3	1.9105	1.9200	2.0719				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.6970	0.9476
	len	4	1.0758	1.0796	1.1946				
100	100	0.1	0.8618	0.9899	0.9697	0.0606	0.0	0.0	0.1674
	len	4	1.2474	1.2524	1.3807				
100	100	0	0.9698	1.0	1.0	1.0	1.0	1.0	0.9646
	len	5	5.2170	5.2842	5.4717				
100	100	0.1	0.9616	1.0	1.0	1.0	1.0	1.0	0.8688
	len	5	5.4372	5.5098	5.7000				

Table 4.3. Cov(t,Y), wtype=3, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9406	1.0	1.0	1.0	1.0	0.8889	0.9408
	len	1	0.7137	0.7204	0.9963				
100	100	0.1	0.8594	0.9899	0.9798	0.1010	0.0	0.0	0.0584
	len	1	1.0774	1.0842	1.3391				
100	100	0	0.9532	1.0	1.0	1.0	1.0	0.9798	0.9482
	len	2	1.0308	1.0435	1.4228				
100	100	0.1	0.9158	1.0	0.9899	0.9899	0.9899	0.0	0.3182
	len	2	1.5167	1.5346	1.9025				
100	100	0	0.9442	1.0	1.0	1.0	1.0	0.9495	0.9448
	len	3	0.7095	0.7163	0.9906				
100	100	0.1	0.8426	0.9899	0.1212	0.0	0.0	0.0	0.0602
	len	3	1.0669	1.0749	1.3253				
100	100	0	0.9372	1.0	1.0	1.0	0.9899	0.8384	0.9488
	len	4	0.5319	0.5369	0.7417				
100	100	0.1	0.7414	0.0	0.0	0.0	0.0	0.0	4e-04
	len	4	0.8297	0.8343	1.0203				
100	100	0	0.9676	1.0	1.0	1.0	1.0	1.0	0.9656
	len	5	1.8191	1.8481	2.4701				
100	100	0.1	0.9464	1.0	1.0	1.0	1.0	0.8182	0.5944
	len	5	2.5991	2.6376	3.2597				

Table 4.4. Cov(t,Y), wtype=4, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9424	1.0	1.0	1.0	1.0	0.8182	0.9396
	len	1	0.4152	0.4179	0.5797				
100	100	0	0.9468	1.0	1.0	1.0	1.0	0.7879	0.9430
	len	2	0.4151	0.4185	0.5808				
100	100	0	0.9452	1.0	1.0	1.0	1.0	0.7879	0.9412
	len	3	0.4141	0.4173	0.5800				
100	100	0	0.9456	1.0	1.0	1.0	1.0	0.7778	0.9460
	len	4	0.4153	0.4185	0.5806				
100	100	0	0.9430	1.0	1.0	1.0	1.0	0.7980	0.9476
	len	5	0.4156	0.4186	0.5812				

Table 4.5. Cov(t,Y), wtype=5, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.8081	0.9498
	len	1	25.6886	25.7753	25.8629				
100	100	0.1	0.9592	1.0	1.0	1.0	1.0	0.9899	0.9434
	len	1	22.4987	22.5958	22.6681				
100	100	0	0.9608	1.0	1.0	1.0	1.0	1.0	0.9508
	len	2	41.8712	42.4539	42.9896				
100	100	0.1	0.9624	1.0	1.0	1.0	1.0	1.0	0.9512
	len	2	35.7635	35.9848	36.2287				
100	100	0	0.9596	1.0	1.0	1.0	1.0	0.9899	0.9502
	len	3	25.2238	25.3569	25.4858				
100	100	0.1	0.962	1.0	1.0	1.0	1.0	1.0	0.9518
	len	3	21.7499	21.8569	21.9813				
100	100	0	0.9546	1.0	1.0	1.0	1.0	0.6970	0.9446
	len	4	14.9002	14.9325	14.9651				
100	100	0.1	0.9600	1.0	1.0	1.0	1.0	1.0	0.9386
	len	4	13.0557	13.1212	13.1664				
100	100	0	0.9676	1.0	1.0	1.0	1.0	1.0	0.9566
	len	5	78.8742	79.5146	80.5798				
100	100	0.1	0.9720	1.0	1.0	1.0	1.0	1.0	0.9696
	len	5	66.2657	67.0948	67.9840				

Table 4.6. Cov(t,Y), wtype=6, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9446	1.0	1.0	1.0	1.0	0.8384	0.9458
	len	1	0.4726	0.4760	0.6230				
100	100	0.1	0.7044	0.0	0.0	0.0	0.0	0.0	0.0062
	len	1	0.8804	0.8862	0.9959				
100	100	0	0.9430	1.0	1.0	1.0	0.8990	0.9438	
	len	2	0.5576	0.5622	0.6926				
100	100	0.1	0.8068	0.9899	0.8788	0.0202	0.0	0.0	0.1016
	len	2	1.1215	1.1300	1.2269				
100	100	0	0.9454	1.0	1.0	1.0	1.0	0.7677	0.9392
	len	3	0.4702	0.4737	0.6203				
100	100	0.1	0.7012	0.0	0.0	0.0	0.0	0.0	0.0110
	len	3	0.8780	0.8827	0.9929				
100	100	0	0.9436	1.0	1.0	1.0	1.0	0.7778	0.9432
	len	4	0.4352	0.4383	0.5967				
100	100	0.1	0.6026	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	0.7499	0.7538	0.8808				
100	100	0	0.9528	1.0	1.0	1.0	1.0	0.9899	0.9536
	len	5	0.8236	0.8301	0.9256				
100	100	0.1	0.8736	0.9899	0.9899	0.9899	0.9192	0.0	0.3582
	len	5	1.7889	1.8044	1.8797				

Table 4.7. Cov(t,Y), wtype=2, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9462	1.0	1.0	1.0	1.0	0.7677	0.9428
	len	1	0.4909	0.4942	0.6393				
100	100	0.1	0.5576	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	0.7089	0.7121	0.8479				
100	100	0	0.9476	1.0	1.0	1.0	1.0	0.8384	0.9466
	len	2	0.6011	0.6053	0.7280				
100	100	0.1	0.6092	0.0	0.0	0.0	0.0	0.0	0.0032
	len	2	0.7670	0.7719	0.8976				
100	100	0	0.9486	1.0	1.0	1.0	1.0	0.8485	0.9454
	len	3	0.4917	0.4948	0.6378				
100	100	0.1	0.5642	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	0.7087	0.7119	0.8486				
100	100	0	0.9412	1.0	1.0	1.0	1.0	0.7677	0.9438
	len	4	0.4424	0.4456	0.6008				
100	100	0.1	0.5308	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	0.6873	0.6905	0.8288				
100	100	0	0.9610	1.0	1.0	1.0	1.0	1.0	0.9520
	len	5	0.9246	0.9322	1.0256				
100	100	0.1	0.7130	0.0	0.0	0.0	0.0	0.0	0.0442
	len	5	0.9714	0.9776	1.0851				

Table 4.8. Cov(t,Y), wtype=1, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.7475	0.7794
	len	1	4.1765	4.191	4.2072				
100	100	0.1	0.6446	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	55.8424	56.0056	56.1395				
100	100	0	0.9572	1.0	1.0	1.0	1.0	0.6970	0.7808
	len	2	4.2080	4.2221	4.2372				
100	100	0.1	0.6544	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	55.8814	56.0213	56.2027				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.7273	0.7728
	len	3	4.1770	4.1874	4.2021				
100	100	0.1	0.6470	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	55.8012	55.9117	56.0715				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.6869	0.7762
	len	4	4.1566	4.1707	4.1815				
100	100	0.1	0.6656	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	55.6662	55.7832	55.9467				
100	100	0	0.9534	1.0	1.0	1.0	1.0	0.7071	0.7934
	len	5	4.3706	4.3883	4.4056				
100	100	0.1	0.6488	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	55.7942	55.9609	56.1033				

Table 4.9. Cov(t,Y), wtype=2, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.8586	0.8806
	len	1	6.0201	6.0561	6.0852				
100	100	0.1	0.8824	0.0101	0.0	0.0	0.0	0.0	0.0052
	len	1	84.1735	84.5533	84.8488				
100	100	0	0.9594	1.0	1.0	1.0	1.0	0.9798	0.9098
	len	2	8.2496	8.3314	8.4043				
100	100	0.1	0.9206	1.0	1.0	0.202	0.0	0.0	0.1148
	len	2	116.3386	116.9796	117.8122				
100	100	0	0.9550	1.0	1.0	1.0	1.0	0.8990	0.8724
	len	3	6.0025	6.0375	6.0724				
100	100	0.1	0.8686	0.0	0.0	0.0	0.0	0.0	0.0108
	len	3	83.3535	83.7542	84.1129				
100	100	0	0.9550	1.0	1.0	1.0	1.0	0.7778	0.8282
	len	4	4.8565	4.8795	4.9060				
100	100	0.1	0.7960	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	66.2198	66.5422	66.8346				
100	100	0	0.9672	1.0	1.0	1.0	1.0	1.0	0.9320
	len	5	13.5602	13.7399	13.9179				
100	100	0.1	0.9464	1.0	1.0	1.0	1.0	0.0	0.3876
	len	5	190.8627	193.5616	195.2268				

Table 4.10. Cov(t,Y), wtype=3, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9542	1.0	1.0	1.0	1.0	0.6465	0.7736
	len	1	4.1875	4.2033	4.2424				
100	100	0.1	0.6546	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	55.6332	55.7922	55.9306				
100	100	0	0.9536	1.0	1.0	1.0	1.0	0.7172	0.7750
	len	2	4.2733	4.2880	4.3857				
100	100	0.1	0.6548	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	55.7211	55.8550	56.0081				
100	100	0	0.9528	1.0	1.0	1.0	1.0	0.6566	0.7764
	len	3	4.1907	4.2048	4.2464				
100	100	0.1	0.6456	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	55.5695	55.7683	55.9422				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.6869	0.7744
	len	4	4.1605	4.1753	4.1899				
100	100	0.1	0.6518	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	55.8532	55.9591	56.1145				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.7071	0.8034
	len	5	4.5595	4.5785	4.9064				
100	100	0.1	0.6518	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	55.9470	56.0910	56.2531				

Table 4.11. Cov(t,Y), wtype=4, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9548	1.0	1.0	1.0	1.0	0.7071	0.7734
	len	1	4.1464	4.1661	4.1864				
100	100	0	0.9518	1.0	1.0	1.0	1.0	0.6869	0.7698
	len	2	4.1451	4.158	4.1718				
100	100	0	0.9536	1.0	1.0	1.0	1.0	0.6768	0.7672
	len	3	4.1474	4.1624	4.1800				
100	100	0	0.9550	1.0	1.0	1.0	1.0	0.6566	0.7834
	len	4	4.1514	4.1657	4.1805				
100	100	0	0.9530	1.0	1.0	1.0	1.0	0.6869	0.7758
	len	5	4.1539	4.1698	4.1909				

Table 4.12. Cov(t,Y), wtype=5, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9554	1.0	1.0	1.0	1.0	0.8384	0.9342
	len	1	25.9718	26.0702	26.1539				
100	100	0.1	0.7080	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	59.8454	60.0393	60.1912				
100	100	0	0.9614	1.0	1.0	1.0	1.0	1.0	0.9504
	len	2	41.9568	42.2773	42.6367				
100	100	0.1	0.7610	0.0	0.0	0.0	0.0	0.0	0.0014
	len	2	66.5732	66.7847	67.0689				
100	100	0	0.9592	1.0	1.0	1.0	1.0	0.9899	0.9466
	len	3	25.4799	25.6927	25.8542				
100	100	0.1	0.6870	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	59.9161	60.1259	60.3537				
100	100	0	0.9554	1.0	1.0	1.0	1.0	0.6061	0.9386
	len	4	15.4473	15.4940	15.5340				
100	100	0.1	0.6688	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	57.1978	57.3393	57.5555				
100	100	0	0.9706	1.0	1.0	1.0	1.0	1.0	0.9652
	len	5	78.6911	79.5101	80.5015				
100	100	0.1	0.8254	0.0	0.0	0.0	0.0	0.0	0.0374
	len	5	88.1044	88.8691	89.4105				

Table 4.13. Cov(t,Y), wtype=6, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9544	1.0	1.0	1.0	1.0	0.6970	0.7876
	len	1	4.1519	4.1701	4.1862				
100	100	0.1	0.6476	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	55.8405	55.9836	56.1013				
100	100	0	0.9546	1.0	1.0	1.0	1.0	0.7172	0.7686
	len	2	4.1665	4.1840	4.1979				
100	100	0.1	0.6526	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	55.652	55.8189	5.9662				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.6768	0.7760
	len	3	4.1638	4.1765	4.1896				
100	100	0.1	0.6482	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	55.7288	55.8413	56.0507				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.8081	0.7708
	len	4	4.1552	4.1688	4.1834				
100	100	0.1	0.6544	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	55.5730	55.7176	55.8253				
100	100	0	0.9540	1.0	1.0	1.0	1.0	0.6566	0.7828
	len	5	4.2118	4.2288	4.2457				
100	100	0.1	0.6534	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	55.8898	55.9881	56.1200				

Table 4.14. Cov(t,Y), wtype=7, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9516	1.0	1.0	1.0	1.0	0.7071	0.7788
	len	1	4.1488	4.1652	4.1801				
100	100	0.1	0.6458	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	55.6886	55.8332	55.9948				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.7475	0.7666
	len	2	4.1813	4.1951	4.2114				
100	100	0.1	0.6572	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	55.6032	55.7499	55.9100				
100	100	0	0.9538	1.0	1.0	1.0	1.0	0.6970	0.7862
	len	3	4.1675	4.1805	4.1952				
100	100	0.1	0.6530	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	55.6779	55.8306	55.9538				
100	100	0	0.9532	1.0	1.0	1.0	1.0	0.7273	0.7634
	len	4	4.1521	4.1646	4.1768				
100	100	0.1	0.6474	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	55.6705	55.8728	56.0141				
100	100	0	0.9536	1.0	1.0	1.0	1.0	0.6768	0.7756
	len	5	4.2364	4.2489	4.2645				
100	100	0.1	0.6592	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	55.7683	55.9199	56.0564				

The simulation used Theorem 2c) for testing with nominal level 0.05. For Table 4.1, when $\psi = 0$, H_0 was true except for $\beta_{1,MMLE}(\boldsymbol{t}, Y) = \eta_1(\boldsymbol{t})$. However, the interval $[L_{1n}/s_1, U_{1n}/s_1]$ tended to contain $\eta_1(\boldsymbol{t}) = \eta_1/\sigma_1$ near 95% of the time. The maximum average interval length 0.7147 on the 2nd line of Table 4.1 corresponded to the first interval for $\eta_1(\boldsymbol{t})$. When $\psi = 0.1$ H_0 was never true. Then the minimum average coverage 0.6362 on the third line of Table 4.1 corresponded to $\eta_1(\boldsymbol{t})$. The remaining coverages were all near 0.84. Hence none of the 99 intervals had coverage over 0.9. The low coverages in the last column for testcov mean that the test for $H_0 : (\eta_{98}(\boldsymbol{t}), \eta_{99}(\boldsymbol{t}))^T = \boldsymbol{0}$ had good power. The power 0.7324= 1-0.2676 was worst for etype=5.

CHAPTER 5

CONCLUSIONS

The response plot of $\hat{\boldsymbol{\phi}}_{OPLS}$ versus Y and the EE plot of $\hat{\boldsymbol{\phi}}^T_{OPLS}\boldsymbol{x}$ versus $\hat{\boldsymbol{\phi}}^T_{OLS}\boldsymbol{x}$ can be used to check whether OPLS is useful. See Olive (2013) for more on these two plots.

Software

The *R* software was used in the simulations. See R Core Team (2020). Programs are in the Olive (2023) collections of *R* functions *slpack.txt*, available from (http://parker.ad.siu.edu/Olive/slpack.txt). The function `mmlesim2` was used to make the tables.

REFERENCES

[1] Basa, J., Cook, R.D., Forzani, L., and Marcos, M. (2022), "Asymptotic Distribution of One-Component Partial Least Squares Regression Estimators in High Dimensions," *The Canadian Journal of Statistics*, to appear.

[2] 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.

[3] Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004), "Least Angle Regression," (with discussion), *The Annals of Statistics,* 32, 407-451.

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

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

[6] Hebbler, B. (1847), "Statistics of Prussia," *Journal of the Royal Statistical Society, A*, 10, 154-186.

[7] Meinshausen, N. (2007), "Relaxed Lasso," *Computational Statistics* & *Data Analysis*, 52, 374-393.

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

[9] Olive, D.J. (2023), *Prediction and Statistical Learning*, online course notes, see (http://parker.ad.siu.edu/Olive/slearnbk.htm).

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

[11] 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.

[12] Pötscher, B.M., and Preinerstorfer, D. (2023), "How Reliable are Bootstrap-Based Heteroskedasticity Robust Tests?" *Econometric Theory*, 39, 789-847.

[13] Qi, X., Luo, R., Carroll, R.J., and Zhao, H. (2015), "Sparse Regression by Projection and Sparse Discriminant Analysis," *Journal of Computational and Graphical Statistics,* 24, 416-438.

[14] R Core Team (2020), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

[15] 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.

[16] Rathnayake, R.C., and Olive, D.J. (2023), "Bootstrapping Some GLMs and Survival Regression Models after Variable Selection," *Communications in Statistics: Theory and Methods*, 52, 2625-2645.

[17] Rinaldo, A., Wasserman, L., and G'Sell, M. (2019), "Bootstrapping and Sample Splitting for High-Dimensional, Assumption-Lean Inference," *The Annals of Statistics*, 47, 3438-3469.

[18] Romano, J.P., and Wolf, M. (2017), "Resurrecting Weighted Least Squares," *Journal of Econometrics*, 197, 1-19.

[19] Tay, J.K., Narasimhan, B. and Hastie, T. (2023), "Elastic Net Regularization Paths for All Generalized Linear Models," *Journal of Statistical Software*, 106, 1-31.

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

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

[22] Zhou, L., Cook, R.D., and Zou, H. (2023), "Enveloped Huber Regression," *Journal of the American Statistical Association*, to appear.

[23] Zou, H., and Hastie, T. (2005), "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society Series, B,* 67, 301-320.

VITA

Graduate School
Southern Illinois University

Kasun G. Pathiranage

kasun7733@gmail.com

University of Kelaniya, Sri Lanka
Bachelor of Science, Statistics, April 2021

Research Paper Title:

Testing Multiple Linear Regression with the One Component Partial Least Squares Estimator

Major Professor: Dr. David J. Olive