Summer 2024

TESTING MULTIPLE LINEAR REGRESSION WITH THE MMLE

Lakni Hettige
lakni.hettige@siu.edu

TESTING MULTIPLE LINEAR REGRESSION WITH THE MMLE

by

Lakni A. W. Hettige

B.S., University of Kelaniya, Sri Lanka, 2022

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

School of Mathematical and Statistical Sciences
in the Graduate School
Southern Illinois University Carbondale
August, 2024

RESEARCH PAPER APPROVAL

TESTING MULTIPLE LINEAR REGRESSION WITH THE MMLE

by

Lakni A. W. Hettige

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
July 1, 2024

AN ABSTRACT OF THE RESEARCH PAPER OF

LAKNI A. W. HETTIGE, for the Master of Science degree in MATHEMATICS, presented on JULY 1, 2024, at Southern Illinois University Carbondale.

TITLE: TESTING MULTIPLE LINEAR REGRESSION WITH THE MMLE

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 $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. 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 husband and my family for all of their support and encouragement. You have pushed me to succeed and I could not have done it without you!

TABLE OF CONTENTS

CHAPTER PAGE

ABSTRACT . . . . **i**

ACKNOWLEDGMENTS . . . . **ii**

LIST OF TABLES . . . . **iv**

CHAPTERS

1 Introduction **1**

2 Large Sample Theory **5**

3 Testing **9**

4 Regression with Heterogeneity **10**

5 Example and Simulations **11**

6 Conclusion **30**

REFERENCES **31**

VITA **34**

LIST OF TABLES

TABLE		PAGE
Table 5.1	Cov(x,Y), wtype=1, k=1	16
Table 5.2	Cov(x,Y), wtype=2, k=1	17
Table 5.3	Cov(x,Y), wtype=3, k=1	18
Table 5.4	Cov(x,Y), wtype=4, k=1	19
Table 5.5	Cov(x,Y), wtype=5, k=1	20
Table 5.6	Cov(x,Y), wtype=6, k=1	21
Table 5.7	Cov(x,Y), wtype=7, k=1	22
Table 5.8	Cov(x,Y), wtype=1, k=99	23
Table 5.9	Cov(x,Y), wtype=1, k=99	24
Table 5.10	Cov(x,Y), wtype=3, k=99	25
Table 5.11	Cov(x,Y), wtype=4, k=99	26
Table 5.12	Cov(x,Y), wtype=5, k=99	27
Table 5.13	Cov(x,Y), wtype=6, k=99	28
Table 5.14	Cov(x,Y), wtype=7, k=99	29

CHAPTER 1

INTRODUCTION

This chapter 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 $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, $\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}_{OLS}^T\overline{\boldsymbol{x}}$, 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

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

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 (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 $W = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T\boldsymbol{x}$, 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

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

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.

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)

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

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

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 $k \le p$, 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 $\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma_w})$ 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 $\text{Cov}(\boldsymbol{w}) = \boldsymbol{\Sigma_w}$. Then it can be shown that $n\tilde{\boldsymbol{\eta}}_n =$ $\sum_{i=1}^n (\boldsymbol{x}_i - \boldsymbol{\mu_x} + \boldsymbol{\mu_x} - \overline{\boldsymbol{x}})(Y_i - \mu_Y + \mu_Y - \overline{Y}) = \sum_i (\boldsymbol{x}_i - \boldsymbol{\mu_x})(Y_i - \mu_Y) = \sum_i \boldsymbol{w}_i - n\boldsymbol{a}_n =$ $\sum_i \boldsymbol{w}_i - n(\boldsymbol{\mu_x} - \overline{\boldsymbol{x}})(\mu_Y - \overline{Y})$.

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

$$\text{Thus } \sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma_w})$$

by Slutsky's theorem.

c) If H_0 is true, then $\boldsymbol{A\eta} = \boldsymbol{0}$, and

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

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

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

For iid cases, $\boldsymbol{\beta}_{MMLE} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma_x}\boldsymbol{\beta}_{OLS}$ where $\boldsymbol{V} = diag(\sigma_1^2, ..., \sigma_p^2) = diag(\boldsymbol{\Sigma_x})$. 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

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

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)

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

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

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

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

where by Theorem 1,

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

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

$$\boldsymbol{A}\begin{pmatrix} \sqrt{n}\left(\frac{1}{s_1^2} - \frac{1}{\sigma_1^2}\right)\hat{\boldsymbol{\Sigma}}_{x_1Y} \\ \vdots \\ \sqrt{n}\left(\frac{1}{s_p^2} - \frac{1}{\sigma_p^2}\right)\hat{\boldsymbol{\Sigma}}_{x_pY} \end{pmatrix} = \begin{pmatrix} \sqrt{n}\left(\frac{1}{s_{i_1}^2} - \frac{1}{\sigma_{i_1}^2}\right)\hat{\boldsymbol{\Sigma}}_{x_{i_1}Y} \\ \vdots \\ \sqrt{n}\left(\frac{1}{s_{i_k}^2} - \frac{1}{\sigma_{i_k}^2}\right)\hat{\boldsymbol{\Sigma}}_{x_{i_k}Y} \end{pmatrix} = o_P(1)$$

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

It can be shown that if $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}} = (c_{ij})$, then $\hat{\boldsymbol{D}}^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}} \hat{\boldsymbol{D}}^2 = (b_{ij})$ where $b_{ij} = c_{ij}/(s_i^2 s_j^2)$.

CHAPTER 3

TESTING

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 4

REGRESSION WITH HETEROGENEITY

A multiple linear regression model with heterogeneity is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i \tag{4.1}$$

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

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

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. Also $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma}_{\boldsymbol{e}} = diag(\sigma_i^2) = diag(\sigma_1^2, ..., \sigma_n^2)$ is an $n \times n$ positive definite matrix. In Section 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 5

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 $\hat{\boldsymbol{\beta}}_{OPLS} = (0.1727, 0.0311, 0.00018, 0.00018)^T$. 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. Then the program checked whether $\beta_{i,MMLE} = \sigma_{iY}/\sigma_i^2$ was in the interval $(1/s_i^2)[L_{in}, U_{in}]$. 5000 intervals were generated for each $\beta_{i,MMLE}$, and the coverage was the proportion of times $\beta_{i,MMLE}$ was in its interval. Hence if $\beta_{1,MMLE}$ was in its interval $4750/5000 = 0.95$, then the observed coverage was 0.95. This procedure correspond to a large sample test for $H_0 : \beta_{i,MMLE} = 0$ only if $\beta_{i,MMLE} = 0$. 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 $H_0 : \boldsymbol{\beta}_{I,MMLE} = \boldsymbol{0}$ was recorded and a correction factor was not used. Here $I = \{98, 99\}$.

Suppose $\boldsymbol{A}\boldsymbol{\beta}_{MMLE} = (\beta_{i_1,MMLE}, ..., \beta_{i_k,MMLE})^T = \boldsymbol{\beta}_{I,MMLE}$ where $I = \{i_1, ..., i_k\}$. Let $\hat{D}_I^2 = diag(1/\hat{\sigma}_{i_1}^2, ..., 1/\hat{\sigma}_{i_k}^2)$. Let $\boldsymbol{u} = \boldsymbol{x}_I = (x_{i1}, ..., x_{ik})^T$. The test statistic for the test $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{MMLE} = \boldsymbol{\beta}_{I,MMLE} = \boldsymbol{0}$ is $T_n = n\hat{\boldsymbol{\beta}}_{MMLE}^T \boldsymbol{A}^T (\boldsymbol{A}\hat{\boldsymbol{D}}^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}} \hat{\boldsymbol{D}}^2 \boldsymbol{A}^T)^{-1} \boldsymbol{A}\hat{\boldsymbol{\beta}}_{MMLE} = n\hat{\boldsymbol{\beta}}_{I,MMLE}^T (\hat{\boldsymbol{D}}_I^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{z}_I} \hat{\boldsymbol{D}}_I^2)^{-1} \hat{\boldsymbol{\beta}}_{I,MMLE} \xrightarrow{D} \chi_k^2$ when H_0 is true. The simulation used $I = \{98, 99\}$ and tested $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{MMLE} = (\beta_{98,MMLE}, \beta_{99,MMLE})^T = \boldsymbol{0}$. 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 $\boldsymbol{\beta}_{OLS} = \boldsymbol{1}$

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(mmlesim)
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)

mmlesim(n=100,p=100,k=1,nruns=5000,etype=1,wtype=1,psi=0,indices = c(98,99))
$lens
[1] 0.5924335 0.5958773 0.7172263
$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 1.
#then repeat to get Tables 2-7 corresponding to wtype =2,...,7
#do not use psi=0.1 for wtype=4

#then repeat with k=99 to get Tables 8-14
#so the first two line of table 8 use the following R command

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

The simulation used Theorem 2b) for testing with nominal level 0.05. For Table 5.1, when $\psi = 0$, H_0 was true except for $\beta_{1,MMLE}$. However, the interval $[L_{1n}/s_1^2, U_{1n}/s_1^2]$ tended to contain $\beta_{1,MMLE} = \eta_1/\sigma_1^2$ near 95% of the time. The maximum average interval length 0.7172 on the 2nd line of Table 5.1 corresponded to the first interval for $\beta_{1,MMLE}$. When $\psi = 0.1$ H_0 was never true. Then the minimum average coverage 0.0616 on the third line of Table 5.1 corresponded to $\beta_{1,MMLE}$. The remaining coverages were all near 0.47. 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 $H_0 : (\beta_{98,MMLE}, \beta_{99,MMLE})^T = \mathbf{0}$ had good power. The power 0.7284 = 1-0.2716 was worst for etype=5.

Table 5.1. Cov(x,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	0.5924	0.5959	0.7172				
100	100	0.1	0.0616	0.0	0.0	0.0	0.0	0.0	0.0002
	len	1	0.5634	0.5675	0.6526				
100	100	0	0.9468	1.0	1.0	1.0	1.0	0.9091	0.9506
	len	2	0.7993	0.8063	0.8977				
100	100	0.1	0.2124	0.0	0.0	0.0	0.0	0.0	0.0304
	len	2	0.6877	0.6933	0.7654				
100	100	0	0.9476	1.0	1.0	1.0	1.0	0.8283	0.9476
	len	3	0.5883	0.5927	0.7143				
100	100	0.1	0.0704	0.0	0.0	0.0	0.0	0.0	0.0006
	len	3	0.5629	0.5659	0.6530				
100	100	0	0.9450	1.0	1.0	1.0	1.0	0.7475	0.9392
	len	4	0.4813	0.4860	0.6297				
100	100	0.1	0.0314	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	0.5090	0.5118	0.6074				
100	100	0	0.9668	1.0	1.0	1.0	1.0	1.0	0.9630
	len	5	1.3696	1.3813	1.4494				
100	100	0.1	0.5352	0.0	0.0	0.0	0.0	0.0	0.2716
	len	5	1.0420	1.0518	1.1095				

Table 5.2. Cov(x,Y), wtype=2, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9546	1.0	1.0	1.0	1.0	0.7778	0.9446
	len	1	1.7674	1.7776	1.8998				
100	100	0.1	0.7760	0.0606	0.0	0.0	0.0	0.0	0.5588
	len	1	1.3769	1.3839	1.4835				
100	100	0	0.9592	1.0	1.0	1.0	1.0	0.9798	0.9508
	len	2	2.8057	2.8348	2.9808				
100	100	0.1	0.8818	0.9899	0.9899	0.8889	0.0404	0.0	0.761
	len	2	2.1529	2.1704	2.2933				
100	100	0	0.9586	1.0	1.0	1.0	1.0	0.9293	0.946
	len	3	1.7413	1.7542	1.8670				
100	100	0.1	0.688	0.0	0.0	0.0	0.0	0.0	0.4852
	len	3	1.3672	1.3752	1.4809				
100	100	0	0.9500	1.0	1.0	1.0	1.0	0.7980	0.9442
	len	4	1.0817	1.0876	1.2036				
100	100	0.1	0.509	0.0	0.0	0.0	0.0	0.0	0.171
	len	4	0.8906	0.8938	0.9841				
100	100	0	0.9706	1.0	1.0	1.0	1.0	1.0	0.9616
	len	5	5.2574	5.3112	5.5066				
100	100	0.1	0.9330	1.0	1.0	1.0	0.9899	0.6263	0.872
	len	5	3.9485	3.9880	4.1461				

Table 5.3. Cov(x,Y), wtype=3, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9378	1.0	1.0	1.0	0.9899	0.8889	0.9454
	len	1	0.7194	0.7254	0.9955				
100	100	0.1	0.4686	0.0	0.0	0.0	0.0	0.0	0.0516
	len	1	0.7705	0.7761	0.9581				
100	100	0	0.9554	1.0	1.0	1.0	1.0	0.9697	0.9522
	len	2	1.0474	1.0653	1.4667				
100	100	0.1	0.6988	0.0	0.0	0.0	0.0	0.0	0.319
	len	2	1.0830	1.0933	1.3541				
100	100	0	0.9442	1.0	1.0	1.0	1.0	0.9899	0.9536
	len	3	0.7110	0.7182	0.9824				
100	100	0.1	0.42	0.0	0.0	0.0	0.0	0.0	0.0626
	len	3	0.7591	0.7646	0.9418				
100	100	0	0.9366	1.0	1.0	1.0	0.9899	0.8384	0.9452
	len	4	0.5380	0.5420	0.7454				
100	100	0.1	0.1394	0.0	0.0	0.0	0.0	0.0	0.0004
	len	4	0.5891	0.5936	0.7261				
100	100	0	0.9692	1.0	1.0	1.0	1.0	1.0	0.9638
	len	5	1.8420	1.8642	2.5108				
100	100	0.1	0.8446	0.9899	0.0202	0.0	0.0	0.0	0.5986
	len	5	1.8849	1.9095	2.3632				

Table 5.4. Cov(x,Y), wtype=4, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9482	1.0	1.0	1.0	1.0	0.7778	0.9472
	len	1	0.4183	0.4215	0.5824				
100	100	0	0.9474	1.0	1.0	1.0	1.0	0.7879	0.946
	len	2	0.4186	0.4215	0.5812				
100	100	0	0.9442	1.0	1.0	1.0	1.0	0.7475	0.9396
	len	3	0.4180	0.4213	0.5824				
100	100	0	0.9474	1.0	1.0	1.0	1.0	0.7980	0.9396
	len	4	0.4183	0.4216	0.5842				
100	100	0	0.9432	1.0	1.0	1.0	1.0	0.7980	0.9384
	len	5	0.4182	0.4220	0.5823				

Table 5.5. Cov(x,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	0.9454
	len	1	25.8386	25.9548	26.0704				
100	100	0.1	0.9588	1.0	1.0	1.0	1.0	0.9798	0.9428
	len	1	16.0245	16.1075	16.1926				
100	100	0	0.9582	1.0	1.0	1.0	1.0	0.9899	0.9552
	len	2	41.8352	42.2946	42.7194				
100	100	0.1	0.962	1.0	1.0	1.0	1.0	1.0	0.9496
	len	2	25.3209	25.6154	25.8194				
100	100	0	0.9586	1.0	1.0	1.0	1.0	0.9697	0.9486
	len	3	25.3915	25.5700	25.7321				
100	100	0.1	0.9624	1.0	1.0	1.0	1.0	1.0	0.9556
	len	3	15.5181	15.6073	15.7101				
100	100	0	0.9534	1.0	1.0	1.0	1.0	0.6465	0.9344
	len	4	15.0130	15.0736	15.1274				
100	100	0.1	0.9552	1.0	1.0	1.0	1.0	0.8586	0.9344
	len	4	9.3428	9.3775	9.4283				
100	100	0	0.9712	1.0	1.0	1.0	1.0	1.0	0.9678
	len	5	78.9781	80.0866	81.0103				
100	100	0.1	0.9702	1.0	1.0	1.0	1.0	1.0	0.9666
	len	5	46.9618	47.3816	47.8914				

Table 5.6. Cov(x,Y), wtype=6, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9392	1.0	1.0	1.0	0.9899	0.7576	0.9412
	len	1	0.4758	0.4793	0.6242				
100	100	0.1	0.1196	0.0	0.0	0.0	0.0	0.0	0.0048
	len	1	0.6270	0.6316	0.7083				
100	100	0	0.9460	1.0	1.0	1.0	1.0	0.9091	0.9494
	len	2	0.5615	0.5662	0.6941				
100	100	0.1	0.3294	0.0	0.0	0.0	0.0	0.0	0.1016
	len	2	0.8122	0.8182	0.8810				
100	100	0	0.9472	1.0	1.0	1.0	1.0	0.7677	0.94
	len	3	0.4749	0.4784	0.6240				
100	100	0.1	0.1204	0.0	0.0	0.0	0.0	0.0	0.0124
	len	3	0.6244	0.6274	0.7057				
100	100	0	0.9432	1.0	1.0	1.0	1.0	0.7475	0.9458
	len	4	0.4386	0.4417	0.5948				
100	100	0.1	0.0362	0.0	0.0	0.0	0.0	0.0	0.0004
	len	4	0.5337	0.5373	0.6265				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.9899	0.9562
	len	5	0.8212	0.8308	0.9290				
100	100	0.1	0.582	0.0	0.0	0.0	0.0	0.0	0.348
	len	5	1.2700	1.2790	1.3235				

Table 5.7. Cov(x,Y), wtype=7, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9408	1.0	1.0	1.0	1.0	0.8182	0.9442
	len	1	0.4945	0.4984	0.6379				
100	100	0.1	0.0296	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	0.5059	0.5084	0.6034				
100	100	0	0.9496	1.0	1.0	1.0	1.0	0.8081	0.9508
	len	2	0.6044	0.6095	0.7335				
100	100	0.1	0.0642	0.0	0.0	0.0	0.0	0.0	0.0032
	len	2	0.5489	0.5526	0.6412				
100	100	0	0.9448	1.0	1.0	1.0	1.0	0.7980	0.9482
	len	3	0.4939	0.4973	0.6371				
100	100	0.1	0.0322	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	0.5073	0.5101	0.6064				
100	100	0	0.9398	1.0	1.0	1.0	0.9899	0.7475	0.9448
	len	4	0.4452	0.4484	0.6029				
100	100	0.1	0.0216	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	0.4881	0.4908	0.5889				
100	100	0	0.9554	1.0	1.0	1.0	1.0	0.9899	0.953
	len	5	0.9289	0.9377	1.0237				
100	100	0.1	0.2136	0.0	0.0	0.0	0.0	0.0	0.048
	len	5	0.6935	0.6979	0.7683				

Table 5.8. Cov(x,Y), wtype=1, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9522	1.0	1.0	1.0	1.0	0.6364	0.7606
	len	1	4.2015	4.2193	4.2395				
100	100	0.1	0.0856	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	39.6266	39.7735	39.8931				
100	100	0	0.9536	1.0	1.0	1.0	1.0	0.6566	0.7804
	len	2	4.2380	4.2573	4.2847				
100	100	0.1	0.0868	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	39.7353	39.8570	39.9980				
100	100	0	0.9542	1.0	1.0	1.0	1.0	0.6667	0.763
	len	3	4.2053	4.2244	4.2489				
100	100	0.1	0.0866	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	39.6370	39.8321	39.9596				
100	100	0	0.9534	1.0	1.0	1.0	1.0	0.6162	0.7714
	len	4	4.1830	4.1981	4.2145				
100	100	0.1	0.0878	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	39.6609	39.7963	39.9213				
100	100	0	0.9544	1.0	1.0	1.0	1.0	0.6566	0.797
	len	5	4.3955	4.4142	4.4411				
100	100	0.1	0.0906	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	39.7994	39.9387	40.0825				

Table 5.9. Cov(x,Y), wtype=1, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9566	1.0	1.0	1.0	1.0	0.8384	0.8714
	len	1	6.0682	6.1104	6.1547				
100	100	0.1	0.5324	0.0	0.0	0.0	0.0	0.0	0.001
	len	1	59.8775	60.1838	60.4348				
100	100	0	0.9588	1.0	1.0	1.0	1.0	0.9798	0.9084
	len	2	8.2322	8.3247	8.4038				
100	100	0.1	0.7222	0.0	0.0	0.0	0.0	0.0	0.112
	len	2	82.6070	83.1059	83.6569				
100	100	0	0.9560	1.0	1.0	1.0	1.0	0.8788	0.857
	len	3	6.0430	6.0813	6.1275				
100	100	0.1	0.4908	0.0	0.0	0.0	0.0	0.0	0.0084
	len	3	59.4855	59.8701	60.2404				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.7677	0.8312
	len	4	4.8889	4.9096	4.9323				
100	100	0.1	0.249	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	47.3216	47.4682	47.6399				
100	100	0	0.9672	1.0	1.0	1.0	1.0	1.0	0.9358
	len	5	13.7658	13.9827	14.1819				
100	100	0.1	0.8364	0.0	0.0	0.0	0.0	0.0	0.379
	len	5	136.5261	137.9956	139.6727				

Table 5.10. Cov(x,Y), wtype=3, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9532	1.0	1.0	1.0	1.0	0.7475	0.7806
	len	1	4.2175	4.2378	4.2579				
100	100	0.1	0.0868	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	39.7822	39.8852	40.0177				
100	100	0	0.954	1.0	1.0	1.0	1.0	0.6869	0.787
	len	2	4.3028	4.3193	4.4224				
100	100	0.1	0.089	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	39.7079	39.8377	39.9930				
100	100	0	0.9560	1.0	1.0	1.0	1.0	0.6869	0.7798
	len	3	4.2230	4.2387	4.2745				
100	100	0.1	0.0908	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	39.8010	39.9154	40.0738				
100	100	0	0.9520	1.0	1.0	1.0	1.0	0.7071	0.7808
	len	4	4.1897	4.2070	4.2234				
100	100	0.1	0.0886	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	39.7975	39.9045	40.0552				
100	100	0	0.9542	1.0	1.0	1.0	1.0	0.7879	0.8054
	len	5	4.5927	4.6159	4.9178				
100	100	0.1	0.089	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	39.8697	40.0144	40.1339				

Table 5.11. Cov(x,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	0.6667	0.7654
	len	1	4.1817	4.1997	4.2172				
100	100	0	0.9548	1.0	1.0	1.0	1.0	0.6970	0.7822
	len	2	4.1837	4.1964	4.2137				
100	100	0	0.9530	1.0	1.0	1.0	1.0	0.6465	0.7736
	len	3	4.1745	4.1944	4.2111				
100	100	0	0.9530	1.0	1.0	1.0	1.0	0.6162	0.7682
	len	4	4.1829	4.2015	4.2165				
100	100	0	0.9560	1.0	1.0	1.0	1.0	0.6465	0.774
	len	5	4.1768	4.1948	4.2152				

Table 5.12. Cov(x,Y), wtype=5, 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	0.8384	0.9368
	len	1	26.2274	26.3463	26.4535				
100	100	0.1	0.1484	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	42.7226	42.90669	43.04279				
100	100	0	0.9596	1.0	1.0	1.0	1.0	0.9899	0.9468
	len	2	42.2069	42.6059	42.9960				
100	100	0.1	0.249	0.0	0.0	0.0	0.0	0.0	0.0016
	len	2	47.5802	47.7618	48.1024				
100	100	0	0.9582	1.0	1.0	1.0	1.0	0.9697	0.9468
	len	3	25.6196	25.7906	25.9427				
100	100	0.1	0.1576	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	42.6949	42.8347	42.9833				
100	100	0	0.9546	1.0	1.0	1.0	1.0	0.6263	0.9324
	len	4	15.5986	15.6459	15.6971				
100	100	0.1	0.1108	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	40.7900	40.8949	41.0656				
100	100	0	0.9706	1.0	1.0	1.0	1.0	1.0	0.9618
	len	5	79.1203	80.1179	81.2339				
100	100	0.1	0.4888	0.0	0.0	0.0	0.0	0.0	0.0312
	len	5	63.0536	63.4663	63.9176				

Table 5.13. Cov(x,Y), wtype=6, 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	0.6364	0.7662
	len	1	4.1789	4.1953	4.2146				
100	100	0.1	0.0884	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	39.7041	39.8250	39.9557				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.7374	0.7812
	len	2	4.1980	4.2180	4.2351				
100	100	0.1	0.0894	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	39.6567	39.8099	39.9254				
100	100	0	0.9538	1.0	1.0	1.0	1.0	0.6566	0.7762
	len	3	4.1906	4.2056	4.2291				
100	100	0.1	0.0878	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	39.7497	39.8436	39.9748				
100	100	0	0.9542	1.0	1.0	1.0	1.0	0.6061	0.7724
	len	4	4.1826	4.2015	4.2200				
100	100	0.1	0.0906	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	39.7650	39.8915	40.0254				
100	100	0	0.9544	1.0	1.0	1.0	1.0	0.7677	0.7646
	len	5	4.2335	4.2583	4.2725				
100	100	0.1	0.0854	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	39.7333	39.8840	39.9717				

Table 5.14. Cov(x,Y), wtype=7, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9564	1.0	1.0	1.0	1.0	0.7374	0.7698
	len	1	4.1871	4.2082	4.2245				
100	100	0.1	0.0878	0.0	0.0	0.0	0.0	0.0	0.0
	len	1	39.7434	39.8520	39.97510				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.6061	0.7774
	len	2	4.2087	4.2269	4.2431				
100	100	0.1	0.0942	0.0	0.0	0.0	0.0	0.0	0.0
	len	2	39.7524	39.8647	39.9743				
100	100	0	0.9538	1.0	1.0	1.0	1.0	0.6970	0.77
	len	3	4.1933	4.2078	4.2290				
100	100	0.1	0.0948	0.0	0.0	0.0	0.0	0.0	0.0
	len	3	39.7562	39.9038	40.0663				
100	100	0	0.9526	1.0	1.0	1.0	1.0	0.6364	0.781
	len	4	4.1778	4.1969	4.2174				
100	100	0.1	0.0818	0.0	0.0	0.0	0.0	0.0	0.0
	len	4	39.7661	39.8912	39.9916				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.6465	0.7884
	len	5	4.2710	4.2871	4.3051				
100	100	0.1	0.0918	0.0	0.0	0.0	0.0	0.0	0.0
	len	5	39.6105	39.7635	39.8746				

CHAPTER 6

CONCLUSION

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 available from the Olive (2023) collections of *R* functions *slpack.txt*, available from (http://parker.ad.siu.edu/Olive/slpack.txt). The function `mmlesim` 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. 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

Lakni A. W. Hettige

lakni.hettige@gmail.com

University of Kelaniya, Sri Lanka
Bachelor of Science, Statistics, July 2022

Research Paper Title:
Testing Multiple Linear Regression with the MMLE

Major Professor: Dr. David J. Olive