TESTING WITH THE ONE COMPONENT PARTIAL LEAST SQUARES AND THE MARGINAL MAXIMUM LIKELIHOOD ESTIMATORS

by

Abdulaziz A. Alshammari

B.Sc., Northern Borders University, 2013
M.S., University of Massachusetts Lowell, 2017

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Doctor of Philosophy Degree

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

AN ABSTRACT OF THE DISSERTATION OF

Abdulaziz A. Alshammari, for the Doctor of Philosophy degree in Statistics, presented on May 28, 2024, at Southern Illinois University Carbondale.

TITLE: TESTING WITH THE ONE COMPONENT PARTIAL LEAST SQUARES AND THE MARGINAL MAXIMUM LIKELIHOOD ESTIMATORS

MAJOR PROFESSOR: Dr. David Olive

We derive some large sample theory for the marginal maximum likelihood estimator for multiple linear regression. Then testing is considered for that estimator and the one component partial least squares estimator, including some high dimensional tests. Testing with these two estimators for the multiple linear regression model with heterogeneity and for the single index model is also considered.

KEY WORDS: Data splitting, dimension reduction, high dimensional data, lasso, single index model.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. David Olive for his invaluable assistance and insights leading to the writing of this paper. He was always available to meet and discuss when I've questions. My sincere thanks also goes to professors of my committee for their proper suggestions. Last but not least, I would like to thank the faculty and staff of the School of Mathematical and Statistical Sciences. I've enjoyed studying and working with you.

DEDICATION

I want to dedicate this work to my parents for always being there for me. They've been my biggest supporters and inspiration. To my amazing wife, thank you for your endless patience and belief in me. To my family, your encouragement has kept me going.

Contents

List of Tables

CHAPTER 1

INTRODUCTION

This section reviews multiple linear regression models, including variable selection and data splitting. Consider a multiple linear regression model with response variable Y and predictors $\boldsymbol{x} = (x_1, ..., x_p)^T$. 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

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

for $i = 1, ..., n$. Here n is the sample size and the random variable e_i is the ith error. Assume that the e_i are independent and identically distributed (iid) with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma^2$. In matrix notation, these n equations become $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors.

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

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

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

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

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 Cov$(\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_n$ where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. For a multiple linear regression model with heterogeneity, assume model (3) holds with $E(\boldsymbol{e}) = \boldsymbol{0}$ and 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. When the σ_i^2 are known, weighted least squares (WLS) is often used.

Under regularity conditions, $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ can be shown to be a consistent estimator of $\boldsymbol{\phi}$ with Cov$(\hat{\boldsymbol{\phi}}) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\Sigma}_{\boldsymbol{e}}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ and $E(\hat{\boldsymbol{\phi}}) = \boldsymbol{\phi}$. See, for example, White (1980). Assume nCov$(\hat{\boldsymbol{\phi}}) \to \boldsymbol{V}$ as $n \to \infty$. If $\boldsymbol{X}^T\boldsymbol{X}/n \to \boldsymbol{W}^{-1}$ and $\boldsymbol{X}^T\boldsymbol{\Sigma}_{\boldsymbol{e}}\boldsymbol{X}/n \to \boldsymbol{U}$, then $\boldsymbol{V} = \boldsymbol{W}\boldsymbol{U}\boldsymbol{W}$. We assume that α is in the model so that the OLS residuals sum to 0.

Some other models are a)

$$Y_i|\boldsymbol{x}_i = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e_i$$

with $V(e_i) = V(Y_i|\boldsymbol{x}_i) = \sigma_i^2 = \sigma^2(\boldsymbol{x}_i)$, b)

$$Y_i|(\boldsymbol{x}_i, \boldsymbol{\beta}^T\boldsymbol{x}_i) = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e_i$$

with $V(e_i) = V(Y_i|\boldsymbol{x}_i, \boldsymbol{\beta}^T\boldsymbol{x}_i) = \sigma_i^2 = \sigma^2(\boldsymbol{x}_i)$, and c)

$$Y_i|\boldsymbol{\beta}^T\boldsymbol{x}_i = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e_i$$

with $V(e_i) = V(Y_i|\boldsymbol{\beta}^T\boldsymbol{x}_i) = \tau_i^2 = \tau^2(\boldsymbol{x}_i)$. See Rajapaksha and Olive (2024). Variants of these models use $e_i = \sigma(\boldsymbol{x}_i)\epsilon_i$ or $e_i = \tau(\boldsymbol{x}_i)\epsilon_i$ where the ϵ_i are iid with $E(\epsilon_i) = 0$ and $V(\epsilon_i) = 1$. Another variant uses iid cases $(\boldsymbol{x}_i, Y_i)$. Suppose the ϵ_i are iid and independent of the iid $(\boldsymbol{x}_i, Y_i)$. Then the $(\boldsymbol{x}_i, Y_i, \epsilon_i)$ are iid, and the above models can be formed, e.g., $Y_i|(\boldsymbol{x}_i, \boldsymbol{\beta}^T\boldsymbol{x}_i) = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + \sigma(\boldsymbol{x}_i)\epsilon_i$.

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

Let

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

and

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

Then the OLS estimators for model (3) are $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, $\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}_{OLS}^T\overline{\boldsymbol{x}}$, and

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

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

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

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

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

As noted by Rajapaksha and Olive (2024), the *nonparametric bootstrap = pairs bootstrap* samples the cases $(\boldsymbol{x}_i^T, Y_i)^T$ with replacement, and uses

$$\boldsymbol{Y}^* = \boldsymbol{X}^* \hat{\boldsymbol{\phi}} + \boldsymbol{r}^* \tag{5}$$

where the $(\boldsymbol{x}_i^T, Y_i, r_i)^T$ are selected with replacement to form $\boldsymbol{Y}^*, \boldsymbol{X}^*$ and $\boldsymbol{r}^*$. Note that with respect to the bootstrap distribution, the $(\boldsymbol{x}_i^{*T}, Y_i^*, r_i)^T$ are iid. Hence Equation (5) is an iid regression model. Freedman (1981) showed that the nonparametric bootstrap with $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}_{OLS}$ can be useful for model (1) when the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid. Since the residuals from $\hat{\boldsymbol{\beta}}_{OPLS}$ sum to zero, the nonparametric bootstrap may be useful for OPLS.

The nonparametric bootstrap for $\tilde{\eta}$ samples the cases with replacement and computes $\tilde{\boldsymbol{\eta}}^*$ from the resulting bootstrap data set. Then

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

Hence the tests from the nonparametric bootstrap and the much faster tests from Chapter 2 should be similar.

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}. \tag{6}$$

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \boldsymbol{I}^{-1} \hat{\Sigma}_{\boldsymbol{t},Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{t}, Y) \tag{7}$$

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

Variable selection estimators include forward selection or backward elimination when $n \geq 10p$.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 section 2.3 considers tests of hypotheses.

CHAPTER 2

LARGE SAMPLE THEORY AND TESTING

2.1 OLS Theory

Let the **MLR model 1** be

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{8}$$

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 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 **MLR model 2** be

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{9}$$

for $i = 1, ..., n$. For this model, we may use $\boldsymbol{\phi} = (\alpha, \boldsymbol{\beta}^T)^T$ with $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}$.

Ordinary least squares (OLS) large sample theory will be useful. Let $\boldsymbol{X} = (\boldsymbol{1} \ \ \boldsymbol{X}_1)$. For model (8), the ith row of $\boldsymbol{X}$ is $(1, x_{i,2}, ..., x_{i,p})$ while for model (9), the ith row of $\boldsymbol{X}$ is $(1, x_{i,1}, ..., x_{i,p})$, and $\boldsymbol{Y} = \alpha\boldsymbol{1} + \boldsymbol{X}_1\boldsymbol{\beta} + \boldsymbol{e} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}$.

Definition 2.1 Using the above notation for model (8), let $\boldsymbol{x}_i^T = (x_{i1}, ..., x_{ip})$, let α be the intercept, and let the slopes vector $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$. Let the population covariance matrices

$$\text{Cov}(\boldsymbol{x}) = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T] = \boldsymbol{\Sigma}_{\boldsymbol{x}}, \ \ \text{and}$$

$$\text{Cov}(\boldsymbol{x}, Y) = E[(\boldsymbol{x} - E(\boldsymbol{x}))(Y - E(Y))] = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}.$$

If the cases $(\boldsymbol{x}_i, Y_i)$ are iid from some population where $\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ exists and $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ is nonsingular, then the population coefficients from an OLS regression of Y on $\boldsymbol{x}$ (even if a linear model does not hold) are

$$\alpha = \alpha_{OLS} = E(Y) - \boldsymbol{\beta}^T E(\boldsymbol{x}) \text{ and } \boldsymbol{\beta} = \boldsymbol{\beta}_{\text{OLS}} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}\text{Y}}.$$

Definition 2.2 Let the sample covariance matrices be

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

Let the method of moments estimators be $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i Y_i - \overline{\boldsymbol{x}}\, \overline{Y}$.

The method of moment estimators are often called the maximum likelihood estimators, but are the MLE if the $(Y_i, \boldsymbol{x}_i^T)^T$ are iid from a multivariate normal distribution, a very strong assumption. In Theorem 2.1, note that $D = \boldsymbol{X}_1^T \boldsymbol{X}_1 - n\overline{\boldsymbol{x}}\, \overline{\boldsymbol{x}}^T = (n-1)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}$.

Theorem 2.1: Seber and Lee (2003, p. 106). Let $\boldsymbol{X} = (\boldsymbol{1} \;\; \boldsymbol{X}_1)$.

Then $\boldsymbol{X}^T \boldsymbol{Y} = \begin{pmatrix} n\overline{Y} \\ \boldsymbol{X}_1^T \boldsymbol{Y} \end{pmatrix} = \begin{pmatrix} n\overline{Y} \\ \sum_{i=1}^n \boldsymbol{x}_i Y_i \end{pmatrix}$, $\boldsymbol{X}^T \boldsymbol{X} = \begin{pmatrix} n & n\overline{\boldsymbol{x}}^T \\ n\overline{\boldsymbol{x}} & \boldsymbol{X}_1^T \boldsymbol{X}_1 \end{pmatrix}$,

$$\text{and } (\boldsymbol{X}^T \boldsymbol{X})^{-1} = \begin{pmatrix} \frac{1}{n} + \overline{\boldsymbol{x}}^T D^{-1} \overline{\boldsymbol{x}} & -\overline{\boldsymbol{x}}^T D^{-1} \\ -D^{-1}\overline{\boldsymbol{x}} & D^{-1} \end{pmatrix}$$

where the $p \times p$ matrix $D^{-1} = [(n-1)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}]^{-1} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}/(n-1)$.

Under model (9), $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$.

Theorem 2.2: Second way to compute $\hat{\phi}$:

a) If $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}$ exists, then $\hat{\alpha} = \overline{Y} - \hat{\boldsymbol{\beta}}^T \overline{\boldsymbol{x}}$ and

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

b) Suppose that $(Y_i, \boldsymbol{x}_i^T)^T$ are iid random vectors such that σ_Y^2, $\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}$, and $\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ exist. Then $\hat{\alpha} \xrightarrow{P} \alpha$ and

$$\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta} \text{ as n} \to \infty$$

where α and $\boldsymbol{\beta}$ are given by Definition 2.1.

Proof. Note that

$$\boldsymbol{Y}^T \boldsymbol{X}_1 = (Y_1 \cdots Y_n) \begin{bmatrix} \boldsymbol{x}_1^T \\ \vdots \\ \boldsymbol{x}_n^T \end{bmatrix} = \sum_{i=1}^n Y_i \boldsymbol{x}_i^T$$

and

$$\boldsymbol{X}_1^T \boldsymbol{Y} = [\boldsymbol{x}_1 \cdots \boldsymbol{x}_n] \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \sum_{i=1}^n \boldsymbol{x}_i Y_i.$$

So

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} + \overline{\boldsymbol{x}}^T D^{-1} \overline{\boldsymbol{x}} & -\overline{\boldsymbol{x}}^T D^{-1} \\ -D^{-1} \overline{\boldsymbol{x}} & D^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{1}^T \\ \boldsymbol{X}_1^T \end{bmatrix} \boldsymbol{Y} =$$

$$\begin{bmatrix} \frac{1}{n} + \overline{\boldsymbol{x}}^T D^{-1} \overline{\boldsymbol{x}} & -\overline{\boldsymbol{x}}^T D^{-1} \\ -D^{-1} \overline{\boldsymbol{x}} & D^{-1} \end{bmatrix} \begin{bmatrix} n\overline{Y} \\ \boldsymbol{X}_1^T \boldsymbol{Y} \end{bmatrix}.$$

Thus $\hat{\boldsymbol{\beta}} = -nD^{-1} \overline{\boldsymbol{x}}\, \overline{Y} + D^{-1} \boldsymbol{X}_1^T \boldsymbol{Y} = D^{-1}(\boldsymbol{X}_1^T \boldsymbol{Y} - n\overline{\boldsymbol{x}}\, \overline{Y}) =$

$$D^{-1} \left[\sum_{i=1}^n \boldsymbol{u}_i Y_i - n\overline{\boldsymbol{x}}\, \overline{Y} \right] = \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1}}{n-1} n \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \frac{n}{n-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}. \text{ Then}$$

$\hat{\alpha} = \overline{Y} + n\overline{\boldsymbol{x}}^T D^{-1}\overline{\boldsymbol{x}}\,\overline{Y} - \overline{\boldsymbol{x}}^T D^{-1}\boldsymbol{X}_1^T\boldsymbol{Y} = \overline{Y} + [n\overline{Y}\overline{\boldsymbol{x}}^T D^{-1} - \boldsymbol{Y}^T\boldsymbol{X}_1 D^{-1}]\overline{\boldsymbol{x}} = \overline{Y} - \hat{\boldsymbol{\beta}}^T\overline{\boldsymbol{x}}$. The convergence in probability results hold since sample means and sample covariance matrices are consistent estimators of the population means and population covariance matrices. $\square$

It is important to note that the convergence in probability results are for iid $(Y_i, \boldsymbol{x}_i^T)^T$ with second moments and nonsingular $\boldsymbol{\Sigma}_{\boldsymbol{x}}$: a linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ does not need to hold. When the linear model does hold, the second method for computing $\hat{\boldsymbol{\beta}}$ is still valid even if $\boldsymbol{X}$ is a constant matrix, and $\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{\beta}$ by Theorem 2.3 b). Note that for Theorem 2.2 b) with iid cases and $\boldsymbol{\mu}_{\boldsymbol{x}} = E(\boldsymbol{x})$,

$$n(\boldsymbol{X}^T\boldsymbol{X})^{-1} \xrightarrow{P} \boldsymbol{V} = \begin{bmatrix} 1 + \boldsymbol{\mu}_{\boldsymbol{x}}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\mu}_{\boldsymbol{x}} & -\boldsymbol{\mu}_{\boldsymbol{x}}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \\ -\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\mu}_{\boldsymbol{x}} & \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \end{bmatrix}$$

There are many large sample theory results for ordinary least squares. The following theorem is important. See, for example, Sen and Singer (1993, p. 280).

Theorem 2.3, OLS CLTs. Consider the MLR model and assume that the zero mean errors are iid with $E(e_i) = 0$ and VAR$(e_i) = \sigma^2$. If the $\boldsymbol{x}_i$ are random vectors, assume that the cases $(\boldsymbol{x}_i, Y_i)$ are independent, and that the $\boldsymbol{e}_i$ and $\boldsymbol{x}_i$ are independent. Also assume that $\max_i(h_1, ..., h_n) \to 0$ and

$$\frac{\boldsymbol{X}^T\boldsymbol{X}}{n} \to \boldsymbol{V}^{-1}$$

as $n \to \infty$ where the convergence is in probability if the $\boldsymbol{x}_i$ are random vectors (instead of nonstochastic constant vectors).

a) For Equation (8), the OLS estimator $\hat{\boldsymbol{\beta}}$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\,\boldsymbol{V}). \tag{10}$$

b) For Equation (9), the OLS estimator $\hat{\boldsymbol{\phi}}$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{D} N_{p+1}(\mathbf{0}, \sigma^2 \, \boldsymbol{V}). \tag{11}$$

c) Suppose the cases $(\boldsymbol{x}_i, Y_i)$ are iid from some population and the Equation (9) MLR model $Y_i = \alpha + \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i$ holds. Assume that $\boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x},Y}$ exist. Then Equation (11) holds and

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \sigma^2 \, \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}) \tag{12}$$

where $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$.

Remark 2.1 Consider Theorem 2.3 For a) and b), the theory acts as if the $\boldsymbol{x}_i$ are constant even if the $\boldsymbol{x}_i$ are random vectors. The literature says the $\boldsymbol{x}_i$ can be constants, or condition on $\boldsymbol{x}_i$ if the $\boldsymbol{x}_i$ are random vectors. The main assumptions for a) and b) are that the errors are iid with second moments and the $n(\boldsymbol{X}^T \boldsymbol{X})^{-1}$ is well behaved. The strong assumptions for c) are much stronger than those for a) and b), but the assumption of iid cases is often reasonable if the cases come from some population.

Remark 2.2 Consider MLR model (9). Let $\boldsymbol{w}_i = \boldsymbol{A}_n \boldsymbol{x}_i$ for $i = 1, ..., n$ where $\boldsymbol{A}_n$ is a full rank $k \times p$ matrix with $1 \leq k \leq p$.

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

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

$\boldsymbol{\Sigma}_{\boldsymbol{w}Y} = \boldsymbol{A}_n \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$.

c) Let $\hat{\boldsymbol{\beta}}(\boldsymbol{u}, Y)$ and $\boldsymbol{\beta}(\boldsymbol{u}, Y)$ be the estimator and parameter from the OLS regression of Y on $\boldsymbol{u}$. The constant parameter vector should not depend on n. Suppose the cases are iid and $\boldsymbol{A}$ is a constant matrix that does not depend on n. By Theorem 2.2, $\hat{\boldsymbol{\beta}}(\boldsymbol{w}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}Y} = [\boldsymbol{A}_n \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \boldsymbol{A}_n]^{-1} \boldsymbol{A}_n \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = [\boldsymbol{A}_n \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \boldsymbol{A}_n]^{-1} \boldsymbol{A}_n \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \hat{\boldsymbol{\beta}}(\boldsymbol{x}, Y)$. If $\boldsymbol{A}_n \xrightarrow{P} \boldsymbol{A}$, $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{x}}$, and $\hat{\boldsymbol{\beta}}(\boldsymbol{x}, Y) \xrightarrow{P} \boldsymbol{\beta}(\boldsymbol{x}, Y)$, then $\hat{\boldsymbol{\beta}}(\boldsymbol{w}, Y) \xrightarrow{P} \boldsymbol{\beta}(\boldsymbol{w}, Y) = [\boldsymbol{A} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{A}]^{-1} \boldsymbol{A} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}(\boldsymbol{x}, Y)$.

A problem with OLS, is that $\boldsymbol{V}$ generally can't be estimated if $p > n$ since typically $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ does not exist. If $p > n$, using $\hat{\boldsymbol{\phi}} = (\boldsymbol{X}^T\boldsymbol{X})^{-}\boldsymbol{X}^T\boldsymbol{Y}$ is a poor estimator that interpolates the data, where $\boldsymbol{A}^-$ is a generalized inverse of $\boldsymbol{A}$. Often the software will not compute $\hat{\boldsymbol{\phi}}$ if $p > n$.

2.2 OPLS and $\hat{\Sigma}_{x,Y}$ 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.

Remark 2.3 The following result is useful for several multiple linear regression estimators. Let $\boldsymbol{w}_i = \boldsymbol{A}_n\boldsymbol{x}_i$ for $i = 1, ..., n$ where $\boldsymbol{A}_n$ is a full rank $k \times p$ matrix with $1 \leq k \leq p$.

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

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

The following Olive and Zhang (2024) theorem gives the large sample theory for $\hat{\boldsymbol{\eta}} = \widehat{\text{Cov}}(\boldsymbol{x}, Y)$, but the proof in this dissertation is new. 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

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

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

Theorem 2.4 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 $\boldsymbol{\mu_x} = E(\boldsymbol{x})$ and $\mu_Y = E(Y)$. Let $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu_x})(Y_i - \boldsymbol{\mu}_Y)$ with sample mean $\overline{\boldsymbol{w}}_n$. Let $\boldsymbol{\eta} = \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$. Then a)

$$\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma_w}),\ \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma_w}), \tag{13}$$

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

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

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) \xrightarrow{D} N_k(\boldsymbol{0}, \lambda^2 \boldsymbol{A}\boldsymbol{\Sigma_w}\boldsymbol{A}^T). \tag{14}$$

Proof. Part a) is a special case of Theorem 2.5.

b) $\boldsymbol{w}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}} + \overline{\boldsymbol{x}} - \boldsymbol{\mu_x})(Y_i - \overline{Y} + \overline{Y} - \mu_Y) =$

$$\boldsymbol{v}_i + (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\overline{Y} - \mu_Y) + (\overline{\boldsymbol{x}} - \boldsymbol{\mu_x})(Y_i - \overline{Y}) + (\overline{\boldsymbol{x}} - \boldsymbol{\mu_x})(\overline{Y} - \mu_Y).$$

Thus $\boldsymbol{w}_i - \overline{\boldsymbol{w}} = \boldsymbol{v}_i - \overline{\boldsymbol{v}} + \boldsymbol{a}_i$ where

$$\boldsymbol{a}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\overline{Y} - \mu_Y) + (\overline{\boldsymbol{x}} - \boldsymbol{\mu_x})(Y_i - \overline{Y}) = O_P(n^{-1/2}).$$

Thus

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{w}_i - \overline{\boldsymbol{w}})(\boldsymbol{w}_i - \overline{\boldsymbol{w}})^T = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{v}_i - \overline{\boldsymbol{v}})(\boldsymbol{v}_i - \overline{\boldsymbol{v}})^T + O_P(n^{-1/2}) = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{v}} + O_P(n^{-1/2}).$$

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

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) = \sqrt{n}\boldsymbol{A}\hat{\boldsymbol{\eta}} \xrightarrow{D} N_k(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{\Sigma_w}\boldsymbol{A}^T).$$

Then $\lambda \boldsymbol{A} \eta = \boldsymbol{0}$ under H_0, and

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

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

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

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

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

For general vectors of elements, the ordering of the vectors will all be the same and be denoted vectors such as $\tilde{c} = [\tilde{\sigma}_{jk}]$, $c = [\sigma_{jk}]$, $\boldsymbol{z}_i = [(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)]$, and

$\boldsymbol{w}_i = [(x_{ij} - \mu_j)(x_{ik} - \mu_k)]$. Let $\overline{\boldsymbol{w}}_n = \sum_{i=1}^n \boldsymbol{w}_i / n$ be the sample mean of the $\boldsymbol{w}_i$. Assuming that $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{w}}$ exists, then $E(\boldsymbol{w}_i) = E(\overline{\boldsymbol{w}}_n) = \boldsymbol{c}$.

The following theorem proves that sample covariance matrices are asymptotically normal. The theorem may be a special case of the Su and Cook (2012) theory for the multivariate linear regression estimator when there are no predictors. When $p = 1$, the theory gives the large sample theory for the sample variance. See Olive (2014, pp. 276-277) and Bickel and Doksum (2007, p. 279). The Olive and Zhang (2024) large sample theory for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ and $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ is also a special case. We use $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_d$ to avoid confusion with the $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ used in Theorems 2.4 and 3.1

Theorem 2.5 Assume the cases $\boldsymbol{x}_i$ are iid and that $Cov(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_d$ exists. Using the above notation with $\boldsymbol{c}$ a $k \times 1$ vector,

i) $\sqrt{n}(\tilde{\boldsymbol{c}} - \boldsymbol{c}) \stackrel{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_d)$.

ii) $\sqrt{n}(\hat{\boldsymbol{c}} - \boldsymbol{c}) \stackrel{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_d)$.

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

Proof. Note that $\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{c}) \stackrel{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_d)$ by the multivariate central limit theorem. i) Then

$$n\, \tilde{\boldsymbol{c}} = \sum_i [(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)] = \sum_i [(x_{ij} - \mu_j + \mu_j - \overline{x}_j)(x_{ik} - \mu_k + \mu_k - \overline{x}_k)] =$$

$$\sum_i [(x_{ij} - \mu_j)(x_{ik} - \mu_k)] + \sum_i [(x_{ij} - \mu_j)(\mu_k - \overline{x}_k)] +$$

$$\sum_i [\mu_j - \overline{x}_j)(x_{ik} - \mu_k] + \sum_i [(\mu_j - \overline{x}_j)(\mu_k - \overline{x}_k)] = \sum_i \boldsymbol{w}_i - \boldsymbol{a}_n$$

where $\boldsymbol{a}_n = [n(\overline{x}_j - \mu_j)(\overline{x}_k - \mu_k)] = [\sqrt{n}(\overline{x}_j - \mu_j)\sqrt{n}(\overline{x}_k - \mu_k)] = O_P(1)$.

By the multivariate Slutsky's theorem,

$$\sqrt{n}(\tilde{\boldsymbol{c}} - \boldsymbol{c}) = \sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{c}) + \boldsymbol{a}_n/\sqrt{n} \stackrel{D}{\to} N_k(\boldsymbol{0}, \boldsymbol{\Sigma}_d)$$

since $\boldsymbol{a}_n/\sqrt{n} = o_P(1)$.

iii) $\boldsymbol{w}_i = [(x_{ij} - \mu_j)(x_{ik} - \mu_k)] = [(x_{ij} - \overline{x}_j + \overline{x}_j - \mu_j)(x_{ik} - \overline{x}_k + \overline{x}_k - \mu_k)] =$

$[(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)] + [(x_{ij} - \overline{x}_j)(\overline{x}_k - \mu_k)] + [(\overline{x}_j - \mu_j)(x_{ik} - \overline{x}_k)] + [(\overline{x}_j - \mu_j)(\overline{x}_k - \mu_k)]$.

Hence $\boldsymbol{w}_i - \overline{\boldsymbol{w}} = \boldsymbol{z}_i - \overline{\boldsymbol{z}} + \boldsymbol{a}_i$ where

$$\boldsymbol{a}_i = [(x_{ij} - \overline{x}_j)(\overline{x}_k - \mu_k)] + [(\overline{x}_j - \mu_j)(x_{ik} - \overline{x}_k)] = O_P(n^{-1/2}).$$

Thus

$$\tilde{\boldsymbol{\Sigma}}_d = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{w}_i - \overline{\boldsymbol{w}})(\boldsymbol{w}_i - \overline{\boldsymbol{w}})^T = \frac{1}{n}\sum_{i=1}^n (\boldsymbol{z}_i - \overline{\boldsymbol{z}})(\boldsymbol{z}_i - \overline{\boldsymbol{z}})^T + O_P(n^{-1/2}) = \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{z}} + O_P(n^{-1/2}). \quad \square$$

2.3 HIGH DIMENTIONAL TESTS

As noted by Olive and Zhang (2024), the following simple testing method reduces a possibly high dimensional problem to a low dimensional problem. Testing $H_0 : \boldsymbol{A\beta}_{OPLS} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A\beta}_{OPLS} \neq \boldsymbol{0}$ is equivalent to testing $H_0 : \boldsymbol{A\eta} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A\eta} \neq \boldsymbol{0}$ 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 $\boldsymbol{u} = (x_{i1}, ..., x_{ik})^T$ 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 2.4 to the predictors $\boldsymbol{u}$ used in the hypothesis test, and thus use the sample covariance matrix 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 $\sigma_{x_i,Y} = \text{Cov}(x_i, Y)$.

Note that the tests with $\hat{\boldsymbol{\eta}}$ using k distinct predictors x_{i_j} 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 OPLS 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 $\boldsymbol{u} \perp\!\!\!\perp v$.

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 $\boldsymbol{\eta}^T\boldsymbol{x}$. However, the test statistic for testing $H_0 : \boldsymbol{A\beta}_{OPLS} = \boldsymbol{0}$ from Theorem

2.4c) is the same as the test statistic for testing $H_0 : \boldsymbol{A\Sigma_{x}}_Y = \boldsymbol{0}$ from Theorem 2.4a) since

$$n\hat{\lambda}\hat{\boldsymbol{\Sigma}}^T_{\boldsymbol{x}Y}\boldsymbol{A}^T(\hat{\lambda}^2\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = n\hat{\boldsymbol{\Sigma}}^T_{\boldsymbol{x}Y}\boldsymbol{A}^T(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{D} \chi^2_k$$

if H_0 is true.

Theorem 2.5 can be used to test $H_0 : \boldsymbol{Ac} = \boldsymbol{0}$, which can reduce a high dimensional problem to a low dimensional problem. Suppose $n > 10k$, $p > n$, and $\boldsymbol{A\beta} = (\beta_{i_1}, ..., \beta_{i_k})^T$ with $i_1, i_2, ..., i_k$ distinct. Then Theorem 3.1a) in chapter 3 can be used since no inverse matrices are required, but the asymptotic covariance matrices of Theorem 3.1b) and 3.1c) are much easier to estimate.

Remark 2.4 Theorem 2.4 depends on the theory of both the sample covariance vector and the sample covariance matrix under iid cases, not on any other model such as linearity. Suppose the cases are iid, and the predictors have nonsingular covariance matrix $\boldsymbol{\Sigma_x}$. Suppose a linear model holds with $Y|\boldsymbol{x} = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e$. If the iid errors e are independent of the predictors $\boldsymbol{x}$, then under mild conditions, linearity implies that $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$ and that the covariance structure is $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \boldsymbol{\Sigma_x\beta}_{OLS}$

2.4 MULTIPLE LINEAR REGRESSION WITH HETEROGENEITY

A multiple linear regression model with heterogeneity is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i \tag{15}$$

for $i = 1, ..., n$ where the e_i are independent with $E(e_i) = 0$ and $V(e_i) = \sigma^2_i$. 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 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 Section 2.2, 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, but the large sample theory differs from that given in Section 2.1. The OLS theory for MLR with heterogeneity often assume iid cases. For the following theorem, see Romano and Wolf (2017), Freedman (1981), and White (1980).

Theorem 2.6. Assume $Y_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + e_i$ for $i = 1, ..., n$ where the cases $(Y_i, \boldsymbol{x}_i^T)^T$ are iid with "fourth moments," $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$, the $e_i = e_i(\boldsymbol{x}_i)$ are independent, $E[e_i|\boldsymbol{x}_i] = 0$, $\boldsymbol{V}^{-1} = E[\boldsymbol{x}_i \boldsymbol{x}_i^T]$, $E[e_i^2|\boldsymbol{x}_i] = v(\boldsymbol{x}_i) = \sigma_i^2$, $Cov[\mathbf{e}|\boldsymbol{X}] = diag(v(\boldsymbol{x}_1), ..., v(\boldsymbol{x}_n))$ and $\boldsymbol{\Omega} = E[v(\boldsymbol{x}_i)\boldsymbol{x}_i \boldsymbol{x}_i^T] = E[e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i^T]$. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{V\Omega V}). \tag{16}$$

Remark 2.5. a) White (1980) showed that the iid cases assumption can be weakened. Assume the cases are independent,

$$\boldsymbol{V}_n = \frac{1}{n} \sum_{i=1}^n E[\boldsymbol{x}_i \boldsymbol{x}_i^T] \xrightarrow{P} \boldsymbol{V}^{-1},$$

and

$$\boldsymbol{\Omega}_n = \frac{1}{n} \sum_{i=1}^n E[e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i^T] \xrightarrow{P} \boldsymbol{\Omega}.$$

Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{OLS} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{V}\boldsymbol{\Omega}\boldsymbol{V}).$$

b) Under the assumptions of Theorem 2.6,

$$\frac{1}{n}\boldsymbol{X}^T\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i^T \xrightarrow{P} \boldsymbol{V}^{-1}.$$

Let $\boldsymbol{D} = diag(\sigma_1^2, ..., \sigma_n^2) = \boldsymbol{\Sigma}_{\boldsymbol{e}}$ and $\hat{\boldsymbol{D}} = diag(r_1^2, ..., r_n^2)$ where r_i^2 is the ith residual from OLS regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Then $\hat{\boldsymbol{D}}$ is not a consistent estimator of $\boldsymbol{D}$. The following theorem, due to White (1980), shows that $\hat{\boldsymbol{D}}$ can be used to get a consistent estimator of $\boldsymbol{\Omega}$. This result leads to the sandwich estimators.

Theorem 2.7 Under strong regularity conditions,

$$\frac{1}{n}(\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X}) \xrightarrow{P} \boldsymbol{\Omega} \text{ and } \frac{1}{\text{n}}(\boldsymbol{X}^{\text{T}}\boldsymbol{D}\boldsymbol{X}) \xrightarrow{\text{P}} \boldsymbol{\Omega}.$$

Hence

$$n(\boldsymbol{X}^T\boldsymbol{X})^{-1}(\boldsymbol{X}^T\hat{\boldsymbol{D}}\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{X})^{-1} \xrightarrow{P} \boldsymbol{V}\boldsymbol{\Omega}\boldsymbol{V}.$$

Now write the linear model as $Y = \alpha + \boldsymbol{x}^T\boldsymbol{\beta} + e$. Under iid cases, OPLS theory does not depend on whether the error variance is constant or not. Hence Theorem 2.4 and the Section 2.3 theory still applies. If the cases are iid and linearity holds (with or without heterogeneity), then under reasonable conditions, $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Hence

$$\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}, \tag{17}$$

as noted by Olive and Zhang (2024) for when the iid errors e_i had constant variance. This result is useful for simulation.

CHAPTER 3

SOME LARGE SAMPLE THEORY FOR MMLE

The MMLE is interesting since if each predictor satisfies a marginal model, then the marginal model theory can be used to find a confidence interval for β_i for $i = 1, ..., p$ where β_i is the ith component of $\boldsymbol{\beta}_{MMLE}$. For high dimensional multiple linear regression, the above regularity condition is weaker than the common assumption that the cases $(Y_i, \boldsymbol{x}_i^T)^T$ are iid from a multivariate normal distribution. For multiple linear regression, let $\boldsymbol{V} = diag(\boldsymbol{\Sigma}_{\boldsymbol{x}}) = diag(\sigma_1^2, ..., \sigma_p^2)$. For iid cases, $\boldsymbol{\beta}_{MMLE} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_{OLS}$, and $\boldsymbol{\beta}_{MMLE} = \boldsymbol{\beta}_{OLS}$ if $\boldsymbol{\beta}_{OLS} = \boldsymbol{0}$, or if $(\boldsymbol{V}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1})\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \boldsymbol{0}$, or if $\boldsymbol{\beta}_{OLS}$ is an eigenvector of $\boldsymbol{V}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}}$ with eigenvalue 1 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 $\sqrt{n}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} - \boldsymbol{\Sigma}_{\boldsymbol{u},Y}) = \sqrt{n}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} - \hat{\boldsymbol{\Sigma}}_{\boldsymbol{u},Y}) + \sqrt{n}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u},Y} - \boldsymbol{\Sigma}_{\boldsymbol{u},Y}) = O_P(1) + \sqrt{n}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u},Y} - \boldsymbol{\Sigma}_{\boldsymbol{u},Y})$ under mild regularity conditions for iid cases. Hence $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y}$ is a $\sqrt{n}$ consistent estimator of $\boldsymbol{\Sigma}_{\boldsymbol{u},Y}$ that is not asymptotically equivalent to $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u},Y}$ unless $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \boldsymbol{0}$. 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, $\boldsymbol{\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{t},Y} = \boldsymbol{\eta}_{OPLS}(\boldsymbol{t}, Y)$.

By Theorems 2.4 and 2.5 with iid $\boldsymbol{x}_i$ replaced by iid $(\boldsymbol{x}_i^T, Y_i)^T$,

$$\sqrt{n}\left[\begin{pmatrix} s_1^2 \\ \vdots \\ s_p^2 \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \end{pmatrix} - \begin{pmatrix} \sigma_1^2 \\ \vdots \\ \sigma_p^2 \\ \boldsymbol{\Sigma}_{\boldsymbol{x}Y} \end{pmatrix}\right] = \sqrt{n}(\hat{\boldsymbol{c}} - \boldsymbol{c}) \xrightarrow{D} N_{2p}\left(\boldsymbol{0}, \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{v}} & \boldsymbol{\Sigma}_{\boldsymbol{v},\boldsymbol{w}} \\ \boldsymbol{\Sigma}_{\boldsymbol{w},\boldsymbol{v}} & \boldsymbol{\Sigma}_{\boldsymbol{w}} \end{pmatrix}\right) \quad (18)$$

Let

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

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

$\boldsymbol{D}_2 = diag(1/\sigma_1^2, 1/\sigma_2^2, ..., 1/\sigma_p^2)$. 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 3.1 Let the cases $(\boldsymbol{x}_i^T, Y_i)^T$ be iid such that Equation (18) 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)

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

c) For standardized predictors, assume $H_0 : \boldsymbol{A\beta}_{MMLE}(\boldsymbol{t}, Y) = \boldsymbol{A\Sigma}_{\boldsymbol{t},Y} = \boldsymbol{0}$. 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 3.1a) holds by the multivariate delta method.

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

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

where by Theorem 2.4,

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

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 $(\Sigma_{x_{i_1}Y}, ..., \Sigma_{x_{i_k}Y})^T = \boldsymbol{0}$. 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 2.4 and Remark 2.3,

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

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

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

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

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

The following theorem is from Olive and Zhang (2024). Note that $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ is the correlation matrix of $\boldsymbol{x}$.

Theorem 3.2 *Consider the MMLE for multiple linear regression. Suppose the cases* $(Y_i, \boldsymbol{x}_i^T)^T$ *are iid from some distribution. Let* $\boldsymbol{w}_i$ *be the standardized predictors and assume* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w},Y} \xrightarrow{P}$

$\boldsymbol{\Sigma}_{\boldsymbol{u},Y}$ *and* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{u}}$ *where the* $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$ *are nonsingular for large enough* n *and* $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ *is nonsingular.*

$$a)\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{w}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w},Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{w}, Y) \xrightarrow{P} \boldsymbol{\Sigma}_{\boldsymbol{u},Y} =$$

$$\boldsymbol{\eta}_{OPLS}(\boldsymbol{u}, Y) = \boldsymbol{\beta}_{MMLE} = \boldsymbol{\Sigma}_{\boldsymbol{u}}[\boldsymbol{\Sigma}_{\boldsymbol{u}}]^{-1}\boldsymbol{\Sigma}_{\boldsymbol{u},Y} = \boldsymbol{\Sigma}_{\boldsymbol{u}}\boldsymbol{\beta}_{OLS}(\boldsymbol{u}, Y).$$

b) Let $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OLS}(\boldsymbol{u}, Y)$*. Then* $\boldsymbol{\beta}_{MMLE} = \boldsymbol{\Sigma}_{\boldsymbol{u}}\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OLS}$ *if* $\boldsymbol{\beta}_{OLS} = \boldsymbol{0}$ *or if* $\boldsymbol{\beta}_{OLS}$ *is an eigenvector of* $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ *with eigenvalue = 1.*

The oracle property for model selection, including variable selection, is $P(I_{min} = S) \to 1$ as $n \to \infty$ for model (8). For this property to hold, S needs to be one of the subsets considered by the model selection method with probability going to 1 as $n \to \infty$. For fixed p and "fast" estimators such as lasso and forward selection, the oracle property tends to hold if the predictors are nearly orthogonal. See Wieczorek and Lei (2022) for references. The MMLE can be used for variable selection with OLS by taking the k predictors with the largest $|\hat{\beta}_{j,MMLE}|$. The oracle property for the MMLE tends not to hold for correlated predictors by Theorem 3.2. MMLE variable selection often gives a useful submodel since predictors that satisfy a marginal regression model with the response Y (such as SLR) will often satisfy a regression model with the response Y (such as multiple linear regression).

If $\boldsymbol{\eta} = \boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ and the cases are iid, then inference for the single index model can be done using Theorem 2.4 and Section 2.3.

CHAPTER 4

SINGLE INDEX MODELS

The distribution of $Y|\boldsymbol{\eta}^T\boldsymbol{x}$ follows a single index model

$$Y|\boldsymbol{\eta}^T\boldsymbol{x} = Y = m(\boldsymbol{\eta}^T\boldsymbol{x}) + e$$

where $E(Y|\boldsymbol{\eta}^T\boldsymbol{x}) = m(\boldsymbol{\eta}^T\boldsymbol{x})$, $V(Y|\boldsymbol{\eta}^T\boldsymbol{x}) = v(\boldsymbol{\eta}^T\boldsymbol{x})$, and $e = Y - m(\boldsymbol{\eta}^T\boldsymbol{x})$. Note that the error variance may not be constant. The model is called a single index model since m depends on a single linear combination $\boldsymbol{\eta}^T\boldsymbol{x}$. A multi-index model would use $m(\boldsymbol{\eta}_1^T\boldsymbol{x}, ..., \boldsymbol{\eta}_k^T\boldsymbol{x})$ where $k > 1$.

If $\boldsymbol{\eta} = \boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ and the cases are iid, then inference for the single index model can be done using Theorem 2.4 and Section 2.3. When the cases are iid, the OPLS single index model estimators can have considerable resistance to overfitting, underfitting, heterogeneity, measurement error, highly correlated predictors, and the number of predictors.

If $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ is a good estimator of $\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$, which can occur if $n \geq 10p$, then the OPLS single index model can be visualized with a response plot of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T\boldsymbol{x}$ versus Y on the vertical axis with a scatterplot smoother added as a visual aid. If the variability about the scatterplot smoother is less than that about any horizontal line, then the model may be useful compared to simply doing inference on the $Y_1, ..., Y_n$ without any predictors.

If $Y|\boldsymbol{x} = m(\alpha + \boldsymbol{\beta}^T\boldsymbol{x}) + e$ and if the predictors $\boldsymbol{x}_i$ are iid from a large class of elliptically contoured distributions, then Li and Duan (1989) and Chen and Li (1998) showed that, under regularity conditions, $\boldsymbol{\beta}_{OLS} = c\boldsymbol{\beta}$. Hence $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = c\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}$. Thus $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = d\boldsymbol{\beta}$ if $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \tau^2\boldsymbol{I}_p$ for some constant $\tau^2 > 0$. If $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$ in this case, then $\beta_i = 0$ implies that $Cov(x_i, Y) = 0$.

The constant c is typically nonzero unless m has a lot of symmetry about the distribution of $\alpha + \boldsymbol{\beta}^T \boldsymbol{x}$. Chang and Olive (2010) considered OLS tests for these models. Simulation with $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ can be difficult if the population values of c and d are unknown.

CHAPTER 5

OUTLIER DIAGNOSTICS

Assume the cases $\boldsymbol{w}_i = (Y_i, \boldsymbol{x}_i^T)^T$ are iid. A simple method to get an outlier resistant estimator $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ is to compute an outlier resistant dispersion or covariance estimator

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} = \begin{pmatrix} \hat{\Sigma}_Y & \hat{\boldsymbol{\Sigma}}_{Y\boldsymbol{x}} \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} & \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \end{pmatrix}.$$

The function `rcovxy` uses the Olive (2017) RMVN and covmb2 estimators for $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}$. The RMVN estimator has much greater outlier resistance than the Olive (2004) MBA estimator. Also see Zhang, Olive, and Ye (2012). The covmb2 estimator can be computed in both low and high dimensions.

Another method to get an outlier resistant estimator $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ is to use the following identity. If X and Y are random variables, then

$$\mathrm{Cov(X,Y) = [Var(X+Y) - Var(X-Y)]/4.}$$

Then replace $Var(W)$ by $[\hat{\sigma}(W)]^2$ where $\hat{\sigma}(W)$ is a robust estimator of scale or standard deviation and $W = X + Y$ or $W = X - Y$. We used $\hat{\sigma}(W) = 1.483MAD(W)$ where $MAD(W) = MAD(n) = MAD(W_1, ..., W_n)$. Hence

$$\widehat{Cov}(X,Y) = [[1.483MAD(X+Y)]^2 - [1.483MAD(X-Y)]^2]/4. \tag{19}$$

This technique has been used to obtain robust dispersion matrices. See Maronna and Zamar (2002) and Mehrota (1995).

Some notation is needed to define $MAD(n)$. If the data set $Y_1, ..., Y_n$ is arranged in ascending order from smallest to largest and written as $Y_{(1)} \leq \cdots \leq Y_{(n)}$, then $Y_{(i)}$ is the ith order statistic and the $Y_{(i)}$'s are called the *order statistics*. If the data $Y_1 = 1, Y_2 = 4, Y_3 = 2, Y_4 = 5,$ and $Y_5 = 3$, then $\overline{Y} = 3$, $Y_{(i)} = i$ for $i = 1, ..., 5$ and MED$(n) = 3$ where the sample size $n = 5$. The *sample median*

$$\text{MED}(n) = Y_{((n+1)/2)} \text{ if n is odd,} \tag{20}$$

$$\text{MED}(n) = \frac{Y_{(n/2)} + Y_{((n/2)+1)}}{2} \text{ if n is even.}$$

The notation MED$(n) =$ MED$(n, Y_i) =$ MED$(Y_1, ..., Y_n)$ will also be used.

The *sample median absolute deviation* is

$$\text{MAD}(n) = \text{MED}(|Y_i - \text{MED}(n)|,\ i = 1, \ldots, n). \tag{21}$$

Since MAD$(n) =$ MAD(n, Y_i) is the median of n distances, at least half of the observations are within a distance MAD(n) of MED(n) and at least half of the observations are a distance of MAD(n) or more away from MED(n). Like the standard deviation, MAD(n) is a measure of spread.

Example 1. Let the data be $1, 2, 3, 4, 5, 6, 7, 8, 9$. Then MED$(n) = 5$ and MAD$(n) = 2 =$ MED$\{0, 1, 1, 2, 2, 3, 3, 4, 4\}$.

Then the outlier resistant estimator uses Equation (19) with

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \begin{pmatrix} \widehat{Cov}(X_1, Y) \\ \vdots \\ \widehat{Cov}(X_p, Y) \end{pmatrix}.$$

CHAPTER 6

EXAMPLES AND SIMULATIONS

Example. This example was used by Olive and Zhang (2024). 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 $W = \boldsymbol{x}^T \hat{\boldsymbol{\beta}}^T_{MMLE}$ 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 similar to that done by Rajapaksha and Olive (2024). The simulation used $\psi = 0, 0.5, 1/\sqrt{p}$, and 0.9; and $k = 1, p-2$, 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 $\boldsymbol{w}_i \sim N_{p-1}(\mathbf{0}, \boldsymbol{I})$ 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 2.4 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).

To summarize the $p-1$, confidence intervals, the average length of the $p-1$ confidence intervals over 5000 runs was computed. Then the minimum, mean, and maximum of the average lengths was computed. The proportion of times each confidence interval contained its population parameter was computed. These proportions were the observed coverages of the $p-1$ confidence 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 test $H_0 : (\eta_i, \eta_j)^T = (\eta_{i0}, \eta_{j0})^T$ was also done where H_0 was true. The coverage of the test was recorded and a correction factor was not used.

For Table 1, the simulation used the function `oplsssim` with $n, p, k, etype,$ and $wtype$ as described above, and $\psi = psi$.

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

 args(oplssim)

function (n = 100, p = 4, k = 1, nruns = 100, eps = 0.1, shift = 9,

    type = 1, psi = 0, cfac = "T", indices = c(1, 2), alph = 0.05)
```

```
oplswsim(n=100,p=4,k=1,nruns=5000,etype=1,wtype=1,psi=0)

$covxy  #sample

          [,1]       [,2]      [,3]

[1,] 0.9168954 0.09018221 0.2160901

$etaopls #population Cov(x,y)

     [,1] [,2] [,3]

[1,]    1    0    0

$oplslen

[1] 0.7128272 0.5856759 0.5865421

$oplscov

[1] 0.9464 0.9628 0.9626

$lens #nin, mean, max

[1] 0.5856759 0.6283484 0.7128272

$covprop
```

```
[1] 0.9464000 1.0000000 1.0000000 1.0000000 1.0000000 0.6666667

$testcov

[1] 0.92

$up

[1] 0.98125


oplswsim(n=100,p=100,k=1,nruns=5000,etype=1,wtype=1,psi=0)

$lens

[1] 0.5845339 0.5890543 0.7142725

$covprop

[1] 0.9468000 1.0000000 1.0000000 1.0000000 1.0000000 0.7373737

$testcov

[1] 0.919

$up

[1] 0.98125
```

Two lines per run scenario are given in each table. For the first two lines in Table 1, the simulation used $n = 100$, $p = 100$, $k = 1$, $etype = 1$, and $\psi = psi = 0$. One hundred confidence intervals were made and one test. The first line summarizes the results. The minimum coverage was 0.9468. Then the coverage proportions $\geq$ 0.9, 0.92, 0.93, and 0.94 all turned out to be 1. The coverage proportion ≥ 0.96 was 0.7374. Hence for this simulation scenario, the correction factor was a bit too large. For the test, the coverage was 0.9190, and a correction factor would have helped. Tables 1-14 illustrate Theorem 2.4a). The proportion of times η_i was in the confidence interval $\hat{\eta}_i \pm t_{n-1,up} SE(\hat{\eta}_i)$ was recorded, and the test statistic

$$n(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I} - \boldsymbol{\Sigma}_{\boldsymbol{x}_I})^T (\boldsymbol{A} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} \boldsymbol{A}^T)^{-1} (\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I} - \boldsymbol{\Sigma}_{\boldsymbol{x}_I})^T \xrightarrow{D} \chi^2_2$$

where $\boldsymbol{A}$ was a $2 \times p$ matrix with $A_{11} = A_{22} = 1$ and all other entries = 0.

Table 9 illustrates Theorem 2.4a), used wtype=2 and k=99, and had more variability than most combinations of wtype and k. For the ten different error type and ψ combinations, the minimum coverage of the 99 confidence intervals for $\eta_i = Cov(X_i, Y)$ ranged from 0.922 to 0.970. Most wtype and k combinations had a smaller range of coverages. The confidence intervals used a correction factor and overcoverage wiith coverage near 0.965 was more common than 3% undercoverage that occurs in Table 9. In line 1 of Table 9, the minimum coverage of the 99 CIs was 0.9564. Hence the proportion of the 99 CIs that had observed coverage $\geq 0.9, 0.92, 0.93$ and 0.94 was 1. The proportion of CIs that had coverage ≥ 0.96 was 0.8989 (89/99 CIs). The CI average lengths were much larger for $\psi = 0.1$ than for $\psi = 0$. The test $H_0 : (\eta_i, \eta_j)^T = (\eta_{i0}, \eta_{j0})^T$ did not use a correction factor, and coverage $<$ 0.94 was rather common. The test coverage in Table 9 was worse than that for most combinations of wtype and k.

Table 1: Cov(x,Y), wtype=1, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9468	1.0	1.0	1.0	1.0	0.7374	0.9190
	len	1	0.5845	0.5891	0.7143				
100	100	0.1	0.9426	1.0	1.0	1.0	1.0	0.0	0.9194
	len	1	1.1108	1.1161	1.2891				
100	100	0	0.9488	1.0	1.0	1.0	1.0	0.9596	0.9232
	len	2	0.7951	0.8021	0.8961				
100	100	0.1	0.9462	1.0	1.0	1.0	1.0	0.0101	0.9170
	len	2	1.3471	1.3551	1.5051				
100	100	0	0.9514	1.0	1.0	1.0	1.0	0.8586	0.9278
	len	3	0.5835	0.5871	0.7124				
100	100	0.1	0.9452	1.0	1.0	1.0	1.0	0.0	0.9162
	len	3	1.1115	1.1167	1.2949				
100	100	0	0.9400	1.0	1.0	1.0	1.0	0.7778	0.9094
	len	4	0.4777	0.4810	0.6282				
100	100	0.1	0.9410	1.0	1.0	1.0	1.0	0.0	0.9166
	len	4	1.0067	1.0116	1.2038				
100	100	0	0.9668	1.0	1.0	1.0	1.0	1.0	0.9526
	len	5	1.3452	1.3573	1.4289				
100	100	0.1	0.9588	1.0	1.0	1.0	1.0	0.9798	0.9454
	len	5	2.0437	2.0618	2.1782				

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

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9540	1.0	1.0	1.0	1.0	0.7878	0.9324
	len	1	1.7563	1.7642	1.8868				
100	100	0.1	0.9516	1.0	1.0	1.0	1.0	0.4040	0.9338
	len	1	2.7072	2.7218	2.9342				
100	100	0	0.9578	1.0	1.0	1.0	1.0	0.9898	0.9384
	len	2	2.8027	2.8309	2.9481				
100	100	0.1	0.9574	1.0	1.0	1.0	1.0	0.9595	0.9406
	len	2	4.1905	4.2234	4.4600				
100	100	0	0.9570	1.0	1.0	1.0	1.0	0.9898	0.942
	len	3	1.7261	1.7380	1.8682				
100	100	0.1	0.9574	1.0	1.0	1.0	1.0	0.8585	0.9468
	len	3	2.6824	2.6975	2.8966				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.6667	0.9332
	len	4	1.0700	1.0751	1.1920				
100	100	0.1	0.9496	1.0	1.0	1.0	1.0	0.0202	0.9242
	len	4	1.7528	1.7606	1.9476				
100	100	0	0.9696	1.0	1.0	1.0	1.0	1.0	0.9630
	len	5	5.2304	5.2805	5.4419				
100	4	0.1	0.9710	1.0	1.0	1.0	1.0	1.0	0.9670
	len	5	7.6531	7.7255	7.9985				

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

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9410	1.0	1.0	1.0	1.0	0.8181	0.9128
	len	1	0.7115	0.7195	1.0012				
100	100	0.1	0.9398	1.0	1.0	1.0	0.9898	0.0	0.9012
	len	1	1.5219	1.5328	1.8982				
100	100	0	0.9524	1.0	1.0	1.0	1.0	0.9696	0.9324
	len	2	1.0343	1.0478	1.4465				
100	100	0.1	0.9474	1.0	1.0	1.0	1.0	0.0303	0.9230
	len	2	2.1355	2.1565	2.6681				
100	100	0	0.9502	1.0	1.0	1.0	1.0	0.9292	0.9188
	len	3	0.7042	.7117	0.9828				
100	100	0.1	0.9360	1.0	1.0	1.0	0.9898	0.0	0.9078
	len	3	1.4980	1.5095	1.8627				
100	100	0	0.9316	1.0	1.0	1.0	0.9898	0.7979	0.8960
	len	4	0.5313	0.5358	0.7431				
100	100	0.1	0.9336	1.0	1.0	1.0	0.7474	0.0	0.9008
	len	4	1.1658	1.1732	1.4367				
100	100	0	0.9696	1.0	1.0	1.0	1.0	1.0	0.9576
	len	5	1.8131	1.8370	2.4581				
100	100	0.1	0.9670	1.0	1.0	1.0	1.0	1.0	0.9528
	len	5	3.6615	3.7218	4.6064				

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

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9476	1.0	1.0	1.0	1.0	0.7272	0.9142
	len	1	0.4135	0.4168	0.5810				
100	100	0	0.9456	1.0	1.0	1.0	1.0	0.7373	0.9126
	len	2	0.4127	0.4162	0.5804				
100	100	0	0.9442	1.0	1.0	1.0	1.0	0.7878	0.9114
	len	3	0.4140	0.4171	0.5826				
100	100	0	0.9446	1.0	1.0	1.0	1.0	0.7778	0.9332
	len	4	0.4131	0.4170	0.5814				
100	100	0	0.9492	1.0	1.0	1.0	1.0	0.8282	0.9216
	len	5	0.4131	0.4175	0.5823				

Table 5: Cov(x,Y), wtype=5, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.7676	0.9404
	len	1	25.5793	25.6706	25.7814				
100	100	0.1	0.9600	1.0	1.0	1.0	1.0	1.0	0.9504
	len	1	31.6712	31.7842	31.9120				
100	100	0	0.9596	1.0	1.0	1.0	1.0	0.9898	0.9516
	len	2	41.4167	41.7714	42.0956				
100	100	0.1	0.9626	1.0	1.0	1.0	1.0	1.0	0.9480
	len	2	50.2093	50.9155	51.3319				
100	100	0	0.9588	1.0	1.0	1.0	1.0	0.9090	0.9448
	len	3	25.0352	25.2252	25.3880				
100	100	0.1	0.9626	1.0	1.0	1.0	1.0	1.0	0.9508
	len	3	30.7604	30.9600	31.1437				
100	100	0	0.9524	1.0	1.0	1.0	1.0	0.6969	0.8960
	len	4	14.8623	14.9086	14.9658				
100	100	0.1	0.9584	1.0	1.0	1.0	1.0	0.9590	0.9480
	len	4	18.4957	18.5544	18.6331				
100	100	0	0.9704	1.0	1.0	1.0	1.0	1.0	0.9670
	len	5	78.0315	79.0563	80.0812				
100	100	0.1	0.9710	1.0	1.0	1.0	1.0	1.0	0.9682
	len	5	93.3959	94.3886	95.3620				

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

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9416	1.0	1.0	1.0	1.0	0.8080	0.9096
	len	1	0.4699	0.4739	0.6227				
100	100	0.1	0.9334	1.0	1.0	1.0	0.4747	0.0	0.9014
	len	1	1.2369	1.2454	1.3987				
100	100	0	0.9494	1.0	1.0	1.0	1.0	0.8888	0.9176
	len	2	0.5554	0.5610	0.6942				
100	100	0.1	0.9448	1.0	1.0	1.0	1.0	0.0	0.9218
	len	2	1.6019	1.6122	1.7471				
100	100	0	0.9440	1.0	1.0	1.0	1.0	0.6868	0.9086
	len	3	0.4707	0.4739	0.6246				
100	100	0.1	0.9378	1.0	1.0	1.0	0.8788	0.6464	0.9042
	len	3	1.2358	1.2417	1.4033				
100	100	0	0.9458	1.0	1.0	1.0	1.0	0.7878	0.9128
	len	4	0.4340	0.4372	0.5977				
100	100	0.1	0.9360	1.0	1.0	1.0	0.9393	0.0	0.9088
	len	4	1.0581	1.0636	1.2480				
100	100	0	0.9484	1.0	1.0	1.0	1.0	0.9898	0.9318
	len	5	0.8097	0.8192	0.9165				
100	100	0.1	0.9598	1.0	1.0	1.0	1.0	1.0	0.9370
	len	5	2.5359	2.5600	2.6475				

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

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9430	1.0	1.0	1.0	1.0	0.7878	0.9134
	len	1	0.4907	0.4941	0.6380				
100	100	0.1	0.9380	1.0	1.0	1.0	0.9898	0.0	0.9092
	len	1	0.9913	0.9972	1.1877				
100	100	0	0.9434	1.0	1.0	1.0	1.0	0.9292	0.9184
	len	2	0.5999	0.6052	0.7298				
100	100	0.1	0.9422	1.0	1.0	1.0	1.0	0.0	0.9154
	len	2	1.0775	1.0841	1.2653				
100	100	0	0.9406	1.0	1.0	1.0	1.0	0.7676	0.9130
	len	3	0.4901	0.4934	0.6395				
100	100	0.1	0.9430	1.0	1.0	1.0	1.0	0.0	0.9114
	len	3	0.9984	1.0040	1.1961				
100	100	0	0.9454	1.0	1.0	1.0	1.0	0.8080	0.9186
	len	4	0.4412	0.4447	0.6032				
100	100	0.1	0.9416	1.0	1.0	1.0	1.0	0.0	0.9102
	len	4	0.9632	0.9695	1.1692				
100	100	0	0.9614	1.0	1.0	1.0	1.0	1.0	0.9446
	len	5	0.9216	0.9292	1.0151				
100	100	0.1	0.9542	1.0	1.0	1.0	1.0	0.4949	0.9284
	len	5	1.3608	1.3701	1.5251				

Table 8: Cov(x,Y), wtype=1, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9524	1.0	1.0	1.0	1.0	0.7272	0.9470
	len	1	4.1706	4.1835	4.1997				
100	100	0.1	0.9382	1.0	1.0	1.0	0.9797	0.0	0.9194
	len	1	78.8311	79.0693	79.2988				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.7070	0.9382
	len	2	4.2071	4.2216	4.2372				
100	100	0.1	0.9394	1.0	1.0	1.0	0.9898	0.0	0.9130
	len	2	78.6014	78.8947	79.1805				
100	100	0	0.9532	1.0	1.0	1.0	1.0	0.6565	0.9342
	len	3	4.1505	4.1680	04.1848				
100	100	0.1	0.9410	1.0	1.0	1.0	1.0	0.0	0.9128
	len	3	78.4876	78.7667	79.0276				
100	100	0	0.9528	1.0	1.0	1.0	1.0	0.7575	0.9454
	len	4	4.1470	4.1621	4.1768				
100	100	0.1	0.9362	1.0	1.0	1.0	0.9191	0.0	0.9114
	len	4	78.3503	78.6564	78.9038				
100	100	0	0.9548	1.0	1.0	1.0	1.0	0.7676	0.9396
	len	5	4.3468	4.3664	4.3937				
100	100	0.1	0.9388	1.0	1.0	1.0	0.9797	0.0	0.9196
	len	5	78.5803	78.9242	79.1622				

Table 9: Cov(x,Y), wtype=2, 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	0.8989	0.9500
	len	1	5.9909	6.0312	6.0628				
100	100	0.1	0.9294	1.0	1.0	1.0	0.9294	0.0	0.8996
	len	1	118.7273	119.1710	119.6229				
100	100	0	0.9592	1.0	1.0	1.0	1.0	0.9696	0.9444
	len	2	8.153728	8.2505	8.3430				
100	100	0.1	0.9400	1.0	1.0	1.0	1.0	0.0	0.9122
	len	2	162.7834	164.0093	165.4210				
100	100	0	0.9566	1.0	1.0	1.0	1.0	0.8383	0.9488
	len	3	5.9832	6.0169	6.0547				
100	100	0.1	0.9336	1.0	1.0	1.0	0.6060	0.0	0.9058
	len	3	116.9208	117.5406	118.1560				
100	100	0	0.9566	1.0	1.0	1.0	1.0	0.8181	0.9406
	len	4	4.8333	4.8636	4.8995				
100	100	0.1	0.9224	1.0	1.0	0.7878	0.0	0.0	0.8940
	len	4	93.5389	93.9788	94.40372				
100	100	0	0.9702	1.0	1.0	1.0	1.0	1.0	0.9426
	len	5	13.6393	13.8106	14.1451				
100	100	0.1	0.9606	1.0	1.0	1.0	1.0	1.0	0.9426
	len	5	270.4977	272.8313	275.7340				

Table 10: Cov(x,Y), wtype=3, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9540	1.0	1.0	1.0	1.0	0.6868	0.9408
	len	1	4.1677	4.1883	4.2386				
100	100	0.1	0.9402	1.0	1.0	1.0	1.0	0.0	0.9168
	len	1	78.5653	78.8622	79.1832				
100	100	0	0.9554	1.0	1.0	1.0	1.0	0.6868	0.9414
	len	2	4.2571	4.2790	4.3870				
100	100	0.1	0.9360	1.0	1.0	1.0	0.9797	0.0	0.9110
	len	2	78.46492	78.8270	79.0445				
100	100	0	0.9512	1.0	1.0	1.0	1.0	0.7575	0.9472
	len	3	4.1801	4.1968	4.2446				
100	100	0.1	0.9364	1.0	1.0	1.0	0.9494	0.0	0.9112
	len	3	78.6214	78.9302	79.1963				
100	100	0	0.9546	1.0	1.0	1.0	1.0	0.6363	0.9450
	len	4	4.1486	4.1697	4.1959				
100	100	0.1	0.9366	1.0	1.0	1.0	0.9292	0.0	0.9130
	len	4	78.7364	78.9669	79.2669				
100	100	0	0.9558	1.0	1.0	1.0	1.0	0.7575	0.9414
	len	5	4.5517	4.5770	4.9071				
100	100	0.1	0.9342	1.0	1.0	1.0	0.9595	0.0	0.9144
	len	5	78.7625	78.9909	79.1899				

Table 11: Cov(x,Y), wtype=4, 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	0.7676	0.9398
	len	1	4.1466	4.1640	4.1790				
100	100	0	0.9540	1.0	1.0	1.0	1.0	0.7070	0.9432
	len	2	4.1328	4.1482	4.1704				
100	100	0	0.9534	1.0	1.0	1.0	1.0	0.7070	0.9434
	len	3	4.1328	4.1482	4.1704				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.6363	0.9418
	len	4	4.1361	4.1537	4.1723				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.6767	0.9430
	len	5	4.1453	4.1647	4.1797				

Table 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	0.7878	0.9446
	len	1	25.9105	26.0309	26.1530				
100	100	0.1	0.9370	1.0	1.0	1.0	0.8585	0.0	0.9142
	len	1	84.0396	84.3577	84.6274				
100	100	0	0.9594	1.0	1.0	1.0	1.0	0.9898	0.9532
	len	2	41.7302	42.0439	42.3989				
100	100	0.1	0.9378	1.0	1.0	1.0	0.9494	0.0	0.9120
	len	2	93.6379	93.9699	94.3939				
100	100	0	0.9592	1.0	1.0	1.0	1.0	0.9696	0.9448
	len	3	25.4353	5.5958	25.7366				
100	100	0.1	0.9372	1.0	1.0	1.0	0.8383	0.0	0.9100
	len	3	83.9672	84.2346	84.5139				
100	100	0	0.9526	1.0	1.0	1.0	1.0	0.7474	0.9486
	len	4	15.4105	15.4812	15.52944				
100	100	0.1	0.9380	1.0	1.0	1.0	0.9292	0.0	0.9116
	len	4	80.3214	80.6125	80.8487				
100	100	0	0.9690	1.0	1.0	1.0	1.0	1.0	0.9666
	len	5	78.6979	79.5798	80.5326				
100	100	0.1	0.9518	1.0	1.0	1.0	1.0	0.0909	0.9380
	len	5	123.9543	124.7747	125.4251				

Table 13: Cov(x,Y), wtype=6, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.6969	0.9408
	len	1	4.1424	4.1564	4.1746				
100	100	0.1	0.9368	1.0	1.0	1.0	0.9494	0.0	0.9082
	len	1	78.2762	78.4933	78.7964				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.6969	0.9406
	len	2	4.1590	4.1761	4.1906				
100	100	0.1	0.9392	1.0	1.0	1.0	0.9898	0.0	0.9104
	len	2	78.5195	78.8016	79.1488				
100	100	0	0.9542	1.0	1.0	1.0	1.0	0.6262	0.9340
	len	3	4.1434	4.1608	4.1744				
100	100	0.1	0.9348	1.0	1.0	1.0	0.6767	0.0	0.9086
	len	3	78.2621	78.5305	78.8025				
100	100	0	0.9548	1.0	1.0	1.0	1.0	0.6767	0.9396
	len	4	4.1383	4.1629	4.1805				
100	100	0.1	0.9374	1.0	1.0	1.0	0.9696	0.0	0.9072
	len	4	78.2358	78.5532	78.8016				
100	100	0	0.9552	1.0	1.0	1.0	1.0	0.7272	0.9492
	len	5	4.2113	4.2281	4.2495				
100	100	0.1	0.9378	1.0	1.0	1.0	0.9696	0.0	0.9164
	len	5	78.4399	78.6697	78.9044				

Table 14: Cov(x,Y), wtype=7, 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	0.6868	0.9416
	len	1	4.1431	4.1635	4.1784				
100	100	0.1	0.9362	1.0	1.0	1.0	0.8889	0.0	0.9094
	len	1	78.2762	78.4933	78.7964				
100	100	0	0.9542	1.0	1.0	1.0	1.0	0.6464	0.9420
	len	2	4.1596	4.1832	4.2023				
100	100	0.1	0.9352	1.0	1.0	1.0	0.7575	0.0	0.9188
	len	2	78.2773	78.5771	78.8371				
100	100	0	0.9536	1.0	1.0	1.0	1.0	0.6667	0.9434
	len	3	4.1404	4.1617	4.1850				
100	100	0.1	0.9384	1.0	1.0	1.0	0.9696	0.0	0.9144
	len	3	78.0388	78.3908	78.7165				
100	100	0	0.9556	1.0	1.0	1.0	1.0	0.6667	0.9374
	len	4	4.1301	4.1526	4.1695				
100	100	0.1	0.9394	1.0	1.0	1.0	0.9898	0.0	0.9078
	len	4	78.1104	78.4263	78.6769				
100	100	0	0.9564	1.0	1.0	1.0	1.0	0.7070	0.9424
	len	5	4.2321	4.2479	4.2647				
100	100	0.1	0.9366	1.0	1.0	1.0	0.9797	0.0	0.9184
	len	5	78.3324	78.6694	78.9753				

6.1 Simulation with Theorem 2.4c)

The simulation for Theorem 2.4c) is similar, but now H_0 was often false, and using $\hat{\lambda}$ to estimate λ sometimes caused problems in high dimensions. Now the proportion of times $\lambda\eta_i = \beta_{i,OPLS}$ was in the interval $\hat{\lambda}\hat{\eta}_i \pm \hat{\lambda} t_{n-1,up} SE(\hat{\eta}_i)$ was recorded, but the interval is not a confidence interval unless $\beta_{i,OPLS} = 0$. The test statistic

$$n\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T \boldsymbol{A}^T(\hat{\lambda}^2 \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = n\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T \boldsymbol{A}^T(\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}}\boldsymbol{A}^T)^{-1}\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{D} \chi^2_2$$

provided $\boldsymbol{A}\boldsymbol{\beta}_{OPLS} = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$. With $k = 1$ and $\psi = 0$, all of the $\beta_{i,OPLS} = 0$ except $\beta_{1,OPLS}$. We also tested whether $H_0 : (\beta_{98}, \beta_{99})^T = \boldsymbol{0}$, and H_0 was true with $k = 1$ and $\psi = 0$. In Table 15, the first two lines had $k = 1$ and $\psi = 0$. Then $\hat{\boldsymbol{\beta}}_{1,OPLS} = 1$ was never in its interval $\hat{\lambda}\hat{\eta}_1 \pm \hat{\lambda} t_{n-1,up} SE(\hat{\eta}_1)$ because $\lambda = 1$ but $\hat{\lambda} < 0.5$ was common. Hence the minimum coverage was 0.0. The other 98 intervals and the test satisfied Theorem 2.4c), and the coverages were good. With $\psi = 0.1$, $\hat{\lambda}$ was often a good estimator of λ, but $(\beta_{98}, \boldsymbol{\beta}_{99}) = (1, 1) \neq (0, 0)$, so testcov was near 0.

```
source("http://parker.ad.siu.edu/Olive/slpack.txt")
args(oplssim2)
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)


oplssim2(n=100,p=100,k=1,nruns=5000,etype=1,wtype=1,psi=0,indices=c(98,99))
$lens
[1] 0.2539468 0.2560456 0.3097610
```

```
$covprop
[1] 0.0000000 0.9898990 0.9898990 0.9898990 0.9898990 0.8484848
$testcov
[1] 0.947
$up
[1] 0.98125
$lambda
[1] 1
$lamhat
0.4956325

oplssim2(n=100,p=100,k=1,nruns=5000,etype=1,wtype=1,psi=0.1,
indices=c(98,99))
$lens
[1] 0.009694085 0.009740023 0.011268022
$covprop
[1] 0.9696 1.0000 1.0000 1.0000 1.0000 1.0000
$testcov
[1] 4e-04
$up
[1] 0.98125
$lambda
[1] 0.008613624
```

```
$lamhat
0.008469176
```

Tables 17 and 28 are used to illustrate Theorem 2.4c), and to show that $\hat{\lambda}$ can be a poor estimator of λ in high dimensions. Now the proportion of times $\lambda\eta_i = \beta_{i,OPLS}$ was in the interval $\hat{\lambda}\hat{\eta}_i \pm \hat{\lambda} t_{n-1,up} SE(\hat{\eta}_i) = [\hat{\lambda} L_{in}, \hat{\lambda} U_{in}]$ was recorded, where $[L_{in}, U_{in}]$ is the large sample 95% CI for η_i. If $\eta_i \neq 0$, then the coverage of this interval tends to be low if $\hat{\lambda}$ underestimates λ, and high if $\hat{\lambda}$ overestimates λ. If $\eta_i = 0 = \beta_{i,OPLS}$ and $\hat{\lambda} > 0$, then the interval gives a large sample test for $H_0 : \beta_{i,OPLS} = 0$ since $0 \in [L_{in}, U_{in}]$ if and only if $0 \in [\hat{\lambda} L_{in}, \hat{\lambda} U_{in}]$. Hence Theorem 2.4c) can be used to test $H_0 : \beta_{i,OPLS} = 0$ in low or high dimensions even if $\hat{\lambda} > 0$ is not a good estimator of λ.

For testing $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} = (\beta_{98,OPLS}, \beta_{99,OPLS})^T = \boldsymbol{0}$, the test statistic

$$n\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T \boldsymbol{A}^T (\hat{\lambda}^2 \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} \boldsymbol{A}^T)^{-1} \boldsymbol{A}\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = n\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T \boldsymbol{A}^T (\boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{w}} \boldsymbol{A}^T)^{-1} \boldsymbol{A}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \xrightarrow{D} \chi^2_2$$

provided $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} = \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$ is true. In the simulation, H_0 is true if $k = 1$ and $\psi = 0$.

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 $\lambda = 1/116.64 = 0.008573$, $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OPLS} = \boldsymbol{1}$, and $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = 116.64\ \boldsymbol{1}$. 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 $p = 100$.)

Table 17 used $k = 1$, and the minimum coverage corresponding to $\beta_{1,OPLS}$ tended to be much smaller than 0.95 for $\psi = 0$ since $\hat{\lambda}$ underestimated λ. When $\psi = 0.1$ the coverages for $\beta_{i,OPLS}$ tended to be a bit high since $\hat{\lambda}$ tended to be near or greater than λ. When $\psi = 0.1$, $H_0 : (\beta_{98,OPLS}, \beta_{99,OPLS})^T = (0, 0)^T$ is false. Then low testcov indicates good power for the test. Sometimes n much larger than 100 was needed to make testcov near 0.

Table 28 used $k = 99$. For $\psi = 0$ the coverage for $\beta_{1,OPLS}$ tended to be low since $\hat{\lambda}$ underestimated λ. The other coverages, including testcov, tended to be low. When $\psi = 0.1$ the coverages for $\beta_{i,OPLS}$ tended to be a bit high since $\hat{\lambda}$ tended to be near or greater than λ. When $\psi = 0.1$, $H_0 : (\beta_{98,OPLS}, \beta_{99,OPLS})^T = (0, 0)^T$ is false. Then low testcov indicates good power for the test. Sometimes n much larger than 100 was needed to make testcov near 0.

Table 15: OPLS, wtype=1, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.8485	0.9470
	len	1	0.2539	0.2560	0.3097				
100	100	0.1	0.9696	1.0	1.0	1.0	1.0	1.0	0.0004
	len	1	0.0097	0.0097	0.0113				
100	100	0	0.0128	0.9899	0.9899	0.9899	0.9899	0.9494	0.9492
	len	2	0.3642	0.3676	0.4110				
100	100	0.1	0.9670	1.0	1.0	1.0	1.0	1.0	0.0254
	len	2	0.0121	0.0123	0.0136				
100	100	0	0.0002	0.9899	0.9899	0.9899	0.9899	0.8383	0.9468
	len	3	0.2531	0.2549	0.3077				
100	100	0.1	0.9660	1.0	1.0	1.0	1.0	1.0	0.0006
	len	3	0.0096	0.0097	0.0112				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7272	0.9424
	len	4	0.1991	0.2006	0.2615				
100	100	0.1	0.9662	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.0086	0.0087	0.0103				
100	100	0	0.2016	0.9899	0.9899	0.9899	0.9899	0.9899	0.9620
	len	5	0.6583	0.6646	0.6984				
100	100	0.1	0.9758	1.0	1.0	1.0	1.0	1.0	0.2682
	len	5	0.0248	0.0256	0.0265				

Table 16: OPLS, wtype=2, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.4008	0.9899	0.9899	0.9899	0.9899	0.8383	0.9472
	len	1	0.8663	0.8707	0.9288				
100	100	0.1	0.9806	1.0	1.0	1.0	1.0	1.0	0.5688
	len	1	0.0376	0.0381	0.0410				
100	100	0	0.6730	0.9899	0.9899	0.9899	0.9899	0.9696	0.9518
	len	2	1.3925	1.4100	1.4842				
100	100	0.1	0.9750	1.0	1.0	1.0	1.0	1.0	0.7628
	len	2	0.1400	0.1439	0.1520				
100	100	0	0.3930	0.9899	0.9899	0.9899	0.9899	0.9292	0.9470
	len	3	0.8471	0.8558	0.9175				
100	100	0.1	0.9706	1.0	1.0	1.0	1.0	1.0	0.4848
	len	3	0.0513	0.0526	0.0577				
100	100	0	0.0722	0.9899	0.9899	0.9899	0.9899	0.7676	0.9376
	len	4	0.5104	0.5130	0.5657				
100	100	0.1	0.9722	1.0	1.0	1.0	1.0	1.0	0.1600
	len	4	0.0159	0.0160	0.0177				
100	100	0	0.8374	0.9899	0.9899	0.9899	0.9899	0.9899	0.9688
	len	5	2.6211	2.6501	2.7297				
100	100	0.1	0.9654	1.0	1.0	1.0	1.0	1.0	0.8654
	len	5	0.4329	0.4425	0.4626				

Table 17: OPLS, wtype=3, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.0182	0.9899	0.9899	0.9899	0.9899	0.9191	0.9486
	len	1	0.3209	0.3239	0.4486				
100	100	0.1	0.9664	1.0	1.0	1.0	1.0	1.0	0.0528
	len	1	0.0133	0.0134	0.0166				
100	100	0	0.1768	0.9899	0.9899	0.9899	0.9899	0.9797	0.9474
	len	2	0.4889	0.4956	0.6865				
100	100	0.1	0.9804	1.0	1.0	1.0	1.0	1.0	0.3084
	len	2	0.0255	0.0267	0.0338				
100	100	0	0.0260	0.9899	0.9899	0.9899	0.9899	0.9797	0.9486
	len	3	0.3185	0.3220	0.4445				
100	100	0.1	0.9672	1.0	1.0	1.0	1.0	1.0	0.0538
	len	3	0.0132	0.0133	0.0165				
100	100	0	0.0002	0.9899	0.9899	0.9899	0.9899	0.8181	0.9424
	len	4	0.2262	0.2282	0.3158				
100	100	0.1	0.9524	1.0	1.0	1.0	1.0	0.9899	0
	len	4	0.0101	0.0102	0.0125				
100	100	0	0.4982	0.9899	0.9899	0.9899	0.9899	0.9899	0.9664
	len	5	0.9030	0.9180	1.2506				
100	100	0.1	0.9852	1.0	1.0	1.0	1.0	1.0	0.6048
	len	5	0.1023	0.1052	0.1345				

Table 18: OPLS, wtype=4, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7171	0.9450
	len	1	0.1672	0.1687	0.2346				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7474	0.9532
	len	2	0.1672	0.1686	0.2349				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7575	0.9424
	len	3	0.3185	0.3220	0.4445				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7373	0.9400
	len	4	0.1672	0.1686	0.2343				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.8080	0.9426
	len	5	0.1675	0.2356	0.2356				

Table 19: OPLS, wtype=5, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.9532	1.0	1.0	1.0	1.0	0.8282	0.9416
	len	1	13.2162	13.2678	13.3680				
100	100	0.1	0.9602	1.0	1.0	1.0	1.0	1.0	0.0528
	len	1	2.2759	2.3004	2.3219				
100	100	0	0.9594	1.0	1.0	1.0	1.0	0.9797	0.9580
	len	2	21.3396	21.5434	21.7143				
100	100	0.1	0.9648	1.0	1.0	1.0	1.0	1.0	0.9522
	len	2	3.6175	3.7116	3.7654				
100	100	0	0.9588	1.0	1.0	1.0	1.0	0.9494	0.9526
	len	3	12.9127	13.0112	13.1067				
100	100	0.1	0.9630	1.0	1.0	1.0	1.0	1.0	0.9510
	len	3	2.1969	2.2246	2.2432				
100	100	0	0.9540	1.0	1.0	1.0	1.0	0.6667	0.9450
	len	4	7.6307	7.6528	7.6855				
100	100	0.1	0.9524	1.0	1.0	1.0	1.0	1.0	0.9418
	len	4	1.3193	1.3297	1.3400				
100	100	0	0.9706	1.0	1.0	1.0	1.0	1.0	0.9634
	len	5	40.4202	40.8316	41.2634				
100	100	0.1	0.9710	1.0	1.0	1.0	1.0	1.0	0.9722
	len	5	7.0543	7.1787	7.3326				

Table 20: OPLS, wtype=6, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7778	0.9428
	len	1	0.11955	0.1968	0.2582				
100	100	0.1	0.9592	1.0	1.0	1.0	1.0	0.9899	0.0528
	len	1	0.0107	0.0108	0.0122				
100	100	0	0.0008	0.9899	0.9899	0.9899	0.9899	0.8484	0.9442
	len	2	0.2397	0.2423	0.2977				
100	100	0.1	0.9668	1.0	1.0	1.0	1.0	1.0	0.1018
	len	2	0.0153	0.0157	0.0167				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7979	0.9440
	len	3	0.1959	0.1968	0.2584				
100	100	0.1	0.9642	1.0	1.0	1.0	1.0	1.0	0.0098
	len	3	0.0107	0.0108	0.0122				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.8686	0.9498
	len	4	0.1775	0.1790	0.2437				
100	100	0.1	0.9628	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.0091	0.0092	0.0108				
100	100	0	0.0174	0.9899	0.9899	0.9899	0.9899	0.9899	0.9552
	len	5	0.3732	0.3769	0.4206				
100	100	0.1	0.9842	1.0	1.0	1.0	1.0	1.0	0.3590
	len	5	0.0429	0.0444	0.0458				

Table 21: OPLS, wtype=7, k=1

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7171	0.9418
	len	1	0.2058	0.2072	0.2664				
100	100	0.1	0.9648	1.0	1.0	1.0	1.0	1.0	0
	len	1	0.0086	0.0087	0.0103				
100	100	0	0.0022	0.9899	0.9899	0.9899	0.9899	0.8080	0.9406
	len	2	0.2628	0.2651	0.3180				
100	100	0.1	0.9666	1.0	1.0	1.0	1.0	1.0	0.0014
	len	2	0.0094	0.0095	0.0111				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7676	0.9478
	len	3	0.2056	0.2070	0.2675				
100	100	0.1	0.9630	1.0	1.0	1.0	1.0	1.0	0
	len	3	0.0086	0.0087	0.0103				
100	100	0	0	0.9899	0.9899	0.9899	0.9899	0.7778	0.9438
	len	4	0.1809	0.1824	0.2471				
100	100	0.1	0.9624	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.0083	0.0084	0.0101				
100	100	0	0.0468	0.9899	0.9899	0.9899	0.9899	0.9899	0.9544
	len	5	0.4336	0.4383	0.4765				
100	100	0.1	0.9712	1.0	1.0	1.0	1.0	1.0	0.0478
	len	5	0.0122	0.0123	0.0136				

Table 22: OPLS, wtype=1, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.6918	0	0	0	0	0	0.7796
	len	1	1.6817	1.6878	1.6941				
100	100	0.1	0.9980	1.0	1.0	1.0	1.0	1.0	0
	len	1	0.6693	0.6717	0.6742				
100	100	0	0.6992	0	0	0	0	0	0.7784
	len	2	1.7040	1.7109	0.7192				
100	100	0.1	0.9978	1.0	1.0	1.0	1.0	1.0	0
	len	2	0.6697	0.6721	0.6739				
100	100	0	0.6900	0	0	0	0	0	0.7672
	len	3	1.6838	1.6916	1.6994				
100	100	0.1	0.9978	1.0	1.0	1.0	1.0	1.0	0
	len	3	0.6709	0.6735	0.6755				
100	100	0	0.6796	0	0	0	0	0	0.7682
	len	4	1.6798	1.6864	1.6953				
100	100	0.1	0.9976	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.6691	0.6719	0.6741				
100	100	0	0.7184	0	0	0	0	0	0.7822
	len	5	1.7849	1.7919	1.7996				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	5	0.6692	0.6719	0.6741				

Table 23: OPLS, wtype=2, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.8326	0	0	0	0	0	0.8724
	len	1	2.6356	2.6496	2.6663				
100	100	0.1	0.9674	1.0	1.0	1.0	1.0	1.0	0.0028
	len	1	1.0184	1.0227	1.0300				
100	100	0	0.8920	0.7171	0	0	0	0	0.9160
	len	2	3.7425	3.7873	3.8356				
100	100	0.1	0.9756	1.0	1.0	1.0	1.0	1.0	0.1158
	len	2	1.6167	1.6725	1.7505				
100	100	0	0.8348	0	0	0	0	0	0.8744
	len	3	2.6217	2.6367	2.6540				
100	100	0.1	0.9722	1.0	1.0	1.0	1.0	1.0	0.0120
	len	3	1.0065	1.0121	1.0173				
100	100	0	0.7584	0	0	0	0	0	0.8312
	len	4	2.0203	2.0349	2.0478				
100	100	0.1	0.9784	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.8021	0.8051	0.8091				
100	100	0	0.9370	1.0	1.0	1.0	0.9797	0	0.9380
	len	5	6.6060	6.7327	6.8192				
100	100	0.1	0.9890	1.0	1.0	1.0	1.0	1.0	0.3752
	len	5	5.6795	5.8686	6.0980				

Table 24: OPLS, wtype=3, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.6928	0	0	0	0	0	0.7818
	len	1	1.6933	1.7005	1.7151				
100	100	0.1	0.9978	1.0	1.0	1.0	1.0	1.0	0
	len	1	0.6703	0.6727	0.6754				
100	100	0	0.7034	0	0	0	0	0	0.7840
	len	2	1.7298	1.7376	1.7800				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	2	0.6701	0.6725	0.6746				
100	100	0	0.6952	0	0	0	0	0	0.7838
	len	3	1.6939	1.7007	1.7114				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	3	0.6705	0.6724	1.6752				
100	100	0	0.6849	0	0	0	0	0	0.7714
	len	4	1.6807	1.6891	1.6954				
100	100	0.1	0.9976	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.6705	0.6725	0.6744				
100	100	0	0.7400	0	0	0	0	0	0.8004
	len	5	1.8828	1.8929	2.0324				
100	100	0.1	0.9980	1.0	1.0	1.0	1.0	1.0	0
	len	5	0.6702	0.6724	0.6743				

Table 25: OPLS, wtype=4, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.6852	0	0	0	0	0	0.7722
	len	1	1.6733	1.6805	1.6888				
100	100	0	0.6866	0	0	0	0	0	0.7756
	len	2	1.6734	1.6818	1.6910				
100	100	0	0.6874	0	0	0	0	0	0.7802
	len	3	1.6749	1.6820	1.6899				
100	100	0	0.6888	0	0	0	0	0	0.7654
	len	4	1.6791	1.6841	1.6935				
100	100	0	0.6890	0	0	0	0	0	0.7728
	len	5	1.6703	1.6794	2.6870				

Table 26: OPLS, wtype=5, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.6928	1.0	1.0	1.0	1.0	0.3030	0.7818
	len	1	13.2205	13.2783	13.3739				
100	100	0.1	0.9932	1.0	1.0	1.0	1.0	1.0	0
	len	1	0.7185	0.7212	0.7241				
100	100	0	0.9574	1.0	1.1.0	1.0	1.0	0.9595	0.9458
	len	2	21.5202	21.7361	21.9323				
100	100	0.1	0.9894	1.0	1.0	1.0	1.0	1.0	0.0022
	len	2	0.8109	0.8158	0.8210				
100	100	0	0.9548	1.0	1.0	1.0	1.0	0.6363	0.9380
	len	3	12.8699	12.9893	13.0901				
100	100	0.1	0.9936	1.0	1.0	1.0	1.0	1.0	0
	len	3	0.7196	0.7229	0.7261				
100	100	0	0.9432	1.0	1.0	1.0	1.0	0.0202	0.9290
	len	4	7.6732	7.6951	7.7268				
100	100	0.1	0.9970	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.6868	0.6896	0.6921				
100	100	0	0.9694	1.0	1.0	1.0	1.0	1.0	0.9656
	len	5	40.2310	40.6674	40.0584				
100	100	0.1	0.9014	1.0	1.0	1.0	1.0	1.0	0.0316
	len	5	1.1897	1.2413	1.3311				

Table 27: OPLS, wtype=6, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.6904	0	0	0	0	0	0.7702
	len	1	1.6783	1.6868	1.6930				
100	100	0.1	0.6690	1.0	1.0	1.0	1.0	1.0	0
	len	1	0.6690	0.6715	0.6741				
100	100	0	0.6912	0	0	0	0	0	0.7874
	len	2	1.6825	1.6903	1.6980				
100	100	0.1	0.6695	1.0	1.0	1.0	1.0	1.0	0
	len	2	0.6695	0.6720	0.6740				
100	100	0	0.6840	0	0	0	0	0	0.7708
	len	3	1.6771	1.6850	1.6907				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	3	0.6685	0.6703	0.6729				
100	100	0	0.6918	0	0	0	0	0	0.7712
	len	4	1.6756	1.6819	1.6877				
100	100	0.1	0.9978	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.6695	0.6719	0.6752				
100	100	0	0.6958	0	0	0	0	0	0.7720
	len	5	1.7047	1.7129	1.7209				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	5	0.6724	0.6743	0.6777				

Table 28: OPLS, wtype=7, k=99

n	p	psi/etype	mincov	cov90	cov92	cov93	cov94	cov96	testcov
100	100	0	0.6882	0	0	0	0	0	0.7738
	len	1	1.6805	1.6866	1.6933				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	1	0.6691	0.6718	0.6743				
100	100	0	0.6896	0	0	0	0	0	0.7702
	len	2	1.6866	1.6929	1.7019				
100	100	0.1	0.9974	1.0	1.0	1.0	1.0	1.0	0
	len	2	0.6701	0.6721	0.6743				
100	100	0	0.6880	0	0	0	0	0	0.7746
	len	3	1.6814	1.6873	1.6955				
100	100	0.1	0.9976	1.0	1.0	1.0	1.0	1.0	0
	len	3	0.6701	0.6701	0.6748				
100	100	0	0.6888	0	0	0	0	0	0.7692
	len	4	1.6782	1.6868	1.6948				
100	100	0.1	0.9968	1.0	1.0	1.0	1.0	1.0	0
	len	4	0.6701	0.6722	0.6743				
100	100	0	0.7024	0	0	0	0	0	0.7834
	len	5	1.7181	1.7266	1.7361				
100	100	0.1	0.9970	1.0	1.0	1.0	1.0	1.0	0
	len	5	0.6686	0.6709	0.6731				

CHAPTER 7

CONCLUSIONS

There is a large literature for multiple linear regression models with heterogeneity. See, for example, Buja et al. (2019), Eicker (1963, 1967), Flachaire (2005), Hinkley (1977), Huber (1967), Long and Ervin (2000), MacKinnon and White (1985), Rajapaksha and Olive (2024), Romano and Wolf (2017), and White (1980). The response plot of $\hat{\boldsymbol{\phi}}_{OPLS}$ versus Y and the EE plot of $\hat{\boldsymbol{\phi}}_{OPLS}^T\boldsymbol{x}$ versus $\hat{\boldsymbol{\phi}}_{OLS}^T\boldsymbol{x}$ can be used to check whether OPLS is useful for WLS. See Olive (2013) for more on these two plots.

Tests for high dimensional covariance matrices include Chen, Zhang, and Zhong (2010), and Himeno and Yamada (2014).

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 `OPLSplot` make the response plot and residual plot for multiple linear regression based on one component partial least squares. The function `OPLSEEplot` plots the OPLS fitted values versus the OLS fitted values. Let $up \approx 1 - \alpha/2$ be the correction factor used for the confidence intervals. The function `covxycis` obtains the large sample $100(1-\alpha)\%$ confidence intervals $\approx \hat{\eta}_j \pm t_{n-1,up} SE(\hat{\eta}_j)$ for $\eta_j = \text{Cov}(x_j, Y)$ for $j = 1, ..., p$. The function `oplscis` obtains the large sample $100(1-\alpha)\%$ confidence intervals $\approx \hat{\beta}_j \pm t_{n-1,up} SE(\hat{\beta}_j)$ for $\beta_j = \lambda\text{Cov}(x_j, Y)$ for $j = 1, ..., p$. If $[L_j, U_j]$ is the confidence interval for η_j, then $[\hat{\lambda}L_j, \hat{\lambda}U_j]$ is the confidence interval for β_j. The function `oplswls` generates a weighted least squares data set of types used by the simulation, the OPLS response plot, the OLS response plot, and the plot of the OPLS fitted values versus the OLS fitted values. In the literature, simulated WLS data set often contain outliers and are often not very linear. The response plot can be used to check for these two problems. The function `oplswsim` was used for Table 9. The function `rcovxy` makes the classical and three robust estimators of $\boldsymbol{\eta}$, and makes a scatterplot matrix of the four estimated sufficient predictors $\hat{\boldsymbol{\eta}}^T \boldsymbol{x}$ and Y. Only two robust estimators are made if $n \leq 2.5p$. The function `oplssim2` was used for Tables 17 and 28.

References

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.

Bickel, P.J., and Doksum, K.A. (2007), *Mathematical Statistics: Basic Ideas and Selected Topics,* Vol. 1., 2nd ed., Updated Printing, Pearson Prentice Hall, Upper Saddle River, NJ.

Buja, A., Brown, L., Berk, R., George, E., Pitkin, E., Traskin, M., Zhang, K., and Zhao, L. (2019), "Models as Approximations I: Consequences Illustrated with Linear Regression," *Statistical Science*, 34, 523-544.

Chang, J., and Olive, D.J. (2010), "OLS for 1D Regression Models," *Communications in Statistics: Theory and Methods*, 39, 1869-1882.

Chen, S.X., Zhang, L.X. and Zhong, P.S. (2010), "Tests for High-Dimensional Covariance Matrices, *Journal of the American Statistical Association*, 105(490):810-819.

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

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

Eicker, F. (1963), “Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions,” *Annals of Mathematical Statistics*, 34, 447-456.

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

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

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

Flachaire, E. (2005), “Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap vs. Pairs Bootstrap, *Computational Statistics* & *Data Analysis*, 49, 361-376.

Freedman, D.A. (1981), “Bootstrapping Regression Models,” *The Annals of Statistics*, 9, 1218-1228.

Hebbler, B. (1847), “Statistics of Prussia,” *Journal of the Royal Statistical Society, A*, 10, 154-186.

Himeno, T., and Yamada, T. (2014), “Estimations for Some Functions of Covariance Matrix in High Dimension under Non-Normality and Its Applications,” *Journal of Multivariate Analysis*, 130, 27-44.

Hinkley, D.V. (1977), “Jackknifing in Unbalanced Situations,” *Technometrics*, 19, 285-292.

Huber, P.J. (1967), “The Behavior of Maximum Likelihood Estimation Under Nonstandard Conditions,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, eds. LeCam, L.M., and Neyman, J., University of California Press, Berkeley, CA, 221-223.

Long, J.S., and Ervin, L.H. (2000), "Using Heteroscedasticity Consistent Standard Errors in the Linear Model," *The American Statistician*, 54, 217-224.

MacKinnon, J.G., and White, H. (1985), "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305-325.

Maronna, R.A., and Zamar, R.H. (2002), "Robust Estimates of Location and Dispersion for High-Dimensional Datasets," *Technometrics,* 44, 307-317.

Mehrotra, D.V. (1995), "Robust Elementwise Estimation of a Dispersion Matrix," *Biometrics*, 51, 1344-1351.

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

Olive, D.J. (2004), "A Resistant Estimator of Multivariate Location and Dispersion," *Computational Statistics* & *Data Analysis,* 46, 99-102.

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

Olive, D.J. (2014), *Statistical Theory and Inference*, Springer, New York, NY.

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

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

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.

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

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.

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

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

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.

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.

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

Seber, G.A.F., and Lee, A.J. (2003), *Linear Regression Analysis,* 2nd ed., Wiley, New York, NY.

Sen, P.K., and Singer, J.M. (1993), *Large Sample Methods in Statistics: an Introduction with Applications,* Chapman & Hall, New York, NY.

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

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.

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

White, H. (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.

Wieczorek, J., and Lei, J. (2022), "Model-Selection Properties of Forward Selection and Sequential Cross-Validation for High-Dimensional Regression," *Canadian Journal of Statistics*, 50, 454-470.

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

Zhang, J., Olive, D.J., and Ye, P. (2012), "Robust Covariance Matrix Estimation with Canonical Correlation Analysis," *International Journal of Statistics and Probability*, 1, 119-136.

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

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.