OLS TESTING WITH PREDICTORS SCALED TO HAVE UNIT SAMPLE VARIANCE

by

Sanjuka Johana Lemonge

B.Sc., University of Kelaniya, 2019

M.S., Southern Illinois University Carbondale, 2022

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

School of Mathematical and Statistical Sciences
in the Graduate School
Southern Illinois University Carbondale
August 2025

DISSERTATION APPROVAL

OLS TESTING WITH PREDICTORS SCALED TO HAVE UNIT SAMPLE VARIANCE

by

Sanjuka Johana Lemonge

A Dissertation Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in the field of Mathematics

Approved by:

Dr. David Olive, Chair

Dr. Michael Sullivan

Dr. Seyed Yaser Samadi

Dr. Lindsey-Kay Lauderdale

Dr. Xiaolan Huang

Graduate School
Southern Illinois University Carbondale
July 1, 2025

AN ABSTRACT OF THE DISSERTATION OF

Sanjuka Johana Lemonge, for the Doctor of Philosophy degree in Mathematics, presented on July 1, 2025, at Southern Illinois University Carbondale.

TITLE: OLS TESTING WITH PREDICTORS SCALED TO HAVE UNIT SAMPLE VARIANCE

MAJOR PROFESSOR: Dr. David Olive

We consider hypothesis tests for the multiple linear regression model with ordinary least squares if the predictor variables have been scaled to have unit sample variance. Some tests are unchanged, but confidence intervals, confidence regions, and some tests are no longer valid.

KEY WORDS: Multiple linear regression.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my advisor, Dr. David Olive for his dedicated support and guidance he has given me throughout my study for past five years. Without his support, advise, and encouragement, this research and dissertation would not have happened. I feel blessed to have him as my Ph.D. advisor.

I would also like to thank Dr. Michael Sullivan, Dr. Yaser Samadi, Dr. Lindsey-Kay Lauderdale and Dr. Xiaolan Huang for sitting on my committee and all the professors and staff of School of Mathematical and Statistical Sciences at Southern Illinois University for their instructions and care throughout my time as a graduate student.

Last but not least, I want to say thank you to my parents and my wife, for all of your support and encouragement. You have pushed me to succeed and I could not have done it without you!

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This section reviews multiple linear regression models. Consider a multiple linear regression model with response variable Y and predictors $\boldsymbol{x} = (x_1, ..., x_p)$ where a constant $x_1 \equiv 1$ is in the model. Then there are n cases $(Y_i, \boldsymbol{x}_i^T)^T$, and the sufficient predictor $SP = \boldsymbol{x}^T\boldsymbol{\beta}$. 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 multiple linear regression model be

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

for $i = 1, ..., n$. Here n is the sample size and the random variable e_i is the ith error. Assume that the e_i are independent and identically distributed (iid) with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma^2$. In matrix notation, these n equations become $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ 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 the covariance matrix $\text{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}_n$ where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. The OLS estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, the vector of fitted values is $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$, the vector of residuals is $\boldsymbol{r} = \boldsymbol{Y} - \hat{\boldsymbol{Y}}$, and $\hat{\sigma}^2 = MSE = \sum_{i=1}^n r_i^2/(n-p)$.

There are many multiple linear regression methods, and it is often convenient to use centered or scaled data. See James et al. (2021). Suppose U has observed values $U_1, ..., U_n$. Let g be an integer near 0. If the sample variance of the U_i is

$$\hat{\sigma}_g^2 = \frac{1}{n-g}\sum_{i=1}^{n}(U_i - \overline{U})^2,$$

then the sample standard deviation of U_i is $\hat{\sigma}_g$. If the values of U_i are not all the same, then $\hat{\sigma}_g > 0$. Using $g = 1$ gives an unbiased estimator s^2 of σ^2, while $g = 0$ gives the method of moments estimator.

Next consider scaling the predictors. If $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{X}, \boldsymbol{Y}) + \boldsymbol{e}$, the model with scaled predictors is $\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{\beta}(\boldsymbol{W}, \boldsymbol{Y}) + \boldsymbol{\epsilon}$ where $\boldsymbol{\beta}(\boldsymbol{X}, \boldsymbol{Y})$ denotes the population coefficients from the OLS regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Here $\boldsymbol{W} = \boldsymbol{X}\hat{\boldsymbol{D}}_n$ where the $p \times p$ matrix $\hat{\boldsymbol{D}}_n = diag(1, 1/s_2, ..., 1/s_p)$ where $s_j = \hat{\sigma}_j$ for the jth predictor x_j, and $j = 2, ..., p$. Since OLS is affine equivariant and $\hat{\boldsymbol{D}}_n$ is nonsingular, $\hat{\boldsymbol{\beta}}(\boldsymbol{W}, \boldsymbol{Y}) = \hat{\boldsymbol{\beta}}(\boldsymbol{X}\hat{\boldsymbol{D}}_n, \boldsymbol{Y}) = \hat{\boldsymbol{D}}_n^{-1}\hat{\boldsymbol{\beta}}(\boldsymbol{X}, \boldsymbol{Y})$. Then $\boldsymbol{H}_{\boldsymbol{W}} = \boldsymbol{W}(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \boldsymbol{H}_{\boldsymbol{X}}$, and the residuals and fitted values are the same for both models. See, for example, Olive (2017, p. 413).

Now consider centered data $Y_i - \overline{Y} = \beta_1^* + (x_{i,2} - \overline{x}_2)\beta_2 + \cdots + (x_{i,p} - \overline{x}_p)\beta_p + \epsilon_i$ or $Z_i = \beta_1^* + w_{i,2}\beta_2 + \cdots + w_{i,p}\beta_p + \epsilon_i$. Do the OLS regression. Since the sample means of the centered response and centered predictors are 0, $\hat{\beta}_1^* = 0$. In terms of the original predictors, $\hat{Y}_i = \tilde{\beta}_1 + x_{i,2}\tilde{\beta}_2 + \cdots + x_{i,p}\tilde{\beta}_p$ where $\tilde{\beta}_1 = \overline{Y} - \tilde{\beta}_2\overline{x}_2 - \cdots - \tilde{\beta}_p\overline{x}_p$. Then $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ since OLS estimators minimize the sum of squared residuals (if $\tilde{\boldsymbol{\beta}} \neq \hat{\boldsymbol{\beta}}$, then one of the estimators has a smaller sum of squared residuals, contradicting the fact that both estimators are OLS estimators). Hence centering the response and predictors gives an equivalent method for computing $\hat{\boldsymbol{\beta}}$, and the large sample theory for the equivalent estimators is unchanged.

There are variants to scaling the predictor variables. Often the response variable is also scaled. Often the response and predictors are centered and scaled by obtaining the Z-score of each variable $Z_i = (w_i - \overline{w})/S_w$ where $\overline{w}$ and S_w are the sample mean and sample standard deviation of the predictor or response variable w_i.

Suppose that multiple linear regression output for the scaled data $(\boldsymbol{W}, \boldsymbol{Y})$ is obtained from OLS software such as R. a) It is known that confidence intervals for the β_i are incorrect. See, for example, Yuan and Chang (2011) and Jones and Waller (2015). b) It is also known that the t-statistic for testing $H_0 : \beta_i = 0$ versus $H_1 : \beta_i \neq 0$ is the same for scaled and unscaled data. This is the case since the standard error formulas for the scaled data are only correct when $\beta_i = 0$. See, for example, Schielzeth (2010), van Ginkel (2020), and Yuan and Chang (2011).

Yuan and Chang (2011) obtain the asymptotic distribution for the scaled data where the formulas are simplified by assuming that the nontrivial predictors are iid from an elliptically con-

toured distribution such as the multivariate normal distribution. Jones and Waller (2015) derive the asymptotic distribution for a broader class of distributions and provide *R* software. The asymptotic covariance matrix is difficult to estimate, and the sample size n needs to be large.

Chapter 2 shows that for multiple linear regression with standardized predictors, OLS software tests of the form $H_0 : \boldsymbol{\beta}_O = \mathbf{0}$ versus $H_1 : \boldsymbol{\beta}_O \neq \mathbf{0}$ are valid large sample tests where $\boldsymbol{\beta}_O = (\beta_{i_1}, ..., \beta_{i_k})^T$. Note that such tests include $H_0 : \beta_i = 0$ versus $H_0 : \beta_i \neq 0$, and the Anova F test: $H_0 : (\beta_2, ..., \beta_p)^T = \mathbf{0}$ versus $H_0 : (\beta_2, ..., \beta_p)^T \neq \mathbf{0}$.

CHAPTER 2

LARGE SAMPLE THEORY

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). Let $\boldsymbol{H} = \boldsymbol{H_X}$, and let h_i be the ith diagonal element of $\boldsymbol{H}$. Theorem 1 acts 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. Let the leverages $h_i = \boldsymbol{H}_{ii}$ be the diagonal elements of $\boldsymbol{H}$.

Theorem 1. Consider the MLR model and assume that the zero mean errors are iid with $E(e_i) = 0$ and $\text{VAR}(e_i) = \sigma^2$. 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 non-stochastic constant vectors). Then the OLS estimator $\hat{\boldsymbol{\beta}}$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2\, \boldsymbol{V}). \tag{2.1}$$

Consider testing $H_0 : \boldsymbol{L\beta} = \boldsymbol{c}$ where $\boldsymbol{L}$ is a full rank $k \times p$ constant matrix and $\boldsymbol{c}$ is a $k \times 1$ constant vector. If H_0 is true, then by Theorem 1, $\sqrt{n}\boldsymbol{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \sqrt{n}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c}) \xrightarrow{D} N_k(\boldsymbol{0}, \sigma^2\, \boldsymbol{VL}^T)$. Hence $\sqrt{n}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T(\sigma^2\, \boldsymbol{VL}^T)^{-1}\sqrt{n}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c}) \xrightarrow{D} \chi^2_k$ as $n \to \infty$. Let $\hat{\sigma}^2 = MSE \xrightarrow{P} \sigma^2$ and $\hat{\boldsymbol{V}} = n(\boldsymbol{X}^T\boldsymbol{X})^{-1} \xrightarrow{P} \boldsymbol{V}$ as $n \to \infty$ where convergence in probability indicates a consistent estimator. Then $\sqrt{n}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T(\hat{\sigma}^2\, \boldsymbol{L}\hat{\boldsymbol{V}}\boldsymbol{L}^T)^{-1}\sqrt{n}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c}) =$

$$kF_1 = \frac{1}{MSE}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T[\boldsymbol{L}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}^T]^{-1}(\boldsymbol{L}\hat{\boldsymbol{\beta}} - \boldsymbol{c}) \xrightarrow{D} \chi^2_k \tag{2.2}$$

as $n \to \infty$ if H_0 is true. If H_0 is true, then an $F_{1-\alpha,k,n-p}$ cutoff can be used for $F_1 = kF_1/k$ since $kF_{k,n-p} \xrightarrow{D} \chi^2_k$ as $n \to \infty$. See Seber and Lee (2003, p. 100).

If $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta}(\boldsymbol{X}, \boldsymbol{Y}) + \boldsymbol{e}$, the model with scaled predictors is $\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{\beta}(\boldsymbol{W}, \boldsymbol{Y}) + \boldsymbol{\epsilon}$ where $\boldsymbol{\beta}(\boldsymbol{X}, \boldsymbol{Y})$ denotes the population coefficients from the OLS regression of $\boldsymbol{Y}$ on $\boldsymbol{X}$. Here $\boldsymbol{W} = \boldsymbol{X}\hat{\boldsymbol{D}}_n$. As noted in chapter 1, and the residuals and fitted values are the same for both models. Thus $\hat{Y} =$

$$\hat{\beta}_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p = \hat{\beta}_1 + \hat{\beta}_2 s_2 \frac{x_2}{s_2} + \cdots + \hat{\beta}_p s_p \frac{x_p}{s_p} = \hat{\beta}_1 + \hat{\beta}_2(\boldsymbol{W}, Y) w_2 + \cdots + \hat{\beta}_p(\boldsymbol{W}, Y) w_p.$$

Hence $\hat{\boldsymbol{\beta}}(\boldsymbol{W}, \boldsymbol{Y}) = (\hat{\beta}_1, \hat{\beta}_2 s_2, ..., \hat{\beta}_p s_p)^T = \hat{\boldsymbol{D}}_n^{-1}\hat{\boldsymbol{\beta}}(\boldsymbol{X}, Y)$ where $\hat{\boldsymbol{\beta}}(\boldsymbol{X}, Y) = (\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_p)^T$.

For the scaled predictors, assume $\hat{\boldsymbol{D}}_n \xrightarrow{P} \boldsymbol{D} = diag(1, 1/\sigma_2, ..., 1/\sigma_p)$ where each $\sigma_i > 0$. This assumption often holds if the $\boldsymbol{x}_i$ are iid from some population. Let $\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{X}, \boldsymbol{Y})$. Then

$$\sqrt{n}(\hat{\boldsymbol{\beta}}(\boldsymbol{W}, \boldsymbol{Y}) - \boldsymbol{D}^{-1}\boldsymbol{\beta}) = \sqrt{n}(\hat{\boldsymbol{D}}_n^{-1}\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{D}}_n^{-1}\boldsymbol{\beta} + \hat{\boldsymbol{D}}_n^{-1}\boldsymbol{\beta} - \boldsymbol{D}^{-1}\boldsymbol{\beta})$$

$$= \sqrt{n}\hat{\boldsymbol{D}}_n^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \sqrt{n}(\hat{\boldsymbol{D}}_n^{-1} - \boldsymbol{D}^{-1})\boldsymbol{\beta} = \boldsymbol{z}_n + \boldsymbol{b}_n$$

where $\boldsymbol{z}_n = \sqrt{n}\hat{\boldsymbol{D}}_n^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{D}^{-1}\boldsymbol{V}_{\boldsymbol{x}}\boldsymbol{D}^{-1})$ if $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_{\boldsymbol{x}})$. Note that $\hat{\boldsymbol{D}}_n^{-1}\hat{\boldsymbol{\beta}} \xrightarrow{P} \boldsymbol{D}^{-1}\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{W}, \boldsymbol{Y})$. Now

$$\boldsymbol{b}_n = \begin{pmatrix} 0 \\ \sqrt{n}(\hat{\sigma}_2 - \sigma_2)\beta_2 \\ \vdots \\ \sqrt{n}(\hat{\sigma}_p - \sigma_p)\beta_p \end{pmatrix} = \begin{pmatrix} 0 \\ b_{2,n} \\ \vdots \\ b_{p,n} \end{pmatrix} = O_p(1)$$

if $\sqrt{n}(\hat{\sigma}_i - \sigma_i) \xrightarrow{D} N(0, \tau_i^2)$. Then $b_{i,n} \xrightarrow{D} N(0, \beta_i^2 \tau_i^2)$ for $i = 2, ..., p$. Thus $\sqrt{n}(\hat{\boldsymbol{\beta}}(\boldsymbol{W}, \boldsymbol{Y}) - \boldsymbol{D}^{-1}\boldsymbol{\beta})$ does not converge in distribution to $\boldsymbol{z} \sim N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{D}^{-1}\boldsymbol{V}_{\boldsymbol{x}}\boldsymbol{D}^{-1})$ unless $\boldsymbol{b}_n \xrightarrow{P} \boldsymbol{0}$.

```
A <- matrix(1,ncol=5,nrow=5)
A     [,1] [,2] [,3] [,4] [,5]
[1,]    1    1    1    1    1
[2,]    1    1    1    1    1
[3,]    1    1    1    1    1
[4,]    1    1    1    1    1
[5,]    1    1    1    1    1
D <- diag(1:5)
D%*%A #premultiplying multiplies ith row of A by d_i
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    1    1    1    1
[2,]    2    2    2    2    2
[3,]    3    3    3    3    3
[4,]    4    4    4    4    4
[5,]    5    5    5    5    5
A%*%D #postmultiplying multiplies jth col. of A by d_j
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    2    3    4    5
[2,]    1    2    3    4    5
[3,]    1    2    3    4    5
[4,]    1    2    3    4    5
[5,]    1    2    3    4    5
D%*%A%*%D # ijth element of A gets multiplied by d_i d_j
     [,1] [,2] [,3] [,4] [,5]
[1,]    1    2    3    4    5
[2,]    2    4    6    8   10
[3,]    3    6    9   12   15
```

```
[4,]    4    8   12   16   20
[5,]    5   10   15   20   25
```

Using the scaled data $(\boldsymbol{W}, Y)$ in the OLS software gives an incorrect normal approximation $\hat{\boldsymbol{\beta}}(\boldsymbol{W}, Y) \approx N_p(\boldsymbol{\beta}(\boldsymbol{W}, Y), MSE\ n\ (\boldsymbol{W}^T\boldsymbol{W})^{-1}) =$

$$N_p(\boldsymbol{D}^{-1}\boldsymbol{\beta}(\boldsymbol{X}, Y), MSE\ n\ \hat{\boldsymbol{D}}_n^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\boldsymbol{D}}_n^{-1}).$$

Hence confidence intervals, confidence regions, and many tests of hypotheses will no longer be valid. An important exception occurs for the partial F tests of the form $H_0 : \boldsymbol{L}_O\boldsymbol{\beta} = \boldsymbol{0}$ with $\boldsymbol{c} = \boldsymbol{0}$ and $\boldsymbol{L}_O$ a full rank $k \times p$ matrix where $\boldsymbol{L}_O\boldsymbol{\beta} = \boldsymbol{\beta}_O = (\beta_{i_1}, ..., \beta_{i_k})^T$ and $O = \{i_1, ..., i_k\}$. For such a test, we would like to leave the predictors $\boldsymbol{L}_O\boldsymbol{x} = \boldsymbol{x}_O = (x_{i_1}, ..., x_{i_k})^T$ out of the regression model, resulting in a reduced model. Note that the jth row of L_O has a 1 in the i_jth position, with all other entries equal to 0.

Let the ijth element of a $p \times m$ matrix $\boldsymbol{A}$ be a_{ij}. Then $\boldsymbol{A} = (a_{ij})$. Thus $\boldsymbol{L}_O\boldsymbol{A} = \boldsymbol{A}_O = (a_{i_a,j})$ where the ath row of $\boldsymbol{A}_O$ is the i_ath row of $\boldsymbol{A}$ for $a = 1, ..., k$. Similarly, if $\boldsymbol{C} = (c_{ij})$ is a $p \times p$ matrix, then

$$\boldsymbol{L}_O\boldsymbol{C}\boldsymbol{L}_O^T = \boldsymbol{C}_{OO} = \begin{pmatrix} c_{i_1,i_1} & c_{i_1,i_2} & ... & c_{i_1,i_k} \\ c_{i_2,i_1} & c_{i_2,i_2} & ... & c_{i_2,i_k} \\ \vdots & \vdots & ... & \vdots \\ c_{i_k,i_1} & c_{i_k,i_2} & ... & c_{i_k,i_k} \end{pmatrix} = (c_{i_a,i_b}).$$

Let $\boldsymbol{Q} = diag(d_1, ..., d_p)$ be a $p \times p$ diagonal matrix with diagonal elements $d_1, ..., d_p$. Let $\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{A} = (h_{ij}) = (d_i a_{ij})$. Then $\boldsymbol{L}_O\boldsymbol{Q}\boldsymbol{A} = \boldsymbol{L}_O\boldsymbol{H} = \boldsymbol{H}_O = (h_{i_a,j}) = (d_{i_a}a_{i_a,j}) = \boldsymbol{Q}_{OO}\boldsymbol{A}_{OO}$. Let $\boldsymbol{B} = \boldsymbol{Q}\boldsymbol{C}\boldsymbol{Q} = (b_{ij}) = (d_i d_j c_{ij})$. Then $\boldsymbol{L}_O\boldsymbol{B}\boldsymbol{L}_O^T = \boldsymbol{B}_{OO} = (b_{i_a,i_b}) = (d_{i_a}d_{i_b}c_{i_a,i_b}) = \boldsymbol{Q}_{OO}\boldsymbol{C}_{OO}\boldsymbol{Q}_{OO}$.

Theorem 2. For the test $H_0 : \boldsymbol{L}_O\boldsymbol{\beta} = \boldsymbol{0}$, the partial F test statistics from the scaled data and the unscaled data are the same.

Proof. The result holds if

$$(\boldsymbol{L}_O\hat{\boldsymbol{\beta}})^T[\boldsymbol{L}_O(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}_O^T]^{-1}(\boldsymbol{L}_O\hat{\boldsymbol{\beta}}) = (\boldsymbol{L}_O\hat{\boldsymbol{\beta}}(\boldsymbol{W}, Y))^T[\boldsymbol{L}_O(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{L}_O^T]^{-1}(\boldsymbol{L}_O\hat{\boldsymbol{\beta}}(\boldsymbol{W}, Y)).$$

By the above remarks, $\boldsymbol{L}_O\hat{\boldsymbol{D}}_n\boldsymbol{L}_O^T = \hat{\boldsymbol{D}}_{OO} = diag(1/s_{i_1}, ..., 1/s_{i_k})$ where we define $s_1 = 1$. Let $\boldsymbol{Q} = \boldsymbol{D}_n^{-1}$ and $\boldsymbol{C} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}$.

Then $\boldsymbol{L}_O\hat{\boldsymbol{\beta}}(\boldsymbol{W}, Y) = \boldsymbol{L}_O\hat{\boldsymbol{D}}_n^{-1}\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_O(\boldsymbol{W}, Y) = \hat{\boldsymbol{D}}_{OO}^{-1}\hat{\boldsymbol{\beta}}_O = \hat{\boldsymbol{D}}_{OO}^{-1}\boldsymbol{L}_O\hat{\boldsymbol{\beta}}$, while

$$\boldsymbol{L}_O(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{L}_O^T = \boldsymbol{L}_O(\hat{\boldsymbol{D}}_n\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{D}}_n)^{-1}\boldsymbol{L}_O^T = \boldsymbol{L}_O\hat{\boldsymbol{D}}_n^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\boldsymbol{D}}_n^{-1}\boldsymbol{L}_O^T$$

$$= \hat{\boldsymbol{D}}_{OO}^{-1}(\boldsymbol{X}^T\boldsymbol{X})_{OO}^{-1}\hat{\boldsymbol{D}}_{OO}^{-1} = \hat{\boldsymbol{D}}_{OO}^{-1}\boldsymbol{L}_O(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}_O^T\hat{\boldsymbol{D}}_{OO}^{-1}.$$

Thus $(\boldsymbol{L}_O\hat{\boldsymbol{\beta}}(\boldsymbol{W}, Y))^T[\boldsymbol{L}_O(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{L}_O^T]^{-1}(\boldsymbol{L}_O\hat{\boldsymbol{\beta}}(\boldsymbol{W}, Y)) =$

$$(\hat{\boldsymbol{D}}_{OO}^{-1}\boldsymbol{L}_O\hat{\boldsymbol{\beta}})^T[\hat{\boldsymbol{D}}_{OO}^{-1}\boldsymbol{L}_O(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}_O^T\hat{\boldsymbol{D}}_{OO}^{-1}]^{-1}\hat{\boldsymbol{D}}_{OO}^{-1}\boldsymbol{L}_O\hat{\boldsymbol{\beta}} =$$

$$(\boldsymbol{L}_O\hat{\boldsymbol{\beta}})^T[\boldsymbol{L}_O(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{L}_O^T]^{-1}(\boldsymbol{L}_O\hat{\boldsymbol{\beta}}),$$

proving the theorem. □

Let $\boldsymbol{x}_i^T$ be the ith row of $\boldsymbol{X}$, and let $\boldsymbol{w}_i^T$ be the ith row of $\boldsymbol{W}$. Let $\hat{\beta}_i = \hat{\beta}_i(\boldsymbol{x}, Y)$ be the ith OLS estimator of $\beta_i = \beta_i(\boldsymbol{x}, Y)$ where $(\boldsymbol{x}, Y)$ denotes that the Y were regressed on the $\boldsymbol{x}$. Similarly, $\hat{\beta}_i(\boldsymbol{w}, Y)$ is the estimator when the Y are regressed on the $\boldsymbol{w}_i$. Let $[L_{in}, U_{in}] = \hat{\beta}_i \pm t_{1-\alpha/2,n-p}SE(\hat{\beta}_i)$ be the large sample $100(1-\alpha)\%$ confidence interval CI for β_i. Let $\sigma_i^2 = Var(x_i)$ for $i = 2, ..., p$. Then $\beta_i(\boldsymbol{w}, Y) = \sigma_i\beta_i(\boldsymbol{x}, Y)$ for $i = 2, ..., p$, and the "CI" for $\beta_i(\boldsymbol{w}, Y)$ is $[s_iL_{in}, s_iU_{in}]$. (By "CI", we mean the confidence interval formula for the unstandardized data was applied to the data with standardized predictors, which is not a valid confidence interval.) This result holds since $(\boldsymbol{W}^T\boldsymbol{W})^{-1} = \hat{\boldsymbol{D}}_n^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\boldsymbol{D}}_n^{-1}$. Scaling does not change the MSE, hence $SE[\hat{\beta}_i(\boldsymbol{w}, Y)] = s_iSE[\hat{\beta}_i(\boldsymbol{x}, Y)]$ for $i = 2, ..., p$ where s_i^2

is the usual unbiased estimator of σ_i^2. If $\beta_i(\boldsymbol{w}, Y) = \beta_i(\boldsymbol{x}, Y) = 0$, then $\beta_i = 0$ is in the interval $[L_{in}, U_{in}]$ if and only if $\beta_i(\boldsymbol{w}, Y) = \sigma_i\beta_i(\boldsymbol{x}, Y) = 0$ is in the "CI" $[s_iL_{in}, s_iU_{in}]$ since $s_i > 0$. Hence in the simulation where $\beta_i = 0$, the coverage of the CI for $\beta_i(\boldsymbol{x}, Y)$ and the coverage of the "CI" for $\beta_i(\boldsymbol{w}, Y)$ will be exactly the same. When $\beta_i \neq 0$, we expect that the coverages will differ, and that the "CI" for $\beta_i(\boldsymbol{w}, Y)$ will often have undercoverage. Here the coverage is the observed proportion of intervals that contained the population parameter. Hence if 5000 CIs for β_i were made, and 4750 of the CIs contained β_i, then the (observed) coverage is 4750/5000 = 0.95.

The simulations used $\boldsymbol{L} = \boldsymbol{L}_O$ where $\boldsymbol{L}_O\boldsymbol{\beta} = \boldsymbol{c} = \boldsymbol{\beta}_O = (\beta_{i_1}, ..., \beta_{i_k})^T$ and $O = \{i_1, ..., i_k\}$.

CHAPTER 3

SURVIVAL ANALYSIS

This section follows Olive (2025b) closely.

Definition 1. Let $Y \geq 0$ be the time until an event occurs. Then Y is called the survival time or time until event. The survival time is censored if the event of interest has not been observed. Let Y_i be the ith survival time. Let Z_i be the time the ith observation (possibly an individual or machine) leaves the study for any reason other than the event of interest. Then Z_i is the time until the ith observation is censored. Then the right censored survival time T_i of the ith observation is $T_i = \min(Y_i, Z_i)$. Let $\delta_i = 0$ if T_i is (right) censored ($T_i = Z_i$) and let $\delta_i = 1$ if T_i is not censored ($T_i = Y_i$). Then the univariate survival analysis data is $(T_1, \delta_1), (T_2, \delta_2), ..., (T_n, \delta_n)$. Alternatively, the data is $T_1, T_2^*, T_3, ..., T_{n-1}^*, T_n$ where the * means that the case was (right) censored. Sometimes the asterisk * is replaced by a plus +, and Y_i, y_i or t_i can replace T_i.

Definition 2. i) The cumulative distribution function (cdf) of Y is $F(t) = P(Y \leq t)$. Since $Y \geq 0$, $F(0) = 0$, $F(\infty) = 1$, and $F(t)$ is nondecreasing.

ii) The probability density function (pdf) of Y is $f(t) = F'(t)$.

iii) The survival function of Y is $S(t) = P(Y > t)$. $S(0) = 1, S(\infty) = 0$ and $S(t)$ is nonincreasing.

iv) The hazard function of Y is $h(t) = \dfrac{f(t)}{1 - F(t)}$ for $t > 0$ and $F(t) < 1$. Note that $h(t) \geq 0$ if $F(t) < 1$.

v) The cumulative hazard function of Y is $H(t) = \int_0^t h(u)du$ for $t > 0$. It is true that $H(0) = 0, H(\infty) = \infty$, and $H(t)$ is nondecreasing.

Assume $Y \geq 0$. Then $F(0) = 0$, $S(0) = 1$, and $H(0) = 1$. Note that $S(\infty) = 0$ implies that $H(\infty) = \infty$ where $\lim_{t \to \infty} H(t) = H(\infty)$. Memorize that $0 \leq F(t) \leq 1$, $0 \leq S(t) \leq 1$, $f(t) \geq 0$, $h(t) \geq 0$, and $H(t) \geq 0$.

Given one of $F(t), f(t), S(t), h(t)$ or $H(t)$, the following theorem shows how to find the other 4 quantities for $t > 0$. Each of these five quanities completely determines the distribution of the

random variable. In reliability analysis, the *reliability function* $R(t) = S(t)$, and in economics, Mill's ratio $= 1/h(t)$. In actuarial sciences, $h(t)$ is the *force of mortality*.

Theorem 3.

A) $F(t) = \int_0^t f(u)du = 1 - S(t) = 1 - \exp[-H(t)] = 1 - \exp[-\int_0^t h(u)du]$.

B) $f(t) = F'(t) = -S'(t) = h(t)[1 - F(t)] = h(t)S(t) = h(t)\exp[-H(t)] = H'(t)\exp[-H(t)]$.

C) $S(t) = 1 - F(t) = 1 - \int_0^t f(u)du = \int_t^\infty f(u)du = \exp[-H(t)] = \exp[-\int_0^t h(u)du]$.

D)

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} = \frac{F'(t)}{1 - F(t)} = \frac{-S'(t)}{S(t)} = -\frac{d}{dt}\log[S(t)] = H'(t).$$

E) $H(t) = \int_0^t h(u)du = -\log[S(t)] = -\log[1 - F(t)]$.

Example 1. Suppose $Y \sim EXP(\lambda)$ where $\lambda > 0$, then $h(t) = \lambda$ for $t > 0$, $f(t) = \lambda e^{-\lambda t}$ for $t > 0$, $F(t) = 1 - e^{-\lambda t}$ for $t > 0$, $S(t) = e^{-\lambda t}$ for $t > 0$, $H(t) = \lambda t$ for $t > 0$ and $E(Y) = 1/\lambda$. The exponential distribution is the only distribution of a continuous random variable Y with a constant hazard function since $h(t)$ completely determines the distribution of the random variable Y. Derive $H(t), S(t), F(t)$, and $f(t)$ from the constant hazard function $h(t) = \lambda$ for $t > 0$ and some $\lambda > 0$.

Solution: $H(t) = \int_0^t h(u)du = \int_0^t \lambda du = \lambda t$ for $t > 0$.

$S(t) = e^{-H(t)} = e^{-\lambda t}$, for $t > 0$.

$F(t) = 1 - S(t) = 1 - e^{-\lambda t}$ for $t > 0$.

Finally, $f(t) = h(t)S(t) = \lambda e^{-\lambda t} = F'(t)$ for $t > 0$.

Example 2. If $Y \sim$ Weibull(γ, λ) where $\gamma > 0$ and $\lambda > 0$, then $h(t) = \lambda\gamma t^{\gamma-1}$ for $t > 0$, $f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^\gamma)$ for $t > 0$, $F(t) = 1 - \exp(-\lambda t^\gamma)$ for $t > 0$, $S(t) = \exp(-\lambda t^\gamma)$ for $t > 0$, $H(t) = \lambda t^\gamma$ for $t > 0$. The Weibull$(\lambda, \gamma = 1)$ distribution is the EXP(λ) distribution. The hazard function can be increasing, decreasing or constant. Hence the Weibull distribution often fits reliability data well, and the Weibull distribution is an important distribution in reliability analysis. Derive $H(t), S(t), F(t)$, and $f(t)$ if $Y \sim$ Weibull(λ, γ).

Solution:

$$H(t) = \int_0^t h(u)du = \int_0^t \lambda\gamma u^{\gamma-1}du = \lambda\gamma\frac{u^\gamma}{\gamma}\Big|_0^t = \lambda t^\gamma \text{ for t} > 0.$$

$S(t) = \exp[-H(t)] = \exp[-\lambda t^\gamma]$, for $t > 0$.

$F(t) = 1 - S(t) = 1 - \exp[-\lambda t^\gamma]$ for $t > 0$.

Finally, $f(t) = h(t)S(t) = \lambda\gamma t^{\gamma-1}\exp[-\lambda t^\gamma]$ for $t > 0$.

CHAPTER 4

WEIBULL AND EXPONENTIAL REGRESSION

In a *1D regression model*, the response variable Y is conditionally independent of the $p \times 1$ vector of predictors $\boldsymbol{x}$ given the sufficient predictor $SP = h(\boldsymbol{x})$, written

$$Y \perp\!\!\!\perp \boldsymbol{x} | SP \text{ or } \mathrm{Y} \perp\!\!\!\perp \boldsymbol{x} | \mathrm{h}(\boldsymbol{x}), \tag{4.1}$$

where the real valued function $h : \mathbb{R}^p \to \mathbb{R}$. The estimated sufficient predictor $ESP = \hat{h}(\boldsymbol{x})$. An important special case is a model with a linear predictor $h(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{\beta}$ where $ESP = \boldsymbol{x}^T\hat{\boldsymbol{\beta}}$.

Definition 3. For parametric proportional hazards regression models, the baseline function is parametric and the parameters are estimated via maximum likelihood. Then as a 1D regression model, $SP = \boldsymbol{\beta}_P^T\boldsymbol{x}$, and the hazard function

$$h_{Y|SP}(t) \equiv h_{\boldsymbol{x}}(t) = \exp(\boldsymbol{\beta}_P^T\boldsymbol{x})h_{0,P}(t) = \exp(SP)h_{0,P}(t)$$

where the parametric baseline function $h_{0,P}$ depends on k unknown parameters but does not depend on the predictors $\boldsymbol{x}$. The survival function is

$$S_{\boldsymbol{x}}(t) \equiv S_{Y|SP}(t) = [S_{0,P}(t)]^{\exp(\boldsymbol{\beta}_P^T\boldsymbol{x})} = [S_{0,P}(t)]^{\exp(SP)}, \tag{4.2}$$

and

$$\hat{S}_{\boldsymbol{x}}(t) = [\hat{S}_{0,P}(t)]^{\exp(\hat{\boldsymbol{\beta}}_P^T\boldsymbol{x})} = [\hat{S}_{0,P}(t)]^{\exp(ESP)}. \tag{4.3}$$

The following univariate results will be useful for Exponential and Weibull regression. If Y has a Weibull distribution, $Y \sim W(\gamma, \lambda)$, then $S_Y(t) = \exp(-\lambda t^\gamma)$ where t, λ and γ are positive. If $\gamma = 1$, then Y has an Exponential distribution, $Y \sim EXP(\lambda)$ where $E(Y) = 1/\lambda$. Now V has a

smallest extreme value distribution, $V \sim SEV(\theta, \sigma)$, if

$$S_V(t) = P(V > t) = \exp\left(-\exp\left(\frac{t-\theta}{\sigma}\right)\right)$$

where $\sigma > 0$ while t and θ are real. If $Z \sim SEV(0,1)$, then $V = \theta + \sigma Z \sim SEV(\theta, \sigma)$ since the SEV distribution is a location scale family. Also, $V = \log(Y) \sim SEV(\theta = -\sigma\log(\lambda), \sigma = 1/\gamma)$, and $Y = e^V \sim W(\gamma = 1/\sigma, \lambda = e^{-\theta/\sigma})$.

If Y_i follows a Weibull regression model, then $\log(Y_i)$ follows an accelerated failure time (AFT) model: $\log(Y_i) = \alpha + \boldsymbol{\beta}_A^T \boldsymbol{x}_i + \sigma e_i$ where the e_i are iid $SEV(0,1)$, and $\log(Y)|\boldsymbol{x} \sim SEV(\alpha + \boldsymbol{\beta}_A^T \boldsymbol{x}, \sigma)$.

Definition 4. The Weibull proportional hazards regression (WPH) model or Weibull regression model is a parametric proportional hazards model with $Y|\boldsymbol{x} \sim W(\gamma = 1/\sigma, \lambda_{\boldsymbol{x}})$ where

$$\lambda_{\boldsymbol{x}} = \exp\left[-\left(\frac{\alpha}{\sigma} + \frac{\boldsymbol{\beta}_A^T \boldsymbol{x}}{\sigma}\right)\right] = \lambda_0 \exp(\boldsymbol{\beta}_P^T \boldsymbol{x})$$

with $\lambda_0 = \exp(-\alpha/\sigma)$ and $\boldsymbol{\beta}_P = -\boldsymbol{\beta}_A/\sigma$. Thus for $t > 0$, $P(Y > t|\boldsymbol{x}) =$

$$S_{\boldsymbol{x}}(t) = \exp(-\lambda_{\boldsymbol{x}} t^\gamma) = \exp(-\lambda_0 \exp(\boldsymbol{\beta}_P^T \boldsymbol{x}) t^\gamma) = [\exp(-\lambda_0 t^\gamma)]^{\exp(\boldsymbol{\beta}_P^T \boldsymbol{x})} =$$

$$[S_{0,P}(t)]^{\exp(\boldsymbol{\beta}_P^T \boldsymbol{x})}.$$

As a 1D regression model, $Y|SP \sim W(\gamma, \lambda_0 \exp(SP))$. Also,

$$h_i(t) = h_{Y_i|\boldsymbol{x}_i}(t) = h_{Y_i|\boldsymbol{\beta}_P^T \boldsymbol{x}_i}(t) = \exp(\boldsymbol{\beta}_P^T \boldsymbol{x}_i) h_0(t)$$

where $h_0(t) = h_0(t|\boldsymbol{\theta}) = \lambda_0 \gamma t^{\gamma-1}$ is the Weibull baseline function. Exponential regression is the special case of Weibull regression where $\sigma = 1$. Hence $Y|\boldsymbol{x} \sim W(1, \lambda_{\boldsymbol{x}}) \sim EXP(\lambda_{\boldsymbol{x}})$.

CHAPTER 5

ACCELERATED FAILURE TIME MODELS

Definition 5. For a parametric *accelerated failure time* model,

$$Z_i = \log(Y_i) = \alpha + \boldsymbol{\beta}_A^T \boldsymbol{x}_i + \sigma e_i \tag{5.1}$$

where the e_i are iid from a location scale family. Let $SP = \boldsymbol{\beta}_A^T \boldsymbol{x}$. Then as a 1D regression model, $\log(Y)|SP = \alpha + SP + e$. The parameters are again estimated by maximum likelihood and the survival function is

$$S_{\boldsymbol{x}}(t) \equiv S_{Y|\boldsymbol{x}}(t) = S_0\left(\frac{t}{\exp(\boldsymbol{\beta}_A^T \boldsymbol{x})}\right),$$

and

$$\hat{S}_{\boldsymbol{x}}(t) = \hat{S}_0\left(\frac{t}{\exp(\hat{\boldsymbol{\beta}}_A^T \boldsymbol{x})}\right)$$

where $\hat{S}_0(t)$ depends on $\hat{\alpha}$ and $\hat{\sigma}$.

For the AFT model, $h_i(t) = h_{\boldsymbol{x}}(t) = e^{-SP} h_0(t/e^{SP})$ and $S_i(t) = S_{\boldsymbol{x}}(t) = S_0(t/\exp(SP))$ where $SP = \boldsymbol{\beta}_A^T \boldsymbol{x}$. If $S_{\boldsymbol{x}}(t_{\boldsymbol{x}}(\rho)) = 1 - \rho$ for $0 < \rho < 1$, then $t_{\boldsymbol{x}}(\rho)$ is the ρth percentile. For the accelerated failure time model,

$$t_{\boldsymbol{x}}(\rho) = t_0(\rho) \exp(\boldsymbol{\beta}_A^T \boldsymbol{x})$$

where $t_0(\rho) = \exp(\sigma e_i(\rho) + \alpha)$ and $S_{e_i}(e_i(\rho)) = P(e_i > e_i(\rho)) = 1 - \rho$. Note that the estimated percentile ratio is free of ρ, $\hat{\sigma}$ and $\hat{\alpha}$

$$\frac{\hat{t}_{\boldsymbol{x}_1}(\rho)}{\hat{t}_{\boldsymbol{x}_2}(\rho)} = \exp(\hat{\boldsymbol{\beta}}_A^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)).$$

The *acceleration factor* $= e^{-SP}$ and $t_{0,\rho} = e^{-SP} t_{\boldsymbol{x},\rho}$. The median survival times are related by $t_{0,0.5} = e^{-SP} t_{\boldsymbol{x},0.5}$. If $e^{-SP} < 1$, then the median survival time of $\boldsymbol{x}$ > the median survival time of $\mathbf{0}$, a result that is good if the event is death, but bad if the event is time until recovery. Note that

$H_{\boldsymbol{x}}(t) = -\log S_{\boldsymbol{x}}(t) = -\log(S_0(t/e^{SP})) = H_0(t/e^{SP})$.

Remark 1. Assume $x_i > 0$. Then $\beta_i > 0$ increases $\log(Y_i)$ and Y_i, while $\beta_i < 0$ decreases $\log(Y_i)$ and Y_i. For the Cox PH regression model, $h_{\boldsymbol{x}}(t) = \exp(\boldsymbol{\beta}^T\boldsymbol{x})h_0(t)$. Hence $\beta_i > 0$ increases hazard and decreases Y_i, while $\beta_i < 0$ decreases hazard and increases Y_i. In the WPH model, $\boldsymbol{\beta}_P = -\boldsymbol{\beta}_A/\sigma$.

Definition 6. The *Weibull AFT* satisfies $\log(Y)|(\alpha+\boldsymbol{\beta}_A^T\boldsymbol{x}) \sim SEV(\alpha+\boldsymbol{\beta}_A^T\boldsymbol{x}, \sigma)$. The *Exponential AFT* is the special case with $\sigma = 1$.

Theorem 4. Weibull regression models, including Exponential regression models, are the only models where Y follows a proportional hazards regression model and $\log(Y)$ follows an accelerated failure time model.

If the Weibull PH regression model holds for Y_i, then $\log(Y_i) = \alpha + \boldsymbol{\beta}_A^T\boldsymbol{x}_i + \sigma e_i$ where $e_i \sim SEV(0, 1)$. Thus $\log(Y)|\boldsymbol{x} \sim SEV(\alpha + \boldsymbol{\beta}_A^T\boldsymbol{x}, \sigma)$, and the $\log(Y_i)$ follows a parametric accelerated failure time model. Two other important AFTs are i) the lognormal AFT where $\log(Y)|\boldsymbol{x} \sim N(\alpha + \boldsymbol{\beta}_A^T\boldsymbol{x}, \sigma^2)$ where the Y_i are lognormal and the $e_i \sim N(0, 1)$ are normal, and ii) the loglogistic AFT where $\log(Y)|\boldsymbol{x} \sim L(\alpha + \boldsymbol{\beta}_A^T\boldsymbol{x}, \sigma)$ where the Y_i are loglogistic and the $e_i \sim L(0, 1)$ are logistic. For the loglogistic AFT, Y follows a proportional odds model. Y does not follow a proportional hazards regression model for the loglogistic and lognormal AFTs.

A case consists of the measurements on a person or thing. Let $(\boldsymbol{x}_i^T, Y_i)^T$ be the ith case. For example, people sick from a deadly disease who go to 3 hospitals, where Y_i is the survival time. As noted by Olive (2025b), if the cases are iid and the censoring is independent of the cases, then the uncensored cases $(\boldsymbol{x}_i^T, Y_i)^T$ (where the Y_i are uncensored) may not follow the multiple linear regression model since the censoring causes the Y_i to follow a truncated distribution. However, OLS may be useful for testing $H_0 : \boldsymbol{A\beta} = \boldsymbol{0}$.

CHAPTER 6

OPLS AND THE MMLE FOR MLR

For the AFT, the log transformation of the response results in a multiple linear regression model. Hence multiple linear regression models will be useful. Now let the response variable Y be for multiple linear regression, so Y need not be a nonnegative integer. A useful multiple linear regression model is $Y|\boldsymbol{x}^T\boldsymbol{\beta} = \alpha + \boldsymbol{x}^T\boldsymbol{\beta} + e$ or $Y_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ or

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{6.1}$$

for $i = 1, ..., n$. 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 form, this model is

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}, \tag{6.2}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times (p+1)$ matrix with ith row $(1, \boldsymbol{x}_i^T)$, $\boldsymbol{\phi} = (\alpha, \boldsymbol{\beta}^T)^T$ is a $(p+1) \times 1$ vector , and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. Also $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \sigma^2\boldsymbol{I}_n$ where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix.

For a multiple linear regression model with heterogeneity, assume model (6.2) holds with $E(\boldsymbol{e}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{e}) = \boldsymbol{\Sigma}_{\boldsymbol{e}} = diag(\sigma_i^2) = diag(\sigma_1^2, ..., \sigma_n^2)$ is an $n \times n$ positive definite matrix. When the σ_i^2 are known, weighted least squares (WLS) is often used. Under regularity conditions, the ordinary least squares (OLS) estimator $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$ can be shown to be a consistent estimator of $\boldsymbol{\phi}$. See, for example, White (1980).

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

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

and

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

Then the OLS estimators for model (6.2) are $\hat{\boldsymbol{\phi}}_{OLS} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$, $\hat{\alpha}_{OLS} = \overline{Y} - \hat{\boldsymbol{\beta}}_{OLS}^T\overline{\boldsymbol{x}}$, and

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

For a multiple linear regression model with 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

$$\lambda = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T\boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}} \text{ and } \hat{\lambda} = \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}}{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}} \tag{6.3}$$

for $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq \boldsymbol{0}$. If $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{0}$, then $\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$. Also see Basa, Cook, Forzani, and Marcos (2024), Cook and Forzani (2024), and Wold (1975). Olive and Zhang (2025) 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 $\boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. Olive et al. (2025) showed that for iid cases $(\boldsymbol{x}_i, Y_i)$, these results still hold for multiple linear regression models with heterogeneity.

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, such as Poisson regression, 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

$$\hat{\boldsymbol{\beta}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}. \tag{6.4}$$

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \boldsymbol{I}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{t}, Y) \tag{6.5}$$

where $(\boldsymbol{t}, Y)$ denotes that Y was regressed on $\boldsymbol{t}$, and $\boldsymbol{I}$ is the $p \times p$ identity matrix. Olive et al. (2025) derived large sample theory for the MMLE for the multiple linear regression models, including models with heterogeneity.

6.1 LARGE SAMPLE THEORY

This section reviews the Olive and Zhang (2025) large sample theory for $\hat{\boldsymbol{\eta}}_{OPLS} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$ and OPLS for the multiple linear regression model, including some high dimensional tests for low dimensional quantities such as $H_O : \beta_i = 0$ 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. Also see the large sample theory given in Olive et al. (2025).

Remark 2. 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 (2025) theorem gives the large sample theory for $\hat{\boldsymbol{\eta}} = \widehat{\text{Cov}}(\boldsymbol{x}, Y)$. This theory needs $\boldsymbol{\eta} = \boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$ to exist for $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}$ to be a consistent estimator of $\boldsymbol{\eta}$. Let $\boldsymbol{x}_i = (x_{i1}, ..., x_{ip})^T$ and let $\boldsymbol{w}_i$ and $\boldsymbol{z}_i$ be defined below where

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

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

Theorem 5. 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 $\boldsymbol{\eta} = \boldsymbol{\Sigma_{x,Y}}$. Then a)

$$\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma_w}), \;\; \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma_w}), \tag{6.6}$$

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

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 $k \leq p$, assume $H_0 : \boldsymbol{A\beta}_{OPLS} = \boldsymbol{0}$ is true, and assume $\hat{\lambda} \xrightarrow{P} \lambda \neq 0$. Then

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) \xrightarrow{D} N_k(\boldsymbol{0}, \lambda^2 \boldsymbol{A\Sigma_w A}^T). \tag{6.7}$$

6.2 TESTING

As noted by Olive and Zhang (2025), 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 $\mathrm{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}) = \mathrm{Cov}(\hat{\boldsymbol{\eta}}) = \boldsymbol{\Sigma_w}$ 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 $\mathrm{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y})$, but we can get good nonsingular estimators of $\mathrm{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}) = \mathrm{Cov}((\hat{\eta}_{i1}, ..., \hat{\eta}_{ik})^T)$ with $\boldsymbol{u} = (x_{i1}, ..., x_{ik})^T$ where $n \geq Jk$ with $J \geq 10$. (Values of J much larger than 10 may be needed if some of the k predictors and/or Y are skewed.) Simply apply Theorem 5 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 $\sigma_{x_i,Y} = \mathrm{Cov}(x_i, Y)$.

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 $\mathrm{Cov}(\boldsymbol{x}, Y)$ if $\mathrm{Cov}(\boldsymbol{x}, v) + \mathrm{Cov}(\boldsymbol{u}, Y) + \mathrm{Cov}(\boldsymbol{u}, v) = \boldsymbol{0}$, 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 $\boldsymbol{\eta}^T\boldsymbol{x}$.

CHAPTER 7

ESTIMATING $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$ FOR SOME CENSORED SURVIVAL REGRESSION MODELS

For many survival regression data sets,

$$Z = \log(Y) = \alpha_Z + \boldsymbol{\beta}_Z^T \boldsymbol{x} + e \tag{7.1}$$

follows a multiple linear regression model. The AFT models are a special case where the model is fit using the MLE. If the chosen model is incorrect, e.g. a Weibull AFT is fit when a lognormal AFT should have been used, then a nonparametric method will often perform better than the incorrect parametric model. The Buckley and James (1979) estimator is a nonparametric competitor for the parametric AFTs. When there is no censoring, this estimator is equivalent to the ordinary least squares (OLS) estimator for multiple linear regression.

Definition 7. The Buckley James estimator $(\hat{\alpha}_{BJ}, \hat{\boldsymbol{\beta}}_{BJ})$ is a nonparametric survival regression method for models of the form (7.1).

Let the log transformation $Z_i = \log(Y_i)$ where $Y_i > 0$ is the survival time. This transformation often results in a linear model with heterogeneity:

$$Z_i = \alpha_Z + \boldsymbol{x}_i^T \boldsymbol{\beta}_Z + e_i \tag{7.2}$$

where the e_i are independent with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma_i^2$. For the AFT and the Buckley James estimator, the variance is constant: $V(e_i) = \sigma^2$ does not depend on i.

For more on estimators for model (7.2), see, for example, Heller and Simonoff (1990), Lai and Ying (1991), Lin and Wei (1992), and Yu, Liu, and Chen (2024).

The Harrell (2015) `rms` library is useful for the Buckley James estimator. See the *R* code below.

```
#download R version 4.4.4 2024
```

```
install.packages("rms")
#lognormal AFT = OLS model without censoring if z=log(y)
p<- 5
k<-2
n<-100
q <- p-1
b <- 0 * 1:q
b[1:k] <- 1 #b[1:0] acts like b[1:1] = b[1]
beta <- c(1,b)
x <- matrix(rnorm(n * q), nrow = n, ncol = q)
z <- 1 + x %*% b + rnorm(n)
#beta = (1,1,0,0)
#z = log(y)
y <- exp(z) #lognormal so positive
status <- 0*1:n + 1   #uncensored
tdata <- as.data.frame(cbind(x,y,status))
names(tdata)  #renamed y as V5,
#likely incorrectly uses V5 as a predictor
 "V1"     "V2"     "V3"     "V4"     "V5"     "status"
names(tdata) <- c("V1","V2","V3","V4","y","status")
library(rms)
#bj(Surv(y,status)~.,data=tdata) gives an error
bj(Surv(y,status)~V1+V2+V3+V4,data=tdata)
Buckley-James Censored Data Regression
bj(formula = Surv(y, status) ~ V1 + V2 + V3 + V4, data = tdata)
                                    Discrimination
                                          Indexes
```

```
Obs    100    Regression d.f.4    g        1.375
Events 100          sigma1.0172    gr       3.956
                    d.f.     95
          Coef    S.E.   Wald Z Pr(>|Z|)
Intercept  0.9659 0.1039  9.30  <0.0001
V1         0.8462 0.1055  8.02  <0.0001
V2         0.8732 0.1053  8.30  <0.0001
V3        -0.1606 0.1022 -1.57  0.1159
V4        -0.1441 0.1138 -1.27  0.2053


lsfit(x,z)$coef
Intercept         X1         X2         X3         X4
0.9659463  0.8461943  0.8731705 -0.1606320 -0.1441290
#same with uncensored data
```

Let $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = \text{Cov}(\boldsymbol{x}, Z) = E[(\boldsymbol{x} - E(\boldsymbol{x}))(Z - E(Z))]$. Let $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \text{Cov}(\boldsymbol{x}) = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T]$ be the covariance matrix of $\boldsymbol{x}$. Suppose the cases $(\boldsymbol{x}_i, Y_i)$ are iid from some population. Let the ordinary least squares (OLS) estimator be $\hat{\boldsymbol{\beta}}_{OLS}$. Since model (7.2) is a multiple linear regression (MLR) model, under mild regularity conditions, $\boldsymbol{\beta}_Z = \boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Z}$. Thus $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = \text{Cov}(\boldsymbol{x})\boldsymbol{\beta}_Z = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_Z$. When the response Y_i is censored, several models give consistent estimators $\hat{\boldsymbol{\beta}}_Z$ of $\boldsymbol{\beta}_Z$. Hence

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_Z. \tag{7.3}$$

If a Weibull regression data set is generated with parameter vector $\boldsymbol{\beta}_P$, then the Weibull AFT parameter vector $\boldsymbol{\beta} = \boldsymbol{\beta}_Z = \boldsymbol{\beta}_A = -\sigma\boldsymbol{\beta}_P = -(1/\gamma)\boldsymbol{\beta}_p$. Hence $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = -\gamma\text{Cov}(\boldsymbol{x})\boldsymbol{\beta}_P$. The `survpack` function *BJcovxz* generates a Weibull regression data set with right censored survival times using a method similar to that of Zhou (2001). Then $\boldsymbol{\beta}_A = -(1/\gamma, ..., 1/\gamma, 0, ..., 0)^T$ with $p - k$ zeroes and $\boldsymbol{\beta}_P = (1, ..., 1, 0, ..., 0)^T$ with k ones and $p - k$ zeroes. The population $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_A$ is computed, as

well as $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$ using the uncensored Z_i that are known since the data was simulated. The estimators $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}(A) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_A$ and $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}(B) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_{BJ}$ were also computed. The output below illustrates the estimators.

```
library(rms)
library(survival)
source("http://parker.ad.siu.edu/Olive/survpack.txt")
BJcovxz(n=100,p=4,k=1,psi=0)
$k
[1] 1
$betaA
[1] -1  0  0  0
$bhatwaft  #Weibull AFT
         V1          V2          V3          V4
-1.08161987 -0.12461451  0.18940525  0.03343348
$bhatbj   #Buckley James estimator
         V1          V2          V3          V4
-1.04721364 -0.21876873  0.26093084 -0.09383167
$C   pop    cov(x,z)   AFT est      BJ est
     [,1]        [,2]         [,3]        [,4]
[1,]   -1 -1.17851098 -1.177787472 -1.12324943
[2,]    0 -0.21657094 -0.163023329 -0.24175567
[3,]    0  0.05959813  0.005348028  0.09104925
[4,]    0 -0.01629190  0.098819495 -0.03343374

BJcovxz(n=100,p=4,k=1,psi=0.9)
$k
[1] 1
```

```
$betaA
[1] -1  0  0  0
$bhatwaft  #not too godd for psi > 0.8
        V1          V2          V3          V4
 1.0725896 -3.7225410  0.6832606  0.8570338
$bhatbj
       V1         V2         V3         V4
 2.190735 -3.996070  1.308376 -0.705631

$C         pop    cov(x,z)   AFT est   BJ est
            [,1]       [,2]       [,3]       [,4]
[1,] -1.0000000 -1.205574 -1.158267 -1.259813
[2,] -0.9970845 -1.229621 -1.183020 -1.289138
[3,] -0.9970845 -1.219214 -1.172632 -1.276607
[4,] -0.9970845 -1.220622 -1.168735 -1.277573
```

The last two estimators of $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$ are nonparametric, but require consistent estimators of $\boldsymbol{\beta}_Z = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Z}$, which occurs, for example, if the the cases $(\boldsymbol{x}_i, Y_i)$ are iid from some population with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ and covariance vector $\boldsymbol{\Sigma}_{\boldsymbol{x}Z}$. The survival times Y_i can be right censored, but the predictor variables $x_1, ..., x_p$ are not censored. Note that the predictor variables hat have the highest absolute correlation with Z have the highest values of $|\widehat{\text{Cov}}(x_i, Z)|/\sqrt{\hat{V}(x_i)}$.

In the literature, there are several estimators for the correlation $cor(X, Y)$ where X and Y are survival times. These estimators usually use the MLE or multiple imputation assuming that (X, Y) are iid from a bivariate normal distribution. See, for example, Barchard and Russell (2024), Li, Gillespie, Shedden, and Gillespie (2018), and Lyles, Fan, and Chuachoowong (2001).

Let the log transformation $Z_i = \log(Y_i)$ where $Y_i > 0$ is the survival time. This transformation

often results in a linear model with heterogeneity:

$$Z_i = \alpha_Z + \boldsymbol{x}_i^T \boldsymbol{\beta}_Z + e_i \tag{7.4}$$

where the e_i are independent with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma_i^2$. For the AFT, the variance is constant: $V(e_i) = \sigma^2$ does not depend on i. The Buckley and James (1979) model is another special case. Since model (7.1) is an MLR model, under mild regularity conditions, $\boldsymbol{\beta}_Z = \boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Z}$. Thus $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}_Z$. When the response Y_i is censored, several models give consistent estimators $\hat{\boldsymbol{\beta}}_Z$ of $\boldsymbol{\beta}_Z$. Hence

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \hat{\boldsymbol{\beta}}_Z. \tag{7.5}$$

Another application of the Buckley James estimator is to check AFTs. Make an EE plot of $ESPBJ = \hat{\boldsymbol{\beta}}_{BJ}^T \boldsymbol{x}$ versus $ESPA = \hat{\boldsymbol{\beta}}_{AFT}^T \boldsymbol{x}$. For the Weibull AFT, also plot $ESPPH = \hat{\sigma} \hat{\boldsymbol{\beta}}_{PH}^T \boldsymbol{x}$ versus the above two ESPs.

The ovarian cancer data is from Collett (2003, p. 187-190) and Edmunson et al. (1979). The response variable is the survival time of $n = 26$ ovarian cancer patients in days with predictors *age* in years and *treat* (1 for cyclophosphamide alone and 2 for cyclophosphamide combined with adriamycin). See Figure 7.1 for the three EE plots for the ovarian cancer data, where ESPW=ESPA. The Weibull AFT appears to be appropriate for this data set.

```
#models for the ovarian cancer data
library(survival)
library(rms)
source("http://parker.ad.siu.edu/Olive/survdata.txt")

z <- survreg(Surv(ovar[,1],ovar[,2])~ovar[,3]+ovar[,4],
dist="weibull")
zc <- coxph(Surv(ovar[,1],ovar[,2])~ovar[,3]+ovar[,4])
sighat<-z$scale
zx <- cbind(ovar[,3],ovar[,4])
ESPPH <- -sighat*zx%*%zc$coef
ESPW <- zx%*%z$coef[-1]

ovardatf <- as.data.frame(ovar)
names(ovardatf)
#[1] "time"   "status" "treat"  "age"

outbj <- bj(formula = Surv(time, status) ~ treat + age,
data = ovardatf)
ESPBJ <- zx%*%outbj$coef[-1]

par(mfrow = c(3, 1))
plot(ESPPH,ESPW)
abline(0,1)
plot(ESPBJ,ESPW)
abline(0,1)
plot(ESPBJ,ESPPH)
```

```
abline(0,1)

par(mfrow=c(1,1))
```

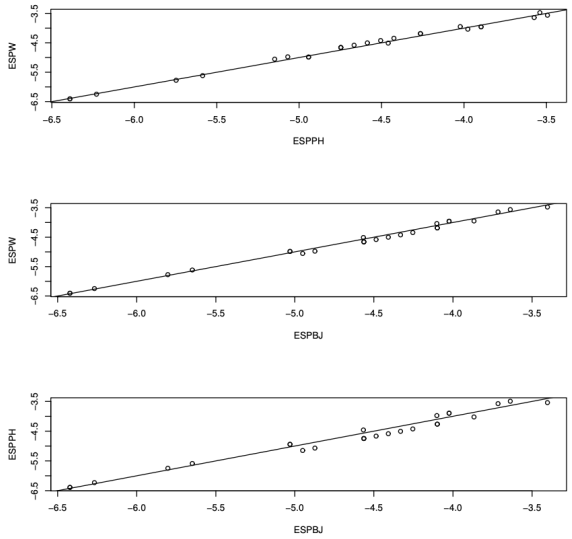

Figure 7.1. Three EE Plots for the Ovarian Cancer Data

CHAPTER 8

EXAMPLE AND SIMULATIONS

Example 1. 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 x_1 and predictors x_2 = the *population of the district in 1843*, x_3 = the *number of married civilian men* in the district, x_4 = the *number of married men in the military* in the district, and x_5 = 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_3 are highly correlated but not equal. Similarly, x_4 and x_5 are highly correlated but not equal. Then $\hat{\boldsymbol{\beta}}_{OLS} = (242.3910, 0.00035, 0.9995, -0.2328, 0.1531)^T$, and forward selection with OLS and the C_p criterion used $\hat{\boldsymbol{\beta}}_{I,0} = (\hat{\beta}_1, 0, 1.0010, 0, 0)^T$. With the scaled data, $\hat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{w}, Y) = (242.3910, 81.0283, 40877.4086, -104.8576, 66.2739)^T$.

Next, we describe a small OLS simulation study. The simulation used $\psi = 0$ and 0.5; and $k = 1$ and $p - 1$ where k and ψ are defined in the following paragraph.

Let $\boldsymbol{x} = (1\ \boldsymbol{u}^T)^T$ where $\boldsymbol{u}$ is the $(p-1) \times 1$ vector of nontrivial predictors. Let $Y = \alpha + \boldsymbol{\phi}^T\boldsymbol{u} + e$ with $\boldsymbol{\beta} = (\alpha, \boldsymbol{\phi}^T)^T$. In the simulations, for $i = 1, ..., n$, we generated $\boldsymbol{w}_i \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$ where the $m = p - 1$ elements of the vector $\boldsymbol{w}_i$ are independent and identically distributed (iid) N(0,1). Let the $m \times m$ matrix $\boldsymbol{A} = (a_{ij})$ with $a_{ii} = 1$ and $a_{ij} = \psi$ where $0 \leq \psi < 1$ for $i \neq j$. Then the vector $\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{w}_i$ so that $Cov(\boldsymbol{u}_i) = \boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{A}^T = (\sigma_{ij})$ 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 $Y_i = 1 + 1x_{i,2} + \cdots + 1x_{i,k+1} + e_i$ for $i = 1, ..., n$. Hence $\alpha = 1$ and $\boldsymbol{\beta} = (1, .., 1, 0, ..., 0)^T$ with k 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.

When $\psi = 0$, the OLS confidence intervals for β_i should have length near $2t_{96,0.975}\sigma/\sqrt{n}$. Hence the scaled CI length $= \sqrt{n}$ CI length $\approx 2(1.96)\sigma = 3.92\sigma$ when the iid zero mean errors have variance σ^2. The simulation gave the average scaled CI lengths.

For the unscaled predictors, the simulation computed the large sample 95% CIs $[L_{in}, U_{in}]$ for β_i and $i = 1, ...p$. The test for $H_0 : (\beta_{i_1}, \beta_{i_2})^T = (\beta_{i_1,0}, \beta_{i_2,0})^T$ was also performed using equation (2.2) with $\{i_1, i_2\} = \{p-1, p\}$. 5000 CIs were generated for each β_i, and the coverage was the proportion of times β_i was in its CI. Hence if β_1 was in its interval 4750/5000 = 0.95, then the observed coverage was 0.95.

For the scaled predictors, the simulation computed the "95% CIs" $[s_i L_{in}, s_i U_{in}]$ for $\sigma_i\beta_i$ and $i = 1, ...p$ with $\{i_1, i_2\} = \{p-1, p\}$. The coverage was the proportion of times $\sigma_i\beta_i$ was in its "CI." The "test" for $H_0 : (\beta_{i_1}(\boldsymbol{w}, Y), \beta_{i_2}(\boldsymbol{w}, Y))^T = (\sigma_{i_1}\beta_{i_1,0}, \sigma_{i_2}\beta_{i_2,0})^T$ was also performed using equation (2.2) on the scaled data $\boldsymbol{W}$. The "test" is a valid large sample test if $(\beta_{i_1}, \beta_{i_2})^T = (0,0)^T$. When $k = 1$, the test is valid and the "95% CI" can be used as a large sample test for $H_0 : \sigma_i\beta_i = 0$ except for β_2 since $\beta_3 = \cdots = \beta_p = 0$. When $k = p-1$ the "test" and "95% CIs" are not valid large sample tests and CIs (except for β_1). The undercoverage can be rather large when the test is not valid.

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

mlrsim(n=100,p=5,k=1,nruns=5000,type=1,psi=0,indices=c(4,5))
$cicov
[1] 0.9488 0.9452 0.9536 0.9482 0.9540
$avelen
[1] 4.038598 4.067681 4.065098 4.063435 4.070485
$testcov
[1] 0.9526
```

```
$beta
[1] 1 1 0 0 0
$cicovs
[1] 0.9488 0.8874 0.9536 0.9482 0.9540
$avelens
[1] 4.038598 4.038187 4.039593 4.039284 4.037956
$testcovs
[1] 0.9526
$betas
[1] 1 1 0 0 0
mlrsim(n=100,p=5,k=1,nruns=5000,type=1,psi=0.5,indices=c(4,5))
$cicov
[1] 0.9484 0.9530 0.9484 0.9510 0.9514
$avelen
[1] 4.041254 7.101501 7.107276 7.090973 7.104132
$testcov
[1] 0.9504
$beta
[1] 1 1 0 0 0
$cicovs
[1] 0.9484 0.9332 0.9484 0.9510 0.9514
$avelens
[1] 4.041254 9.373663 9.380725 9.365281 9.379010
$testcovs
[1] 0.9504
$betas
[1] 1.000000 1.322876 0.000000 0.000000 0.000000
```

```
##repeat with types 2,3,4,5
```

Table 8.1. n=100,p=5,indices=(4,5), k=1

psi	etype	β_1	β_2	β_3	β_4	β_5	testcov
0, cov	1	0.9488	0.9452	0.9536	0.9482	0.9540	0.9526
u, len		4.0386	4.0676	4.0651	4.0634	4.0705	
0, cov	1	0.9488	0.8874	0.9536	0.9482	0.9540	0.9526
s, len		4.0386	4.0382	4.0396	4.0393	4.0380	
0.5, cov	1	0.9484	0.9530	0.9484	0.9510	0.9514	0.9504
u, len		4.0413	7.1015	7.1073	7.0910	7.1041	
0.5, cov	1	0.9484	0.9332	0.9484	0.9510	0.9514	0.9504
s, len		4.0413	9.3737	9.3807	9.3653	9.3790	

Each table has 4 lines for each type. The first line gives the coverages for the β_i while the second line gives the scaled CI lengths. There is a length for testcov since the test corresponds to a confidence region instead of a confidence interval. The third and fourth lines are for the scaled data where cov is the proportion of times $\sigma_i\beta_i$ was in its interval. With 5000 runs, coverage between 0.94 and 0.96 suggests that the actual coverage is near the nominal large sample coverage of 0.95.

For Table 8.1, H_0 is true except for the scaled data with $\sigma_2\beta_2$. With error type 1 and psi = $\psi = 0$, the average scaled CI lengths were near 4.07 which is not too far from 3.92 considering that $n = 100$ and $p = 5$. In the third line under β_2, the coverage is 0.8874. With $\psi = 0.5$, the sixth line under β_2 has coverage 0.9333. Increasing ψ often decreased the undercoverage.

```
mlrsim(n=100,p=5,k=4,nruns=5000,type=1,psi=0,indices=c(4,5))
$cicov
[1] 0.9448 0.9448 0.9536 0.9492 0.9500
$avelen
[1] 4.041916 4.073015 4.079712 4.069030 4.068888
$testcov
[1] 0.9528
```

```
$beta
[1] 1 1 1 1 1
$cicovs
[1] 0.9448 0.8976 0.8984 0.8882 0.8902
$avelens
[1] 4.041916 4.042054 4.041657 4.042183 4.042372
$testcovs
[1] 0.8654
$betas
[1] 1 1 1 1 1
mlrsim(n=100,p=5,k=4,nruns=5000,type=1,psi=0.5,indices=c(4,5))
$cicov
[1] 0.9548 0.9582 0.9486 0.9530 0.9472
$avelen
[1] 4.043087 7.095204 7.106610 7.115100 7.110030
$testcov
[1] 0.9506
$beta
[1] 1 1 1 1 1
$cicovs
[1] 0.9548 0.9360 0.9354 0.9338 0.9310
$avelens
[1] 4.043087 9.355476 9.367924 9.372169 9.364274
$testcovs
[1] 0.913
$betas
[1] 1.000000 1.322876 1.322876 1.322876 1.322876
```

```
##repeat with types 2,3,4,5
```

Table 8.2. n=100,p=5,indices=(4,5), k=5

psi	etype	β_1	β_2	β_3	β_4	β_5	testcov
0, cov	1	0.9448	0.9448	0.9536	0.9492	0.9500	0.9528
u, len		4.0419	4.0730	4.0797	4.0690	4.0689	
0, cov	1	0.9448	0.8976	0.8984	0.8882	0.8902	0.8654
s, len		4.0419	4.0421	4.0417	4.0422	4.0424	
0.5, cov	1	0.9548	0.9582	0.9486	0.9530	0.9472	0.9506
u, len		4.0431	7.0952	7.1066	7.1151	7.1100	
0.5, cov	1	0.9548	0.9360	0.9354	0.9338	0.9310	0.9130
s, len		4.0431	9.3555	9.3679	9.3722	9.3643	

For Table 8.2 with the scaled data, H_0 is only true for β_1. For the scaled data, the "CI" undercoverage was more severe for $\psi = 0$ than for $\psi = 0.5$, and the testcov was worse than that for the CIs. With the unscaled data, H_0 was always true.

8.1 $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$ SIMULATION

```
library(survival)
library(rms)
source("http://parker.ad.siu.edu/Olive/survdata.txt")
args(BJcovxzsim)
function (n = 100, p = 4, nruns = 100, k = 1, psi = 0, a = 1,
    gam = 1, clam = 0.1)
BJcovxzsim(n=100,p=4,nruns=5000,k=1,psi=0,a=1,gam=1,clam=0.1)
$k
[1] 1
$betaA
[1] -1  0  0  0
$bhatwaft
```

```
         V1          V2          V3          V4
-1.20494486 -0.13980676 -0.10482867  0.09373143
$bhatbj
        V1          V2          V3          V4
-1.2251305 -0.1713720 -0.1115955  0.1222867
$covxz
     [,1] [,2] [,3] [,4]
[1,]   -1    0    0    0
$covxzhatmn
[1] -0.999312340 -0.002244755 -0.002427508 -0.001046906
$covxzhatsd
[1] 0.1935163 0.1619503 0.1622831 0.1629504
$covxzwaftmn
[1] -1.002959e+00 -1.391445e-03 -1.447495e-03 -3.198343e-05
$covxzwaftsd
[1] 0.1855422 0.1491370 0.1496335 0.1505240
$covxzBJmn
[1] -1.0019151047 -0.0023403747 -0.0015671645 -0.0009214999
$covxzBJsd
[1] 0.2024268 0.1687888 0.1689074 0.1693748
```

The function `BJcovxzsim` simulates estimators of $\boldsymbol{\Sigma}_{\boldsymbol{x}Z}$. The value of $\boldsymbol{\Sigma}_{\boldsymbol{x}Z}$ is given. Then 5000 runs are used to get the estimators. The means and standard deviations of the estimators are given. In the simulation, the uncensored values of Z are known. Hence the first estimator is the usual sample covariance vector $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$. For real data, only censored values of Z are known, so $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$ can not be computed. The second estimator is $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}(A) = \hat{\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}}_A$ from the Weibull AFT. The third estimator is $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}(BJ) = \hat{\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}}_{BJ}$ using the Buckley James estimator. Let $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = (\sigma_{1Z}, ..., \sigma_{pZ})^T$. Table 8.3 gives 2 lines per simulation scenario. The first line gives the means while the second line

gives the standard deviations. A value of 0+ means the absolute value was less that 0.00005.

Table 8.3. $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = (-1, 0, 0, 0)^T$

(n, p, ψ, k)	est	σ_{1Z}	σ_{2Z}	σ_{3Z}	σ_{4Z}
(100,4,0,1)	samp	-0.9993	-0.0022	-0.0024	-0.0010
	SD	0.1935	0.1620	0.1623	0.1630
(100,4,0,1)	AFT	-1.0030	-0.0014	-0.0014	0+
	SD	0.1855	0.1491	0.1496	0.1505
(100,4,0,1)	BJ	-1.0019	-0.0023	-0.0016	-0.0009
	SD	0.2024	0.1688	0.1689	0.1694

CHAPTER 9

CONCLUSIONS

For multiple linear regression with standardized predictors, OLS software tests of the form $H_0 : \boldsymbol{\beta}_O = \mathbf{0}$ are valid large sample tests where $\boldsymbol{\beta}_O = (\beta_{i_1}, ..., \beta_{i_k})^T$. However, OLS software does not give correct confidence intervals for $\beta_i(\boldsymbol{w}, Y) = \sigma_i \beta_i$ for $i = 2, ..., p$ unless $\beta_i = 0$. Note that the unstandardized data can be used for inference such as testing, confidence intervals, and confidence regions.

Software

The *R* software was used in the simulations. See R Core Team (2024). Programs are in the Olive (2025a) collections of *R* functions *slpack.txt*, available from (http://parker.ad.siu.edu/Olive/slpack.txt). The function `mlrsim` was used to make the tables.

REFERENCES

[1] Barchard, K.A., and Russell, J.A. (2024), "Distorted Correlations among Censored Data: Causes, Effects, and Correction," *Behavior Research Methods*, 56, 1207-1228.

[2] Basa, J., Cook, R.D., Forzani, L., and Marcos, M. (2024), "Asymptotic Distribution of One-Component Partial Least Squares Regression Estimators in High Dimensions," *The Canadian Journal of Statistics*, 52, 118-130.

[3] Buckley, J. and James, I. (1979), "Linear Regression with Censored data," *Biometrika*, 66, 429-436.

[4] Collett, D. (2003), *Modelling Survival Data in Medical Research*, 2nd ed., Chapman & Hall/CRC, Boca Raton, FL.

[5] Cook, R.D., and Forzani, L. (2024), *Partial Least Squares Regression: and Related Dimension Reduction Methods*, Chapman and Hall/CRC, Boca Raton, FL.

[6] 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.

[7] Edmunson, J.H., Fleming, T.R., Decker, D.G., Malkasian, G.D., Jorgenson, E.O., Jeffries, J.A., Webb, M.J., and Kvols, L.K. (1979), "Different Chemotherapeutic Sensitivities and Host Factors Affecting Prognosis in Advanced Ovarian Carcinoma Versus Minimal Residual Disease," *Cancer Treatment Reports*, 63, 241-247.

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

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

[10] Harrell, F.E. (2015), *Regression Modeling Strategies: with Applications to Linear Models, Logistic Regression, and Survival Analysis*, 2nd ed., Springer, New York, NY.

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

[12] Heller, G., and Simonoff, J. S. (1990), "A Comparison of Estimators for Regression with a

Censored Response Variable," *Biometrika*, 77, 515-520.

[13] Herath, H.M.W.B., and Samadi,S.Y. (2025), "Scaled envelope models for multivariate time series," *Journal of Multivariate Analysis*, 205,105370

[14] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021), *An Introduction to Statistical Learning with Applications in R*, 2nd ed., Springer, New York, NY.

[15] Jones, J.A., and Waller, N.G. (2015), "The Normal-Theory and Asymptotic Distribution-Free (ADF) Covariance Matrix of Standardized Regression Coefficients: Theoretical Extensions and Finite Sample Behavior," *Psychometrika*, 80, 365-378.

[16] Lai, T.L., and Ying, Z. (1991), "Large-Sample Theory of a Modified Buckley-James Estimator for Regression Analysis with Censored Data," *Annals of Statistics*, 19, 1370-1402.

[17] Li, Y., Gillespie, B.W., Shedden, K. and Gillespie, J.A. (2018), "Profile Likelihood Estimation of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data," *The R Journal,* 10, 159-179.

[18] Lin, J.S., and Wei, L.J. (1992), "Linear Regression Analysis Based on Buckley-James Estimating Equation," *Biometrics*, 48, 679-681.

[19] Lyles, R.H., Fan, D., and Chuachoowong, R. (2001), "Correlation Coefficient Estimation Involving a Left Censored Laboratory Assay Variable," *Statistics in Medicine*, 20, 2921-2933.

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

[21] Olive, D.J. (2025a), *Prediction and Statistical Learning*, online course notes, see (http://parker.ad.siu.edu/Olive/slearnbk.htm).

[22] Olive, D.J. (2025b), *Survival Analysis*, online course notes, see (http://parker.ad.siu.edu/Olive/survbk.htm).

[23] Olive, D.J., Alshammari, A.A., Pathiranage, K.G., and Hettige, L.A.W. (2025), "Testing with the One Component Partial Least Squares and the Marginal Maximum Likelihood Estimators," is at (http://parker.ad.siu.edu/Olive/pphdwls.pdf).

[24] Olive, D.J., and Zhang, L. (2025), "One Component Partial Least Squares, High Dimensional Regression, Data Splitting, and the Multitude of Models," *Communications in Statistics: The-*

ory and Methods, 54, 130-145.

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

[26] Schielzeth, H. (2010), "Simple Means to Improve Interpretability of Regression Coefficients," *Methods in Ecology and Evolution*, 1, 103-111.

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

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

[29] van Ginkel, J.R. (2020), "Standardized Regression Coefficients and Newly Proposed Estimators for R^2 in Multiply Imputed Data," *Psychometrika*, 85, 185-205.

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

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

[32] Yu, L., Liu, L., and Chen, D.G. (2024), "Extending Buckley James Method for Heteroscedastic Survival Data," *Journal of Statistical Computation and Simulation*, 94, 1776-1792.

[33] Yuan, K.H., and Chan, W. (2011), "Biases and Standard Errors of Standardized Regression Coefficients," *Psychometrika*, 76, 670-690.

[34] Zhou, M. (2001), "Understanding the Cox Regression Models with Time–Change Covariates," *The American Statistician*, 55, 153-155.

VITA

Graduate School
Southern Illinois University Carbondale

Sanjuka Johana Lemonge

sanjuka.johanalemonge@siu.edu

University of Kelaniya, Sri Lanka
Bachelor of Science, (Honors) Statistics, March 2019

Southern Illinois University, Carbondale, IL, USA
Master of Science, Mathematics, August 2022

Special Honors and Awards:
Dissertation Research Assistantship Award (Summer 2025)

John M.H. Olmsted Outstanding Master's Teaching Assistant award (2022)

Dissertation Paper Title:
OLS testing with predictors scaled to have unit sample variance

Major Professor: Dr. David Olive

Publications:

Dassanayake, D. A. R. C.; Wijayawardhana, H. N. A. M.; Prematthilaka, L. M.; Senarath, S. A. K. B.; Weerasekara, W. M. D. D.; Thilakarathne, K. H. K. D. S.; Goonetilleke, H. R.; Sandaruwan, B. H. S.; Winisinghe, W. T. D.; Kumara, J. L. S. M.; Attanayake, A. M. C. H. (2017). Study on the selection of mobile service providers: Special reference to the students at Faculty of Science, University of Kelaniya. International Research Symposium on Pure and Applied Sciences, 2017 Faculty of Science, University of Kelaniya, Sri Lanka. p73.http://repository.kln.ac.lk/handle/123456789/18201

Kumara, J.L.S.M. & Chandrasekara, N.V. (2019). Influence of the Factors Associated with Suicides: A Case Study for the Kelaniya Police Division in Sri Lanka. 12th International Research Conference 2019, General Sir John Kotelawala Defence University, Sri Lanka. http://ir.kdu.ac.lk/handle/345/2356

Kumara, J.L.S.M., Weerasinghe, W.P.M.C.N., Saumyamala, M.G.A., Sachithra, S.A.L. & Chandrasekara, N.V. (2019). Time Series Modelling of Open Market Retail Price of Red Onions in Colombo. 12th International Research Conference 2019, General Sir John Kotelawala Defence University, Sri Lanka. http://ir.kdu.ac.lk/handle/345/2358