TESTING POISSON REGRESSION AND RELATED MODELS WITH THE ONE COMPONENT PARTIAL LEAST SQUARES ESTIMATOR

by

Paul Amarh Quaye

M.S., Southern Illinois University Edwardsville, 2021

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

TESTING POISSON REGRESSION AND RELATED MODELS WITH THE ONE COMPONENT PARTIAL LEAST SQUARES ESTIMATOR

by

Paul Amarh Quaye

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

Approved by:
David J. Olive, PhD, Chair
Seyed Yaser Samadi, PhD
Jennifer Koran, PhD
Michael Sullivan, PhD
Mathew Gluck, PhD

Graduate School
Southern Illinois University Carbondale
June 27, 2025

AN ABSTRACT OF THE DISSERTATION OF

Paul Amarh Quaye, for the Doctor of Philosophy degree in Mathematics, presented on June 27, 2025, at Southern Illinois University Carbondale.

TITLE: TESTING POISSON REGRESSION AND RELATED MODELS WITH THE ONE COMPONENT PARTIAL LEAST SQUARES ESTIMATOR

MAJOR PROFESSOR: Prof. David J. Olive

Poisson regression, negative binomial regression, and related regression methods are often used when the response variable is a count. A log transformation often results in a linear model with heterogeneity. Then testing can be done with the one component partial least squares estimator for multiple linear regression, including some high dimensional tests. For prediction, a simple method that uses information from several estimators, is also considered.

KEY WORDS: Data splitting, dimension reduction, high dimensional data, negative binomial regression, lasso.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my mentor, Prof. David J. Olive, for his tremendous help, steadfast support, and insightful advice throughout this dissertation. His guidance and inspiration have been truly invaluable, and I feel incredibly fortunate to have him as my PhD. advisor.

My profound gratitude also goes to the five members of my dissertation committee for their patience and understanding throughout the process of working on this dissertation. My heartfelt appreciation extends to the professors and staff of the School of Mathematical and Statistical Sciences at Southern Illinois University for their support and guidance throughout my time as a graduate student.

Lastly, I would like to thank Prof. Henri Schurz for his support and encouragement. Your belief in me has been a constant source of motivation, and I could not have achieved this without you.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

This section reviews regression models where the nonnegative integer count response variable is Y that is independent of the $p \times 1$ vector of predictors $\boldsymbol{x} = (x_1, ..., x_p)^T$ given $\boldsymbol{x}^T\boldsymbol{\beta}$, written $Y \perp\!\!\!\perp \boldsymbol{x}|\boldsymbol{x}^T\boldsymbol{\beta}$. Then there are n cases $(Y_i, \boldsymbol{x}_i^T)^T$, and the sufficient predictor $SP = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}$. For the regression models, the conditioning and subscripts, such as i, will often be suppressed. A useful *Poisson regression (PR) model* is $Y \sim \text{Poisson}\left(\text{e}^{\text{SP}}\right)$. This model has $E(Y|SP) = V(Y|SP) = \exp(SP)$.

Some notation is needed for the negative binomial regression model. If Y has a (generalized) negative binomial distribution, $Y \sim NB(\mu, \kappa)$, then the probability mass function (pmf) of Y is

$$P(Y = y) = \frac{\Gamma(y + \kappa)}{\Gamma(\kappa)\Gamma(y + 1)} \left(\frac{\kappa}{\mu + \kappa}\right)^{\kappa} \left(1 - \frac{\kappa}{\mu + \kappa}\right)^{y}$$

for $y = 0, 1, 2, ...$ where $\mu > 0$ and $\kappa > 0$. Then $E(Y) = \mu$ and $\text{V}(Y) = \mu + \mu^2/\kappa$.

The *negative binomial regression model* states that $Y_1, ..., Y_n$ are independent random variables with

$$Y|SP \sim \text{NB}(\exp(\text{SP}), \kappa).$$

This model has $E(Y|SP) = \exp(SP)$ and

$$V(Y|SP) = \exp(SP)\left(1 + \frac{\exp(SP)}{\kappa}\right) = \exp(SP) + \tau \exp(2\ SP).$$

Following Agresti (2002, p. 560), as $\tau \equiv 1/\kappa \to 0$, it can be shown that the negative binomial regression model converges to the Poisson regression model.

The *quasi-Poisson regression model* has $E(Y|SP) = \exp(SP)$ and $V(Y|SP) = \phi \exp(SP)$ where the dispersion parameter $\phi > 0$. Note that this model and the Poisson regression model have the same conditional mean function, and the conditional variance functions are the same if $\phi = 1$.

Next, some notation is needed for the zero truncated Poisson regression model. See Olive (2017, pp. 430–431). Y has a zero truncated Poisson distribution, $Y \sim ZTP(\mu)$, if the probability mass function of Y is

$$f(y) = \frac{e^{-\mu}\,\mu^y}{(1-e^{-\mu})\,y!}$$

for $y = 1, 2, 3, ...$ where $\mu > 0$. The ZTP pmf is obtained from a Poisson distribution where $y = 0$ values are truncated, so not allowed. Now $E(Y) = \mu/(1-e^{-\mu})$, and

$$V(Y) = \frac{\mu^2+\mu}{1-e^{-\mu}} - \left(\frac{\mu}{1-e^{-\mu}}\right)^2.$$

The *zero truncated Poisson regression* model has $Y|SP \sim ZTP(\exp(SP))$. Hence the parameter $\mu(SP) = \exp(SP)$,

$$E(Y|SP) = \frac{\exp(SP)}{1-\exp(-\exp(SP))}, \text{ and}$$

$$V(Y|SP) = \frac{[\exp(SP)]^2+\exp(SP)}{1-\exp(-\exp(SP))} - \left(\frac{\exp(SP)}{1-\exp(-\exp(SP))}\right)^2.$$

Other alternatives include the zero truncated negative binomial regression model, the hurdle or zero inflated Poisson regression model, and the hurdle or zero inflated negative binomial regression model. See Zuur et al. (2009), Simonoff (2003), and Hilbe (2011).

Variable selection estimators include forward selection or backward elimination when $n \geq 10p$. When n/p is not large, the Chen and Chen (2008) EBIC criterion with forward selection can be useful. Sparse regression methods can also be used for variable selection even if n/p is not large: the regression submodel, such as a Nelder and Wedderburn (1972) generalized linear model (GLM), uses the predictors that had nonzero sparse regression estimated coefficients. For Poisson regression, these methods include lasso and elastic net. See Friedman et al. (2007), Friedman, Hastie, and Tibshirani (2010), Tibshirani (1996), and Zou and Hastie (2005).

Following Olive and Hawkins (2005), a *model for variable selection* can be described by

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_E^T\boldsymbol{\beta}_E = \boldsymbol{x}_S^T\boldsymbol{\beta}_S \tag{1.1}$$

where $\boldsymbol{x} = (\boldsymbol{x}_S^T, \boldsymbol{x}_E^T)^T$, $\boldsymbol{x}_S$ is an $a_S \times 1$ vector, and $\boldsymbol{x}_E$ is a $(p - a_S) \times 1$ vector. Given that $\boldsymbol{x}_S$ is in the model, $\boldsymbol{\beta}_E = \boldsymbol{0}$ and E denotes the subset of terms that can be eliminated given that the subset S is in the model. Let $\boldsymbol{x}_I$ be the vector of a terms from a candidate subset indexed by I, and let $\boldsymbol{x}_O$ be the vector of the remaining predictors (out of the candidate submodel). Suppose that S is a subset of I and that model (1.1) holds. Then

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S = \boldsymbol{x}_I^T\boldsymbol{\beta}_I + \boldsymbol{x}_O^T\boldsymbol{0} = \boldsymbol{x}_I^T\boldsymbol{\beta}_I.$$

Thus $\boldsymbol{\beta}_O = \boldsymbol{0}$ if $S \subseteq I$. The model using $\boldsymbol{x}^T\boldsymbol{\beta}$ is the full model.

To clarify notation, suppose $p = 3$, a constant α is always in the model, and $\boldsymbol{\beta} = (\beta_1, 0, 0)^T$. Then the $J = 2^p = 8$ possible subsets of $\{1, 2, ..., p\}$ are $I_1 = \emptyset$, $S = I_2 = \{1\}$, $I_3 = \{2\}$, $I_4 = \{3\}$, $I_5 = \{1, 2\}$, $I_6 = \{1, 3\}$, $I_7 = \{2, 3\}$, and $I_8 = \{1, 2, 3\}$. There are $2^{p-a_S} = 4$ subsets I_2, I_5, I_6, and I_8 such that $S \subseteq I_j$. Let $\hat{\boldsymbol{\beta}}_{I_7} = (\hat{\beta}_2, \hat{\beta}_3)^T$ and $\boldsymbol{x}_{I_7} = (x_2, x_3)^T$.

Let I_{min} correspond to the set of predictors selected by a variable selection method such as forward selection or lasso variable selection. If $\hat{\boldsymbol{\beta}}_I$ is $a \times 1$, use zero padding to form the $p \times 1$ vector $\hat{\boldsymbol{\beta}}_{I,0}$ from $\hat{\boldsymbol{\beta}}_I$ by adding 0s corresponding to the omitted variables. For example, if $p = 4$ and $\hat{\boldsymbol{\beta}}_{I_{min}} = (\hat{\beta}_1, \hat{\beta}_3)^T$, then the observed variable selection estimator $\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_{min},0} = (\hat{\beta}_1, 0, \hat{\beta}_3, 0)^T$. As a statistic, $\hat{\boldsymbol{\beta}}_{VS} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities $\pi_{kn} = P(I_{min} = I_k)$ for $k = 1, ..., J$ where there are J subsets, e.g. $J = 2^p$.

Theory for the variable selection estimator $\hat{\boldsymbol{\beta}}_{VS}$ is complicated. See Pelawa Watagoda and Olive (2023) for multiple linear regression, and Rathnayake and Olive (2023) for models such as generalized linear models. For fixed p, these two papers showed that $\hat{\boldsymbol{\beta}}_{VS}$ is $\sqrt{n}$ consistent with a complicated nonnormal limiting distribution.

Let the log transformation $Z_i = \log(Y_i)$ if $Y_i > 0$ and $Z_i = \log(0.5)$ if $Y_i = 0$. This transformation often results in a linear model with heterogeneity:

$$Z_i = \alpha_Z + \boldsymbol{x}_i^T\boldsymbol{\beta}_Z + e_i \tag{1.2}$$

where the e_i are independent with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma_i^2$. For Poisson regression, the minimum chi-square estimator is the weighted least squares estimator from the regression of Z_i on $\boldsymbol{x}_i$ with weights $w_i = e^{Z_i}$. See Agresti (2002, pp. 611–612) and Olive (2013, 2017: pp. 406–407).

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

$$Y_i = \alpha + x_{i,1}\beta_1 + \cdots + x_{i,p}\beta_p + e_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{1.3}$$

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

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e}, \tag{1.4}$$

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 (1.4) 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

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

and

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

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

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

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

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

$$\lambda = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T\boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}} \text{ and } \hat{\lambda} = \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}}{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}\mathrm{Y}}} \tag{1.5}$$

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. Thus the OPLS regression of Z_i on $\boldsymbol{x}_i$ is useful to model (1.2).

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = [diag(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}})]^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x},Y}. \tag{1.6}$$

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

$$\hat{\boldsymbol{\beta}}_{MMLE} = \hat{\boldsymbol{\beta}}_{MMLE}(\boldsymbol{t}, Y) = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \boldsymbol{I}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{t},Y} = \hat{\boldsymbol{\eta}}_{OPLS}(\boldsymbol{t}, Y) \tag{1.7}$$

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.

If the regression model for Y depends on $\boldsymbol{x}$ only through $\alpha + \boldsymbol{\beta}^T \boldsymbol{x}$, and if the predictors $\boldsymbol{x}_i$ are independent and identically distributed (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. Results from Cameron and Trivedi (1998, p. 89) suggest that if a Poisson regression model is fit using OLS software for multiple linear regression, then a rough approximation is $\hat{\boldsymbol{\beta}}_{PR} \approx \hat{\boldsymbol{\beta}}_{OLS} / \overline{Y}$.

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, Olive and Zhang (2025) and 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.

CHAPTER 2

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 1. 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 $\hat{\boldsymbol{\eta}} = \widehat{\mathrm{Cov}}(\boldsymbol{x}, Y)$. 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

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

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

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

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

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

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

c) Let $\boldsymbol{A}$ be a $k \times p$ full rank constant matrix with $k \leq p$, assume $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$ is true, and assume $\hat{\lambda} \xrightarrow{P} \lambda \neq 0$. Then

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

2.1 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 $H_0 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A}\boldsymbol{\beta}_{OPLS} \neq \boldsymbol{0}$ is equivalent to testing $H_0 : \boldsymbol{A}\boldsymbol{\eta} = \boldsymbol{0}$ versus $H_1 : \boldsymbol{A}\boldsymbol{\eta} \neq \boldsymbol{0}$ where $\boldsymbol{A}$ is a $k \times p$ constant matrix. Let $\text{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}) = \text{Cov}(\hat{\boldsymbol{\eta}}) = \boldsymbol{\Sigma}_{\boldsymbol{W}}$ be the asymptotic covariance matrix of $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$. In high dimensions where $n < 5p$, we can't get a good nonsingular estimator of $\text{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y})$, but we can get good nonsingular estimators of $\text{Cov}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{u}Y}) = \text{Cov}((\hat{\eta}_{i1}, ..., \hat{\eta}_{ik})^T)$ with $\boldsymbol{u} = (x_{i1}, ..., x_{ik})^T$ where $n \geq Jk$ with $J \geq 10$. (Values of J much larger than 10 may be needed if some of the k predictors and/or Y are skewed.) Simply apply Theorem 1 to the predictors $\boldsymbol{u}$ used in the hypothesis test, and thus use the sample covariance matrix 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 $\boldsymbol{\eta}^T\boldsymbol{x}$.

Zhao et al. (2024) have an interesting result for the multiple linear regression model (1.3). Assume that the cases $(\boldsymbol{x}_i^T, Y_i)^T$ are iid with $E(Y) = \mu_Y$, $E(\boldsymbol{x}) = \boldsymbol{\mu_x}$ and nonsingular $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma_x}$. Then testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ is equivalent to testing $H_0 : \boldsymbol{\mu} = \mathbf{0}$ versus $H_1 : \boldsymbol{\mu} \neq \mathbf{0}$ with $\boldsymbol{\mu} = E(\boldsymbol{z}_i) = \boldsymbol{\Sigma_x}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ where $\boldsymbol{z}_i = (\boldsymbol{x}_i - \boldsymbol{\mu_x})(Y_i - \mu_Y - (\boldsymbol{x}_i - \boldsymbol{\mu_x})^T\boldsymbol{\beta}_0)$, and a one sample test can be applied to $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y} - (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T\boldsymbol{\beta}_0)$. Since $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$, $\boldsymbol{\beta} = \mathbf{0}$ if and only if $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$, using $\boldsymbol{\beta}_0 = \mathbf{0}$ gives tests for $H_0 : \boldsymbol{\beta} = \mathbf{0}$, $H_0 : \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \mathbf{0}$, $H_0 : \boldsymbol{\beta}_{OPLS} = \mathbf{0}$, and $H_0 : \boldsymbol{\beta}_{kPLS} = \mathbf{0}$, the k-component partial least squares estimator. Abid and Olive (2025) simplify the theory for the one sample test used by Zhao et al. (2024), resulting in an estimator that is quick to compute when H_0 is true.

CHAPTER 3

INCORPORATING INFORMATION FROM SEVERAL REGRESSION ESTIMATORS

The theory and tests from the previous section can be applied to model (1.2) with Z replacing Y.

There are several ways to compute k-component partial least squares (PLS) estimators for multiple linear regression. A simple way is to do the OLS regression on $W_1, ..., W_k$ where $W_j = \hat{\boldsymbol{\eta}}_j^T \boldsymbol{x}$ and $\hat{\boldsymbol{\eta}}_j = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{j-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$, and $k \leq min(n-2, p)$. Then the one component PLS estimator is OPLS while the 3-component PLS estimator regresses Y on $W_1 = \hat{\boldsymbol{\eta}}_1^T \boldsymbol{x} = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^T \boldsymbol{x}$, $W_2 = \hat{\boldsymbol{\eta}}_2^T \boldsymbol{x} = [\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}]^T \boldsymbol{x}$, and $W_3 = \hat{\boldsymbol{\eta}}_3^T \boldsymbol{x} = [\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}]^T \boldsymbol{x}$. See Helland (1990).

This result suggests computing $W_i = \hat{\boldsymbol{\eta}}_i^T \boldsymbol{x}$ for $i = 1, ..., J$ and fit the OLS model that regresses Z on the W_i or, for example, the Poisson regression model that regresses Y on the W_i. Some interesting choices are $\hat{\boldsymbol{\eta}}_1 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$, $\hat{\boldsymbol{\eta}}_2 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$, $\hat{\boldsymbol{\eta}}_3 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Z}$, $\hat{\boldsymbol{\eta}}_4 = \hat{\boldsymbol{\beta}}_L(\boldsymbol{x}, Z)$= the lasso estimator from regressing Z on $\boldsymbol{x}$, $\hat{\boldsymbol{\eta}}_5 = \hat{\boldsymbol{\beta}}_{RR}(\boldsymbol{x}, Z)$= the ridge regression estimator from regressing Z on $\boldsymbol{x}$, $\hat{\boldsymbol{\eta}}_6 = \hat{\boldsymbol{\beta}}_{LPR}(\boldsymbol{x}, Y)$= the lasso Poisson regression estimator from regressing Y on $\boldsymbol{x}$. Let $\boldsymbol{x}_I$ denote the set of variables selected using $\hat{\boldsymbol{\eta}}_4$. Then $\hat{\boldsymbol{\eta}}_7 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I Z}$, $\hat{\boldsymbol{\eta}}_8 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I Z}$, $\hat{\boldsymbol{\eta}}_9 = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I}^2 \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}_I Z}$, $\hat{\boldsymbol{\eta}}_{10} = \hat{\boldsymbol{\beta}}_{RR}(\boldsymbol{x}_I, Z)$= the ridge regression estimator from regressing Z on $\boldsymbol{x}_I$. Other good choices can easily be obtained. For example, let $\boldsymbol{x}_G$ denote the set of variables selected using $\hat{\boldsymbol{\eta}}_6$.

CHAPTER 4

PLS FOR MULTIPLE LINEAR REGRESSION

This section gives more information about the k-component partial least squares estimators for multiple linear regression $Y_i = \alpha + \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ where the $\boldsymbol{x}_i$ are $p \times 1$ vectors of predictors for $i = 1, ..., n$. As mentioned in the previous section, the k-component partial least squares estimator $\hat{\boldsymbol{\beta}}_{kPLS}$ can be found from the OLS regression on a constant and $W_1, ..., W_k$ where $W_j = \hat{\boldsymbol{\eta}}_j^T\boldsymbol{x}$ and $\hat{\boldsymbol{\eta}}_j = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{j-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}$, and $k \leq min(n-2, p)$. Note that $\hat{\boldsymbol{\beta}}_{OPLS} = \hat{\boldsymbol{\beta}}_{1PLS}$ with $k = 1$ and $\hat{\boldsymbol{\beta}}_{OLS} = \hat{\boldsymbol{\beta}}_{pPLS}$ with $k = p$ if $n > p+1$. Let $\boldsymbol{X}_c$ be the matrix of centered predictors (subtract the sample mean from each predictor) so that $\boldsymbol{D} = \boldsymbol{X}_c^T\boldsymbol{X}_x = (n-1)\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}$ and let $\boldsymbol{y}$ be the vector of centered response variables. Let $\boldsymbol{d} = \boldsymbol{X}_c^T\boldsymbol{y} = (n-1)\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$. The model selection PLS estimator is $\hat{\boldsymbol{\beta}}_{MSPLS}$, and selects k^* with some criterion, such as 10-fold cross validation.

An equivalent way to compute the k-component PLS estimator is to find unit vectors $\hat{\boldsymbol{\eta}}_1, ..., \hat{\boldsymbol{\eta}}_k$ and perform the OLS regression of Y on a constant and the $U_i = \hat{\boldsymbol{\eta}}_i^T\boldsymbol{x}$ for $i = 1, ..., k$. Following Brown (1993, pp. 71-72), first maximize $(\mathbf{c}^T\boldsymbol{d})^2$ subject to the constraint $\mathbf{c}^T\ \mathbf{c} = \|\mathbf{c}\|^2 = 1$. The maximum occurs at $\mathbf{c}_1 = \hat{\boldsymbol{\eta}}_1 = \boldsymbol{d}/\|\boldsymbol{d}\| = \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}/\|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}\| = \hat{\boldsymbol{\eta}}_{OPLS}/\|\hat{\boldsymbol{\eta}}_{OPLS}\|$. Then $\mathbf{c_2} = \hat{\boldsymbol{\eta}}_{\mathbf{2}}$ is found by maximizing $(\mathbf{c}^T\boldsymbol{d})^2$ subject to both $\|\mathbf{c}\| = 1$ and $\mathbf{c}^T\boldsymbol{D}\mathbf{c_1} = 0$ (called $\boldsymbol{D}$-norm orthogonalization) to get $\mathbf{c_2} = \hat{\boldsymbol{\eta}}_{\mathbf{2}}$. Continue in this way to get the remaining vectors $\mathbf{c_3}, ..., \mathbf{c_k}$.

Let $Y = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}_{kPLS} + \epsilon$ be a working model. Let $\boldsymbol{X} = (\mathbf{1}\ \boldsymbol{X}_1)$. An equivalent way to formulate PLS is to form $\boldsymbol{b}_j$ iteratively where $\boldsymbol{b}_k = \arg\max_{\boldsymbol{b}}\{[corr(\boldsymbol{Y}, \boldsymbol{X}_1\boldsymbol{b})]^2 V(\boldsymbol{X}_1\boldsymbol{b})\}$ subject to $\boldsymbol{b}^T\boldsymbol{b} = 1$ and $\boldsymbol{b}^T\boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{b}_j = 0$ for $j = 1, ..., k-1$. Let the $\hat{\boldsymbol{b}}_j$ be the estimates of $\boldsymbol{b}_j$, and perform the OLS regression of $\boldsymbol{Y}$ on $\boldsymbol{X}_1\hat{\boldsymbol{C}}_{k,n}$ and a constant where $\hat{\boldsymbol{C}}_{k,n} = [\hat{\boldsymbol{b}}_1, ..., \hat{\boldsymbol{b}}_k]$ to find $\hat{\boldsymbol{\gamma}}_k$. Then $\hat{\boldsymbol{\beta}}_{kPLS} = \hat{\boldsymbol{C}}_{k,n}\hat{\boldsymbol{\gamma}}_k$.

Again let $Y = \alpha + \boldsymbol{x}^T\boldsymbol{\beta}_{kPLS} + \epsilon$ be a working model. From Naik and Tsai (2000), Helland and Almøy (1994), and Helland (1990), let

$$\hat{\boldsymbol{A}}_{k,n}^T = [\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}, \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^2\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}, ..., \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}^{k-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}].$$

Let $\boldsymbol{w} = \hat{\boldsymbol{A}}_{k,n}\boldsymbol{x}$ with $Y = \alpha + \boldsymbol{w}^T\boldsymbol{\gamma}_k + \epsilon$ the working model so $\hat{\boldsymbol{\beta}}_{kPLS} = \hat{\boldsymbol{A}}_{k,n}^T\hat{\boldsymbol{\gamma}}_k$. Then

$$\hat{\boldsymbol{\beta}}_{kPLS} = \hat{\boldsymbol{A}}_{k,n}^T[\hat{\boldsymbol{A}}_{k,n}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{A}}_{k,n}^T]^{-1}\hat{\boldsymbol{A}}_{k,n}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} = \hat{\boldsymbol{A}}_{k,n}^T[\hat{\boldsymbol{A}}_{k,n}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{A}}_{k,n}^T]^{-1}\hat{\boldsymbol{A}}_{k,n}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}}\hat{\boldsymbol{\beta}}_{OLS}(\boldsymbol{x}, Y).$$

The Mevik et al. (2015) `pls` library is useful for computing PLS.

Much of the PLS literature claims that if the cases are iid, then under mild conditions, $\hat{\boldsymbol{\beta}}_{OPLS}$, $\hat{\boldsymbol{\beta}}_{kPLS}$, and $\hat{\boldsymbol{\beta}}_{MSOPLS}$ estimate $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$. See for example, Basa et al. (2024) and Cook and Forzani (2024). However, they use a very strong regularity condition:

$$Y|\boldsymbol{x} = \alpha_{OPLS} + \boldsymbol{\beta}_{OPLS}^T\boldsymbol{x} + e. \tag{4.1}$$

When $Y|\boldsymbol{x} = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e$, then under mild regularity conditions, $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$. Hence regularity condition (4.1) and iid cases forces $\boldsymbol{\beta}_{OLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \lambda\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\beta}_{OPLS}$. Thus regularity condition (4.1) forces $\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ and $\boldsymbol{\beta}_{OLS} = \lambda\boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ to be eigenvectors of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ if $\lambda \neq 0$. Hence $\boldsymbol{\beta}_{OLS}^T\boldsymbol{x}$ is equivalent (up to a positive constant multiplier) to the population principal component regression (PCR) component $\boldsymbol{\eta}_j^T\boldsymbol{x}$ that is most correlated with Y, where $\boldsymbol{\eta}_j$ is one of the eigenvectors of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$.

Olive and Zhang (2025) noted that under iid cases and $Y|\boldsymbol{x} = \alpha + \boldsymbol{\beta}^T\boldsymbol{x} + e$, then under mild regularity conditions, $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$ and $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}$, and typically $\hat{\boldsymbol{\beta}}_{kPLS}$ estimates $\boldsymbol{\beta}_{kPLS} \neq \boldsymbol{\beta}_{OLS}$ if $k < p$. In particular, typically $\boldsymbol{\beta}_{OPLS} \neq \boldsymbol{\beta}_{OLS}$.

To see some problems with the claim that $\boldsymbol{\beta}_{OPLS} = \boldsymbol{\beta}_{OLS}$, this paragraph and the following two paragraphs are taken from Olive and Zhang (2025). Consider multiple linear regression with $\text{Cov}(\boldsymbol{x}) = diag(1, 2, ..., p)$. First consider OPLS with $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OPLS}$. Then at most one element of $\text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x},Y}$ is nonzero since $\boldsymbol{\Sigma}_{\boldsymbol{x},Y}$ is an eigenvector of $\text{Cov}(\boldsymbol{x})$. Hence at most one predictor is correlated with Y, regardless of the value of p. This restriction is too strong.

If the cases are iid from a multivariate normal distribution, then $Y|\boldsymbol{x} = \alpha_{OLS} + \boldsymbol{\beta}_{OLS}^T\boldsymbol{x} + e$ and $Y|\boldsymbol{\beta}_{OPLS}^T\boldsymbol{x} = \alpha_{OPLS} + \boldsymbol{\beta}_{OPLS}^T\boldsymbol{x} + e$ are both linear models by Olive and Zhang (2025) where e depends on the model. Since $\boldsymbol{\beta}_{OPLS} = \boldsymbol{\beta}_{OLS}$ forces $\boldsymbol{\beta}_{OLS}$ to be an eigenvector of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, *if* $\boldsymbol{\beta}_{OLS} \neq \boldsymbol{0}$ *is not an eigenvector of* $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, *then* $\boldsymbol{\beta}_{OPLS} \neq \boldsymbol{\beta}_{OLS}$. For a computational example, let $\boldsymbol{x} \sim N_p(\boldsymbol{0}, diag(1, 2, 3, 4))$

with $\boldsymbol{\Sigma}_{\boldsymbol{x}} = diag(1, 2, 3, 4)$, and let the population generating model be $Y_i = x_{i1} + x_{i2} + e_i$ for $i = 1, ..., n$ where the e_i are iid $N(0, 1)$ and independent of the $\boldsymbol{x}_i$. Then $\alpha = 0$ and $\boldsymbol{\beta} = (1, 1, 0, 0)^T$. Hence $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta} = (1, 1, 0, 0)^T$, $\boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \boldsymbol{\Sigma}_{\boldsymbol{x}}\boldsymbol{\beta}_{OLS} = (1, 2, 0, 0)^T$, and

$$\lambda = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x},Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x},Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x},Y}^T \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x},Y}} = 5/9.$$

Thus $\boldsymbol{\beta}_{OPLS} = \lambda \boldsymbol{\Sigma}_{\boldsymbol{x},Y} = \lambda \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\beta}_{OLS} = (5/9, 10/9, 0, 0)^T \neq \boldsymbol{\beta}_{OLS}$. Thus OLS and OPLS usually give different valid population multiple linear regression models with $\boldsymbol{\beta}_{OPLS} \neq \boldsymbol{\beta}_{OLS}$. However, model $Y | \boldsymbol{\beta}_{OPLS}^T \boldsymbol{x} = \alpha_{OPLS} + \boldsymbol{\beta}_{OPLS}^T \boldsymbol{x} + e$ is often a useful multiple linear regression model with large sample theory given in Section 2. Thus the claims in the OPLS literature that $\boldsymbol{\beta}_{OLS} = \boldsymbol{\beta}_{OPLS} =$ an eigenvector of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ under mild regularity conditions are incorrect.

In the OLS literature, $\boldsymbol{\beta}_{OLS}$ can be any vector in $\mathbb{R}^p$. If $\boldsymbol{\beta}_{OLS}$, $\boldsymbol{\Sigma}_{\boldsymbol{x},Y}$, and $\boldsymbol{\beta}_{OPLS}$ were restricted to be eigenvectors of $\boldsymbol{\Sigma}_{\boldsymbol{x}}$, then the OLS and OPLS estimators would often not fit the data well.

In the PCR literature, several principal components usually need to be used: there is rarely one principal component that is highly correlated with Y.

In PLS talks for multiple linear regression, empirically the model selection estimator that used $k^* = 3$ tends to work better than $k^* = 1$. If $\boldsymbol{\beta}_{OPLS} = \boldsymbol{\beta}_{OLS}$, then the other PLS components $\hat{\boldsymbol{\eta}}_i^T \boldsymbol{x}$ would be nuisance variables that would degrade the fit rather than improve the fit.

CHAPTER 5

EXAMPLE AND SIMULATIONS

Next, we describe a small simulation study. Let $\boldsymbol{x} \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$ be the $(p-1) \times 1$ vector of nontrivial predictors. Let $ESP_i = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i = 1 + 1x_{i,1} + \cdots + 1x_{i,k}$ for $i = 1, ..., n$. Hence $\alpha = 1$ and $\boldsymbol{\phi} = (\alpha, \boldsymbol{\beta}^T)^T = (1, .., 1, 0, ..., 0)^T$ with $k+1$ ones and $p-k-1$ zeros. Here $\boldsymbol{\beta}$ is the Poisson regression parameter vector $\boldsymbol{\beta}_{PR}$ or the negative binomial regression parameter vector $\boldsymbol{\beta}_{NBR}$. Let $Z_i = \log(Y_i)$ if $Y_i > 0$ and $Z_i = \log(0.5)$ if $Y_i = 0$. Then a multiple linear regression model with heterogeneity is $Z_i = \alpha_Z + \boldsymbol{x}_i^T \boldsymbol{\beta}_Z + e_i$ where the e_i are independent with expected value $E(e_i) = 0$ and variance $V(e_i) = \sigma_i^2$. Since the cases $(\boldsymbol{x}_i, Y_i)$ are iid, the OLS estimator $\boldsymbol{\beta}_{OLS} = c_o \boldsymbol{\beta} = \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{x}Z} = \boldsymbol{\Sigma}_{\boldsymbol{x}Z}$ because $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \boldsymbol{I}_{p-1}$. Thus $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = (c_o, .., c_o, 0, ..., 0)^T$ with the first k values equal to c_o and $p-k-1$ zeros.

Let $\boldsymbol{\eta}_{OPLS} = \boldsymbol{\Sigma}_{\boldsymbol{x}Z} = (\eta_1, ..., \eta_{p-1})^T$. Then the Theorem 1 large sample $100(1-\delta)$ CI is $\hat{\eta}_i \pm t_{n-1,1-\delta/2} SE(\hat{\eta}_i)$ could be computed for each η_i. If 0 is not in the confidence interval, then $H_0 : \eta_i = 0$ and $H_0 : \beta_{iE} = 0$ are both rejected for estimators $E =$ OPLS and MMLE for the multiple linear regression model with Z. 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 zero 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 ≥ 0.9, 0.92, 0.93, 0.94, and 0.96 were also recorded. The test $H_0 : (\eta_i, \eta_j)^T = (0,0)^T$ was also done where H_0 was true. The coverage of the test was recorded and a correction factor was not used. Negative binomial regression and Poisson regression were used, where $\kappa = \infty$ indicates that Poisson regression was used.

Table 5.1. Cov(x, Z), n=100, p=100, k=1, κ = 0.5,1,10,100,1000,10000

κ	mincov	cov90	cov92	cov93	cov94	cov96	testcov
∞	0.0000	0.9899	0.9899	0.9899	0.9899	0.7071	0.9454
len	0.4159	0.4184	0.4865				
0.5	0.0038	0.9899	0.9899	0.9899	0.9899	0.6970	0.9416
len	0.5060	0.5081	0.5687				
1	0.0000	0.9899	0.9899	0.9899	0.9899	0.6667	0.9398
len	0.4807	0.4833	0.5431				
10	0.0000	0.9899	0.9899	0.9899	0.9899	0.6667	0.9370
len	0.4253	0.4281	0.4942				
100	0.0000	0.9899	0.9899	0.9899	0.9899	0.7172	0.9452
len	0.4178	0.4197	0.4904				
1000	0.0000	0.9899	0.9899	0.9899	0.9899	0.6465	0.9350
len	0.4164	0.4181	0.4865				
10000	0.0000	0.9899	0.9899	0.9899	0.9899	0.7071	0.9406
len	0.4162	0.4186	0.4872				

Table 5.1 illustrates Theorem 1a) with Z replacing Y and $k = 1$. Confidence intervals were made for $\eta_i = Cov(x_i, Z)$ for $i = 1, ..., 99$ and the coverage was the percentage of the 5000 CIs that contained 0. Here $\eta_1 \neq 0$, but $\eta_i = 0$ for $i = 2, ..., 99$. The first two lines of Table 5.1 correspond to Poisson regression. The confidence interval for η_1 never contained 0, hence the minimum coverage was 0 with observed power = $1-0 = 1$. The proportion of CIs that had coverage ≥ 0.94 was 0.9899 (98/99 CIs). Hence this was also the proportion of CIs with coverage $\geq 0.90, 0.92$ and 0.93. The proportion of CIs that had coverage ≥ 0.96 was 0.7071 (70/99 CIs). The typical coverage was near 0.965, hence the correction factor was slightly too large. The test $H_0 : (\eta_{98}, \eta_{99})^T = (0,0)^T$ did not

use a correction factor, and coverage was 0.9454. The minimum average CI length was 0.4159, the sample mean of the average CI lengths was 0.4184, and the maximum average length was 0.4865, corresponding to η_1. The second two lines and below for Table 5.1 were for the negative binomial regression with kappa = $\kappa = 0.5, 1, 10, 100, 1000, 10000$.

Table 5.2. Cov(x,Y), n=100, p=100, k=1, κ = 0.5,1,10,100,1000,10000

κ	mincov	cov90	cov92	cov93	cov94	cov96	testcov
∞	0.0160	0.9899	0.9899	0.9899	0.9899	0.9899	0.9540
len	2.0406	2.0776	4.0484				
0.5	0.1668	0.9899	0.9899	0.9899	0.9899	0.9899	0.9588
len	3.6015	3.6879	6.7636				
1	0.0804	0.9899	0.9899	0.9899	0.9899	0.9899	0.9548
len	2.9737	3.0525	5.6898				
10	0.0200	0.9899	0.9899	0.9899	0.9899	0.9899	0.9486
len	2.1610	2.2139	4.3127				
100	0.0172	0.9899	0.9899	0.9899	0.9899	0.9899	0.9586
len	2.0786	2.1165	4.2023				
1000	0.0122	0.9899	0.9899	0.9899	0.9899	0.9899	0.9482
len	2.0429	2.0781	4.0509				
10000	0.0108	0.9899	0.9899	0.9899	0.9899	0.9899	0.9542
len	2.0441	2.0811	4.0519				

Table 5.2 illustrates Theorem 1a) without Z replacing Y (Y stays the same) and $k = 1$. Confidence intervals were made for $\eta_i = Cov(x_i, Y)$ for $i = 1, ..., 99$ and the coverage was the percentage of the 5000 CIs that contained 0. Here $\eta_1 \neq 0$, but $\eta_i = 0$ for $i = 2, ..., 99$. The first two lines of Table 5.2 correspond to Poisson regression. The confidence interval for η_1 never contained 0, hence the minimum coverage was 0.0108 with observed power = 1 − 0.0108 = 0.9892. The proportion of CIs that had coverage $\geq$ 0.94 was 0.9899 (98/99 CIs). Hence this was also the proportion of CIs with coverage $\geq 0.90, 0.92$ and 0.93. The proportion of CIs that had coverage $\geq$ 0.96 was 0.9899 (98/99 CIs). The typical coverage was near 0.965, hence the correction factor was slightly too large. The test $H_0 : (\eta_{98}, \eta_{99})^T = (0, 0)^T$ did not use a correction factor, and coverage was 0.954. The minimum average CI length was 2.0406 , the sample mean of the average CI lengths was 2.0776, and the maximum average length was 4.0484, corresponding to η_1. The

second two lines and below for Table 5.2 were for the negative binomial regression with kappa $= \kappa = 0.5, 1, 10, 100, 1000, 10000$. Upon comparing Table 5.1 and Table 5.2, it is evident that the average lengths for $Cov(x_i, Z)$ and $Cov(x_i, Y)$ differ. This difference is quite understandable, considering that Z represents a logarithmic transformation of Y.

CHAPTER 6

REAL DATA EXAMPLES

From Chapter 5, we aim to apply Poisson regression to real data examples and visualize the OD plots for overdispersion. One such example is Figure 6.1 from Cook and Weisberg (1999, pp. 285-286), which involves species data with a sample size of $n = 29$ and utilizes 7 variables based on the following: Y representing the number of species, endem denoting the number of endemic species, area indicating the island size, elev for the island elevation, distnear as the distance to the nearest island, distsc representing the distance to Santa Cruz, and areanear for the area of the nearest island. The data was sourced from *source*("$http://parker.ad.siu.edu/Olive/sldata.txt$"). Although the dataset is quite small, the plots still provide fairly good results.

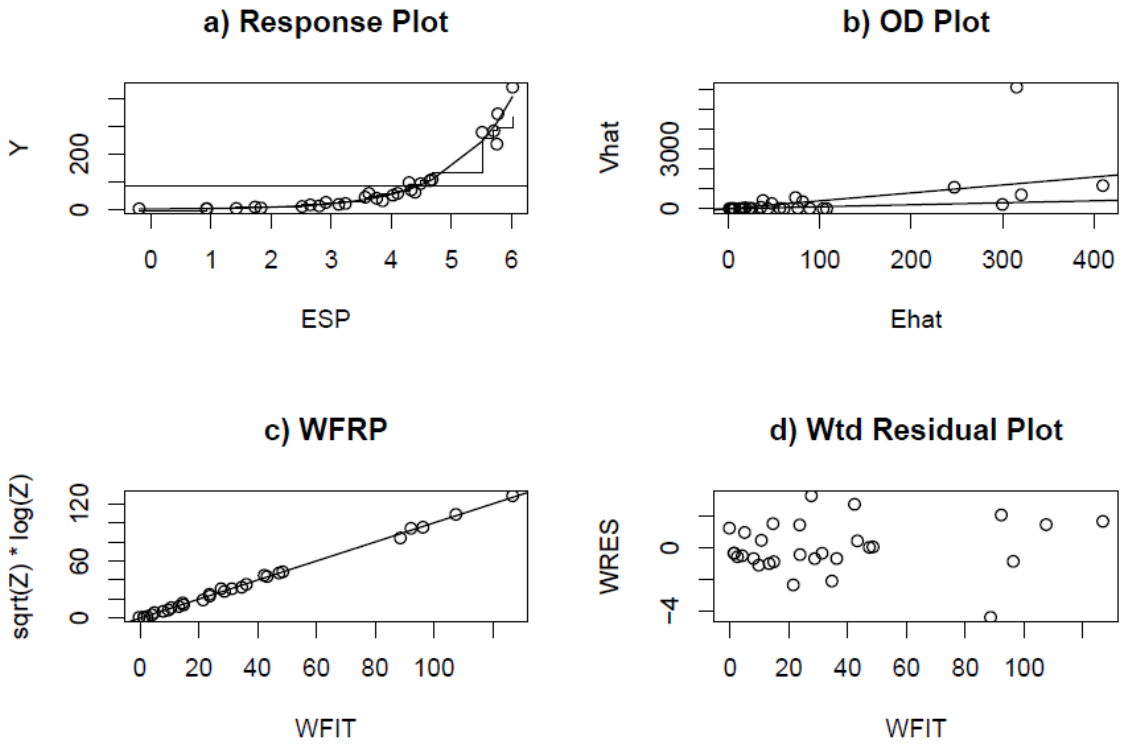

Figure 6.1. Plot for species data

Figure 6.1 presents 4 plots. In the response plot of Figure 6.1a, the lowess curve appears jagged to differentiate it from the estimated mean function. The horizontal line corresponds to the sample mean $\overline{Y}$. The OD plot in Figure 6.1b indicates the presence of overdispersion, as the vertical scale is ten times that of the horizontal scale, and few points are notably large and positioned above the slope 4 line. We observe that the Poisson mean can be approximately estimated by identifying the value of the exponential curve in Figure 6.1a. For instance, at when ESP = 5.5, $Y \approx$ Poisson(250). We also examined the MLR response plot, which appears roughly linear but not ideal. Figure 6.2 displays the MLR plot, which demonstrates roughly linear relationship.

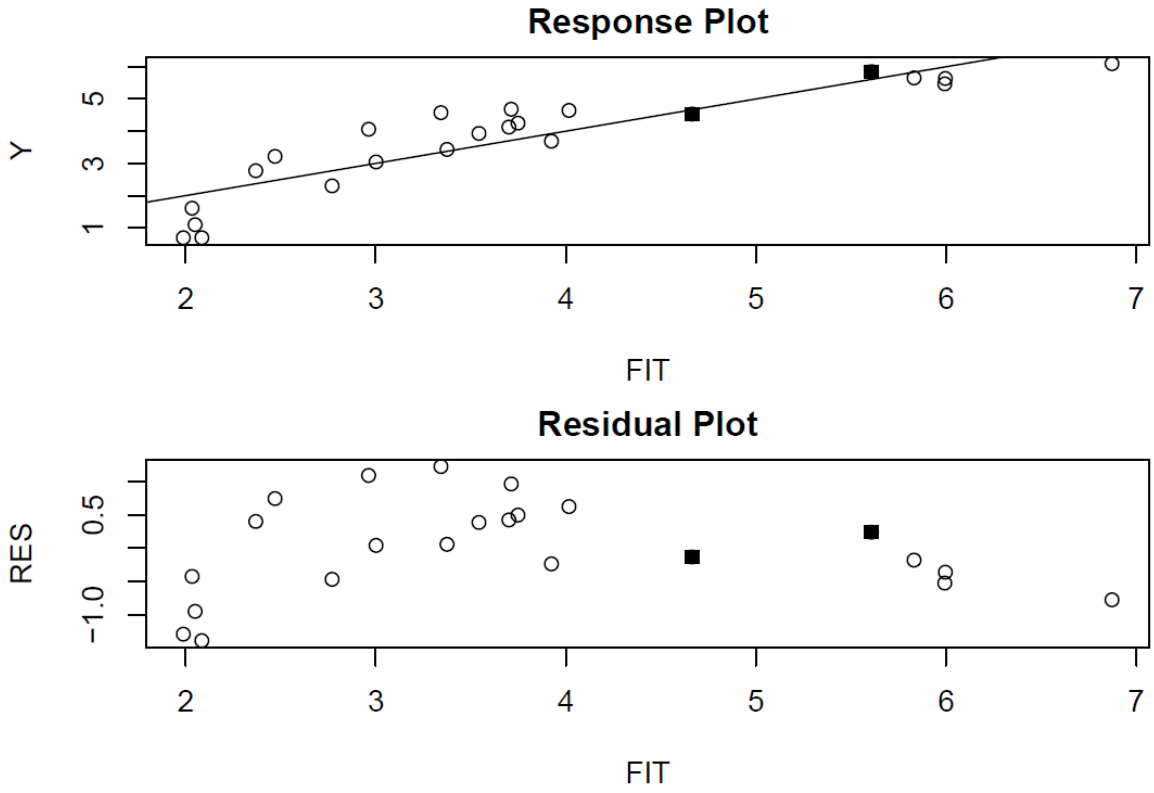

Figure 6.2. MLR plot for species data

In Figure 6.2, we applied some transformations to the predictors to enhance the fit shown in the MLR plot in Figure 6.4. Figure 6.3 presents 4 plots. In the response plot of Figure 6.3a, the lowess curve appears jagged to differentiate it from the estimated mean function. The horizontal line corresponds to the sample mean $\overline{Y}$. The OD plot in Figure 6.3b indicates that there is a little presence of overdispersion, as the vertical scale is less than ten times that of the horizontal scale,

and few points are notably large and positioned above the slope 4 line. We observe that the Poisson mean can be approximately estimated by identifying the value of the exponential curve in Figure 6.3a. For instance, at when ESP = 5.5, $Y \approx$ Poisson(250).

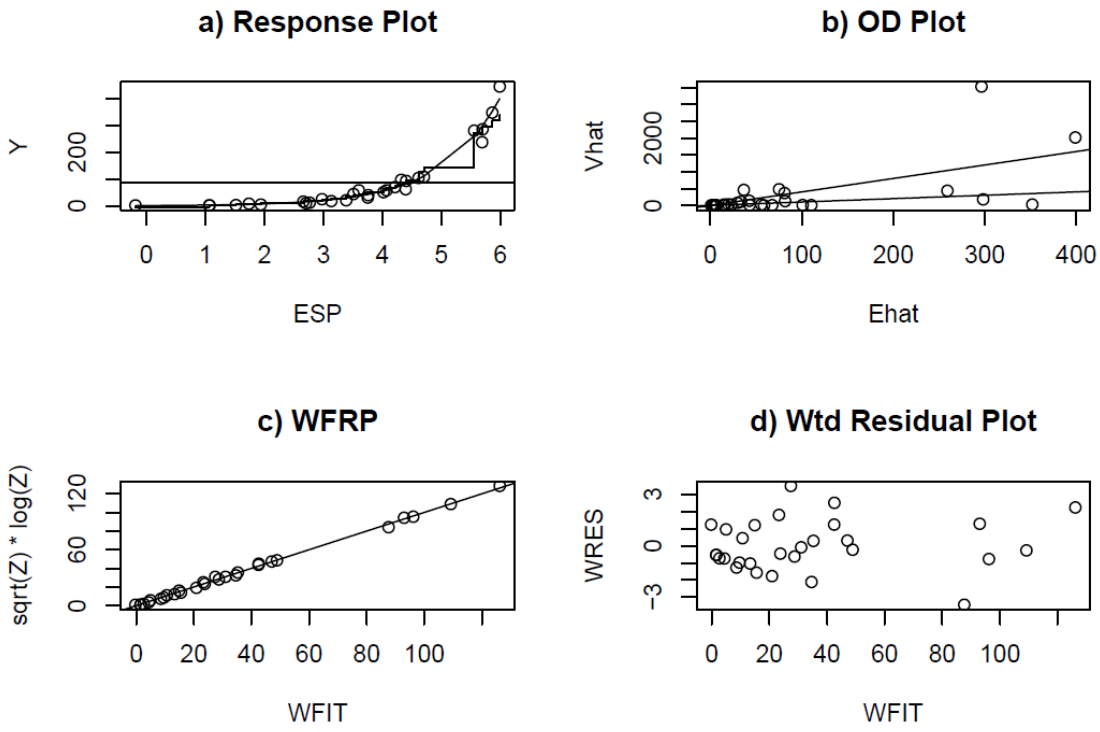

Figure 6.3. Plot for the transformed species data

We also examined the MLR response plot, which appears approximately linear and therefore, transforming the predictors is necessary to improve the fit. Figure 6.4 displays the MLR plot, which demonstrates a more clearly linear relationship.

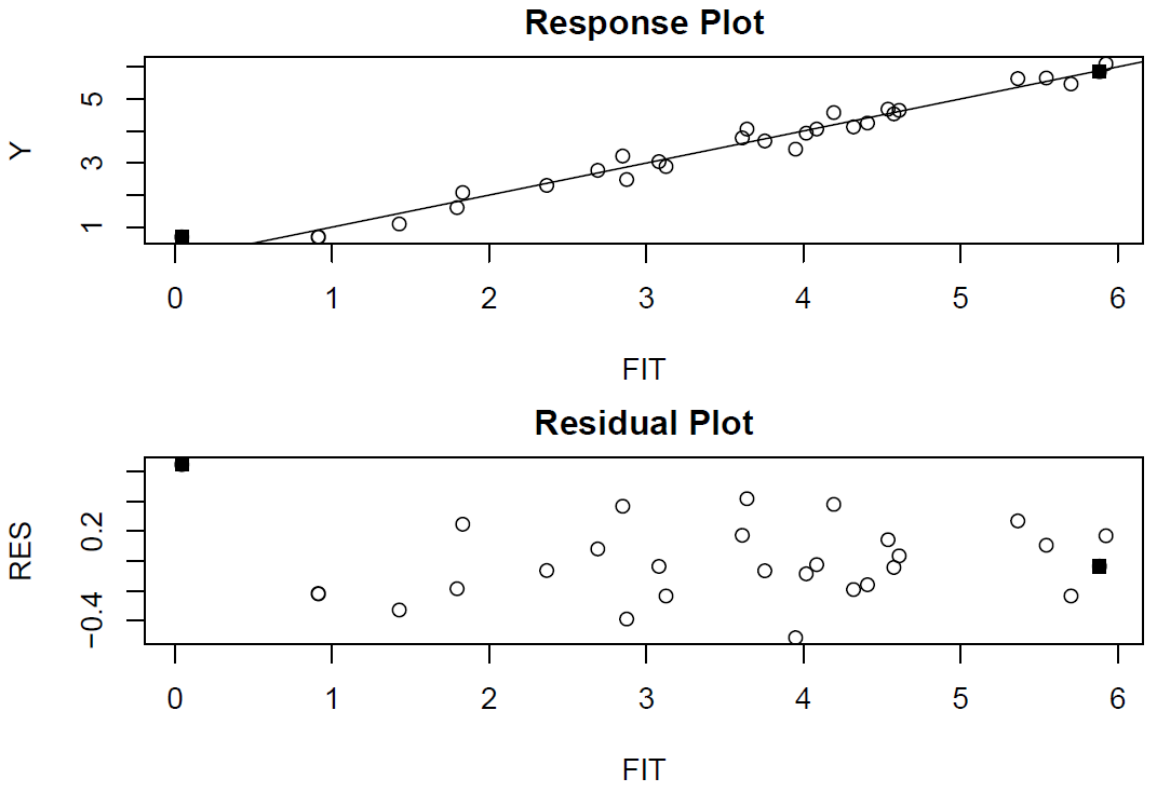

Figure 6.4. MLR plot for the transformed species data

Another example is data from Cook and Weisberg (1999, pp. 351, 433, 447) consisting of mussels data from an ecological study of horse mussels (Mussels' Muscles) in the Marlborough Sounds, located off the coast of New Zealand. The dataset includes a sample size of $n = 79$ and incorporates 4 predictors: shell height, shell width, shell length (each measured in millimeters), and shell mass (measured in grams). The response variable Y represents the mussels' muscle mass in grams. Here Y is not a count data, but log Y follows a multiple linear regression model. The data was sourced from $source(\text{"}http://parker.ad.siu.edu/Olive/sldata.txt\text{"})$.

Figure 6.5 presents 4 plots. In the response plot of Figure 6.5a, the lowess curve appears jagged to differentiate it from the estimated mean function. The horizontal line corresponds to the

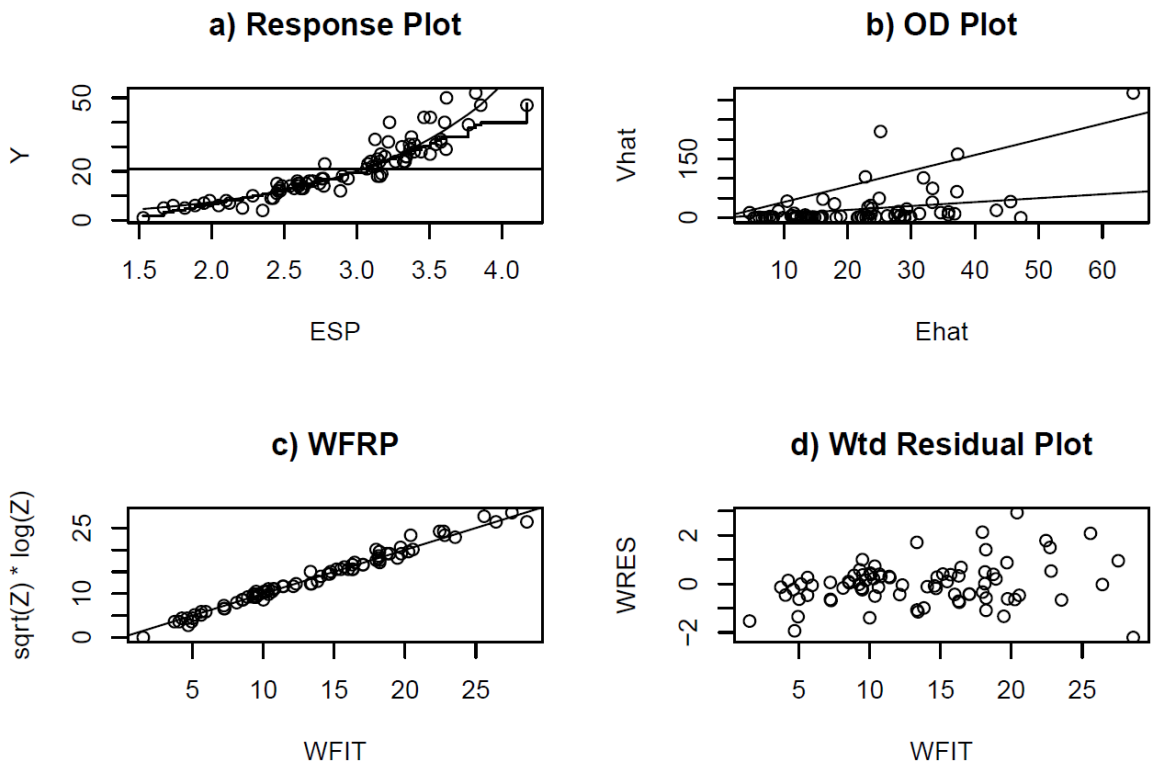

Figure 6.5. Plot for mussels' muscles data

sample mean $\overline{Y}$. The OD plot in Figure 6.5b indicates that there is a little presence of overdispersion, as the vertical scale is less than ten times that of the horizontal scale, and few points are notably large and positioned above the slope 4 line. We observe that the mean can be approximately estimated by identifying the value of the exponential curve in Figure 6.5a. For instance, at when ESP = 4, $E(Y) \approx 35$. We also examined the MLR response plot, which appears roughly linear but not ideal. Figure 6.6 displays the MLR plot, which demonstrates roughly linear relationship.

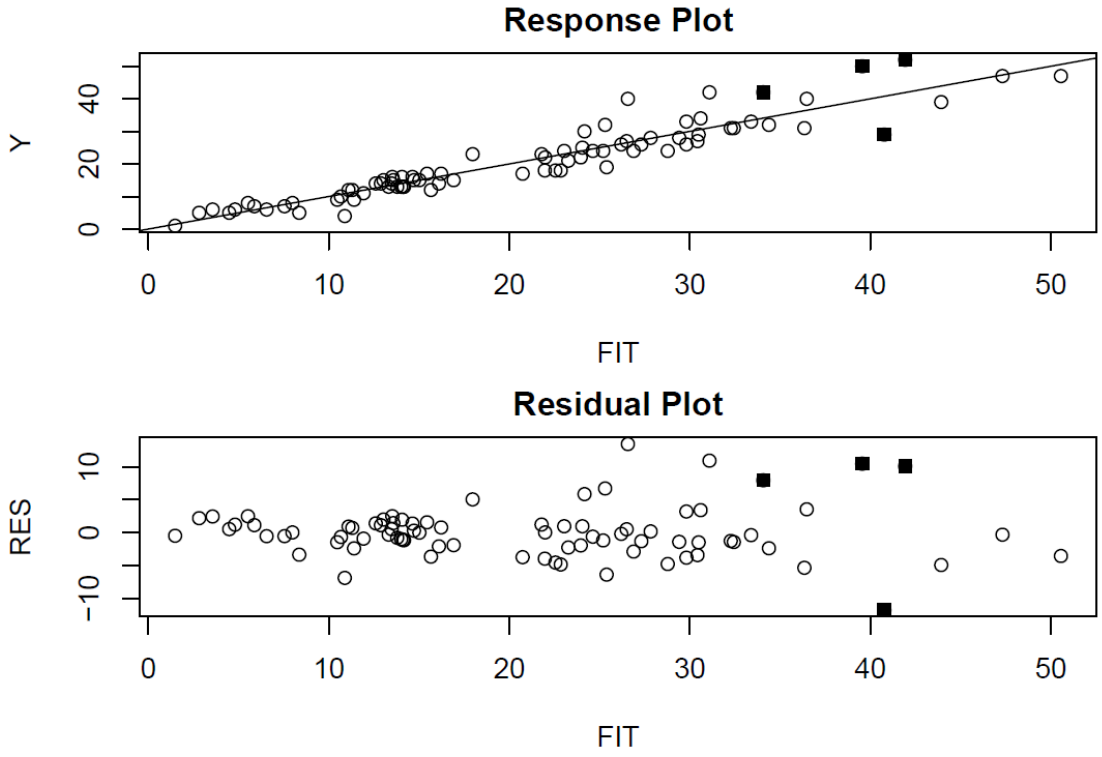

Figure 6.6. MLR plot for mussels' muscles data

In Figure 6.6, we applied some transformations to the predictors to enhance the fit shown in the MLR plot in Figure 6.8. Figure 6.7 presents 4 plots. In the response plot of Figure 6.7a, the lowess curve appears jagged to differentiate it from the estimated mean function. The horizontal line corresponds to the sample mean $\overline{Y}$. The OD plot in Figure 6.7b indicates that there is a little presence of overdispersion, as the vertical scale is less than ten times that of the horizontal scale, and few points are notably large and positioned above the slope 4 line. We observe that the mean

can be approximately estimated by identifying the value of the exponential curve in Figure 6.7a. For instance, at when ESP = 4, $E(Y) \approx 35$.

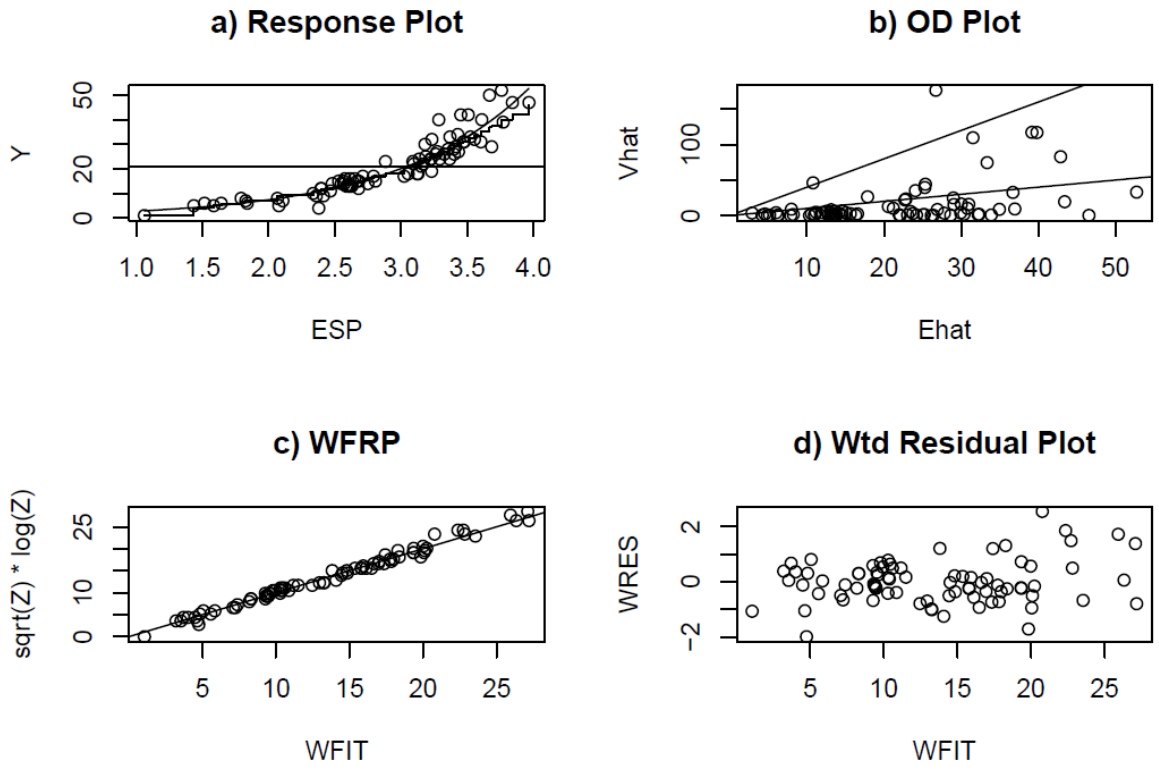

Figure 6.7. Plot for the transformed mussels' muscles data

We also examined the MLR response plot, which appears approximately linear and therefore, transforming the predictors is necessary to improve the fit. Figure 6.8 displays the MLR plot, which demonstrates a more clearly linear relationship.

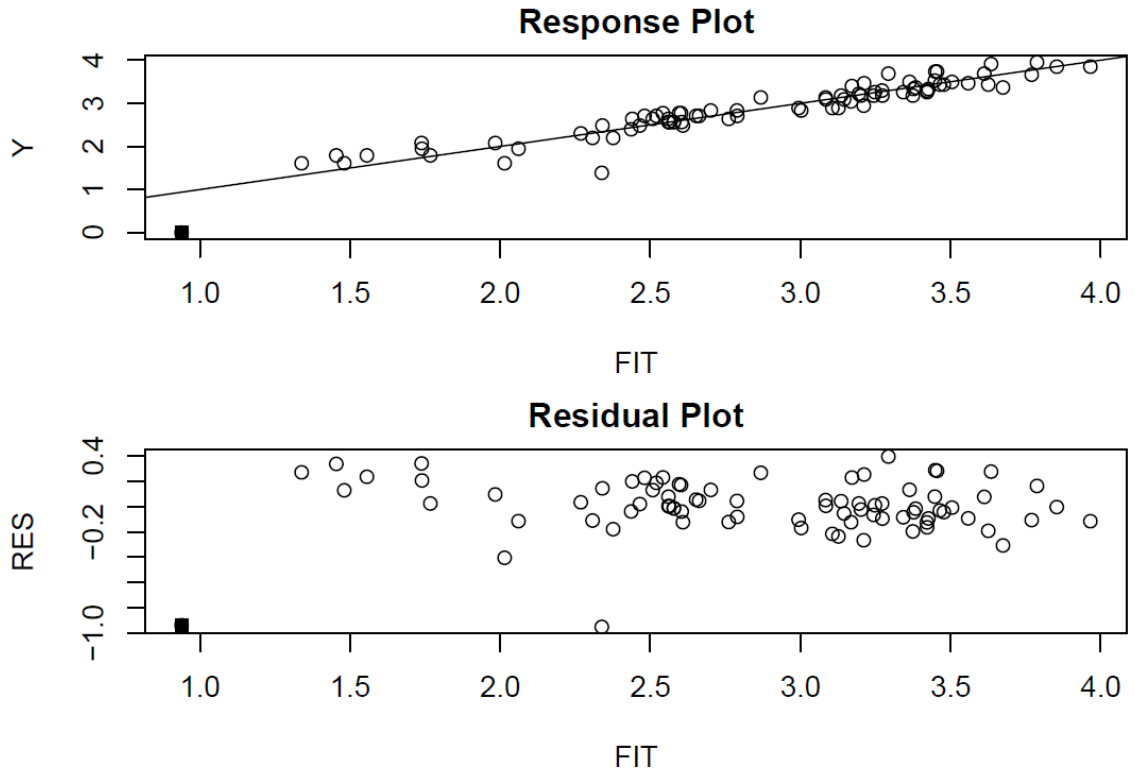

Figure 6.8. MLR plot for the transformed mussels' muscles data

An additional example is drawn from the Ceriodaphnia data presented by Myers, Montgomery, and Vining (2002, pp. 136-139). The response variable, Y, represents the number of Ceriodaphnia organisms counted in a container. The sample size $n = 70$ observations, with 7 different concentrations of jet fuel x_1 and an indicator variable for two strains of the organism x_2 serving as predictors. Since jet fuel was suspected to hinder reproduction, higher concentrations were expected to result in lower organism counts. The data was sourced from $source(\text{"}http://parker.ad.siu.edu/Olive/sldata.txt\text{"})$.

Figure 6.9 presents 4 plots. In the response plot of Figure 6.9a, the lowess curve appears jagged to differentiate it from the estimated mean function. The horizontal line corresponds to the

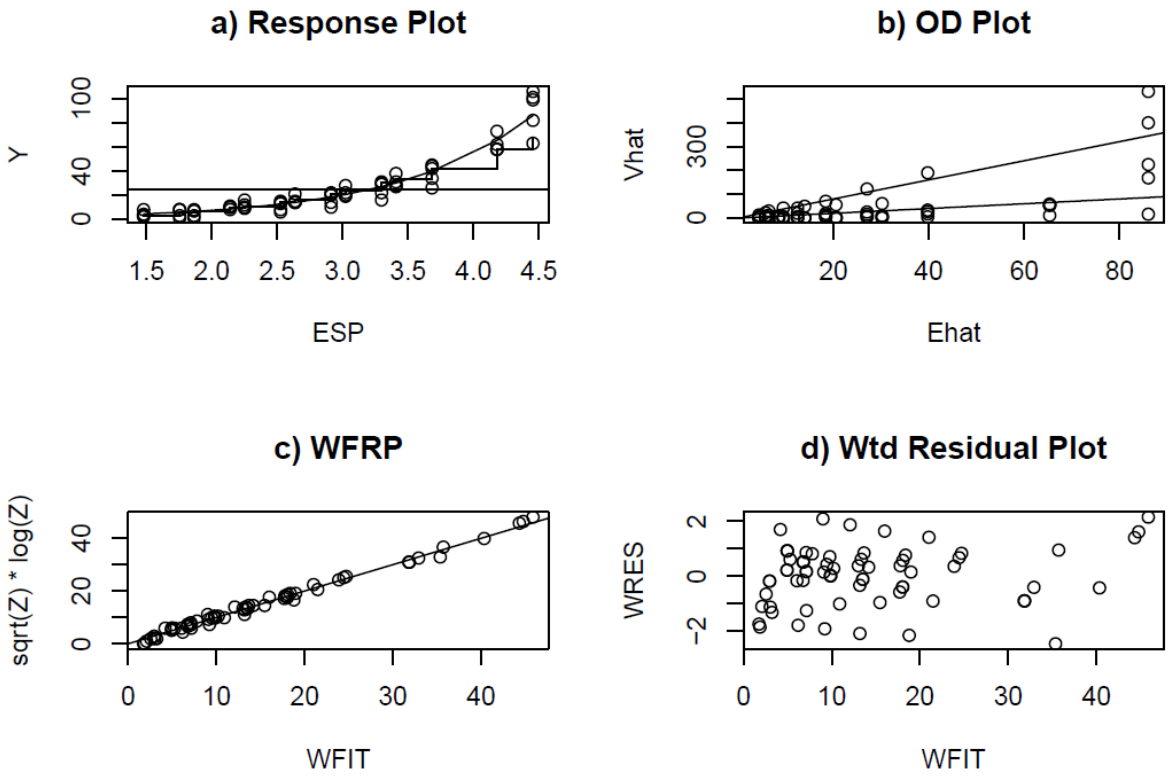

Figure 6.9. Plot for Ceriodaphnia data

sample mean $\overline{Y}$. The OD plot in Figure 6.9b indicates that there is a little presence of overdispersion, as the vertical scale is less than ten times that of the horizontal scale, and all but one of the plotted points are close to the wedge formed by the horizontal axis and slope 4 line. We observe that the Poisson mean can be approximately estimated by identifying the value of the exponential curve in Figure 6.9a.The plotted points scatter about the identity line in Figure 6.9c and there are no unusual points in Figure 6.9d. For instance, at when ESP = 4, $Y \approx$ Poisson(40). We also evaluated the MLR response plot, which looks roughly linear but is not ideal, similar to what occurs during the transformation process. Figure 6.10 displays the MLR plot, which demonstrates roughly linear relationship.

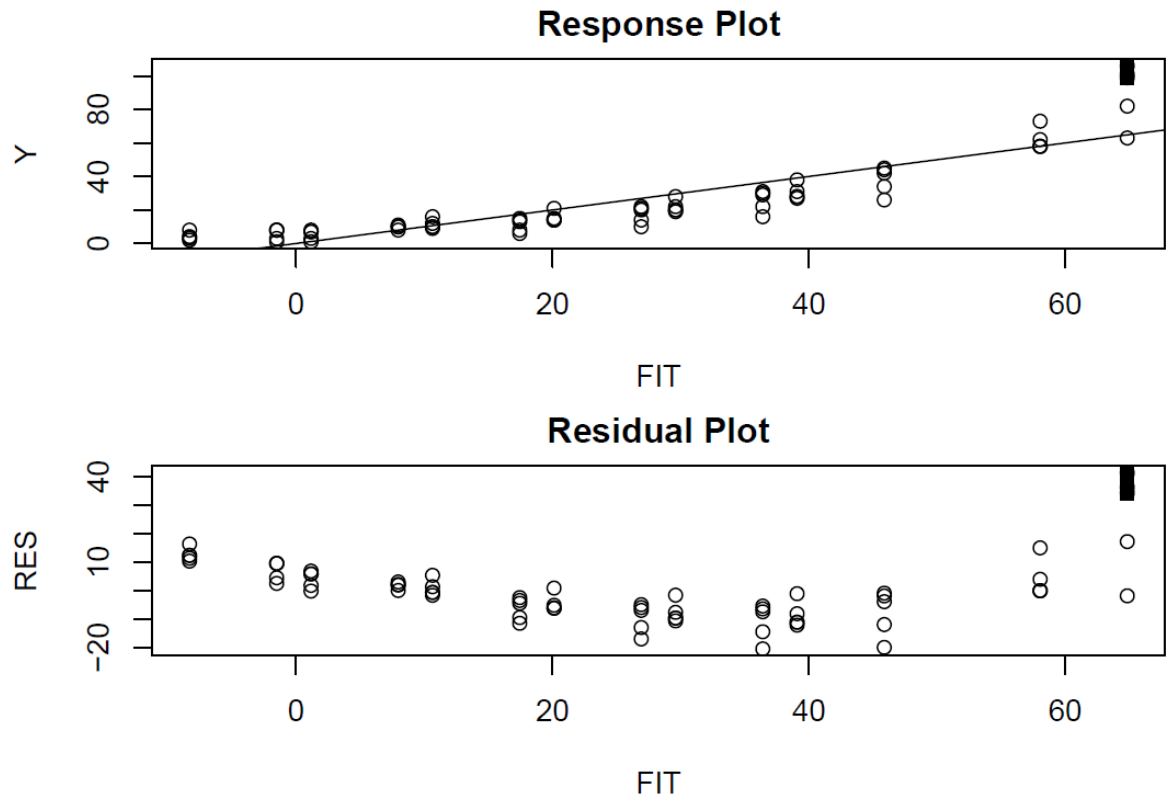

Figure 6.10. MLR plot for Ceriodaphnia data

Finally, we consider the Crab example from Agresti (2002, pp. 126-131), which applies Poisson regression to model the response variable Y , representing the number of satellites (male crabs) surrounding a female crab. The sample size is $n = 173$, and the predictor variables include

color (with categories: 2 for light medium, 3 for medium, 4 for dark medium, and 5 for dark), spine condition (1 for both spines intact, 2 for one worn or broken, and 3 for both worn or broken), carapace width measured in centimeters, and the female crab's weight in grams. The data was sourced from *source*("*http* : //*parker.ad.siu.edu*/*Olive*/*sldata.txt*").

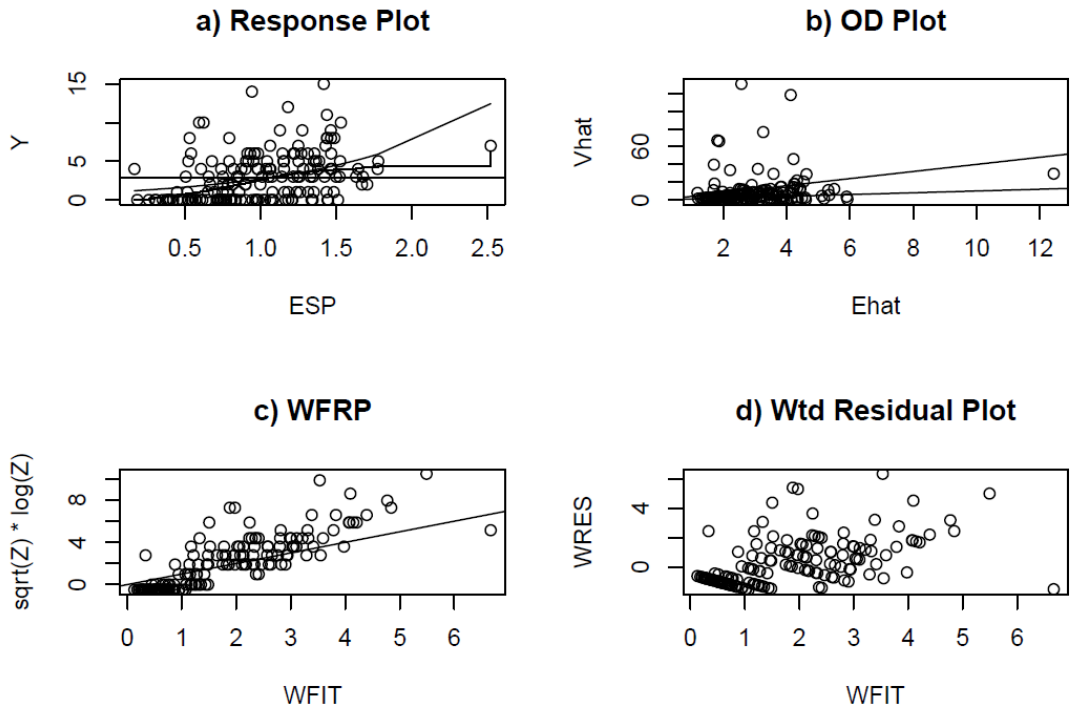

Figure 6.11. Plot for Crab data

The model that generated Figure 6.11 treated the ordinal variables color and spine condition using their original coded values. An alternative approach would be to treat spine condition as a categorical factor. Figure 6.11a indicates the presence of a single case with an unusually high ESP value. Additionally, the lowess curve does not closely follow the exponential curve in this figure. Figure 6.11b indicates the presence of overdispersion, as the vertical scale is approximately ten times larger than the horizontal scale, and numerous plotted points are both large and positioned above the slope 4 line. The lack of fit is more evident in Figure 6.11c, where the plotted points do not align with the identity line. However, the exponential mean function provides a better fit to the lowess curve than the straight line $Y = \overline{Y}$ does. Alternative models proposed by Agresti (2002)

might provide a better fit for the data. We also assessed the MLR response plot, which does not appear linear or ideal, reflecting the behavior seen during the transformation process. Figure 6.12 presents the MLR plot, illustrating a nonlinear relationship.

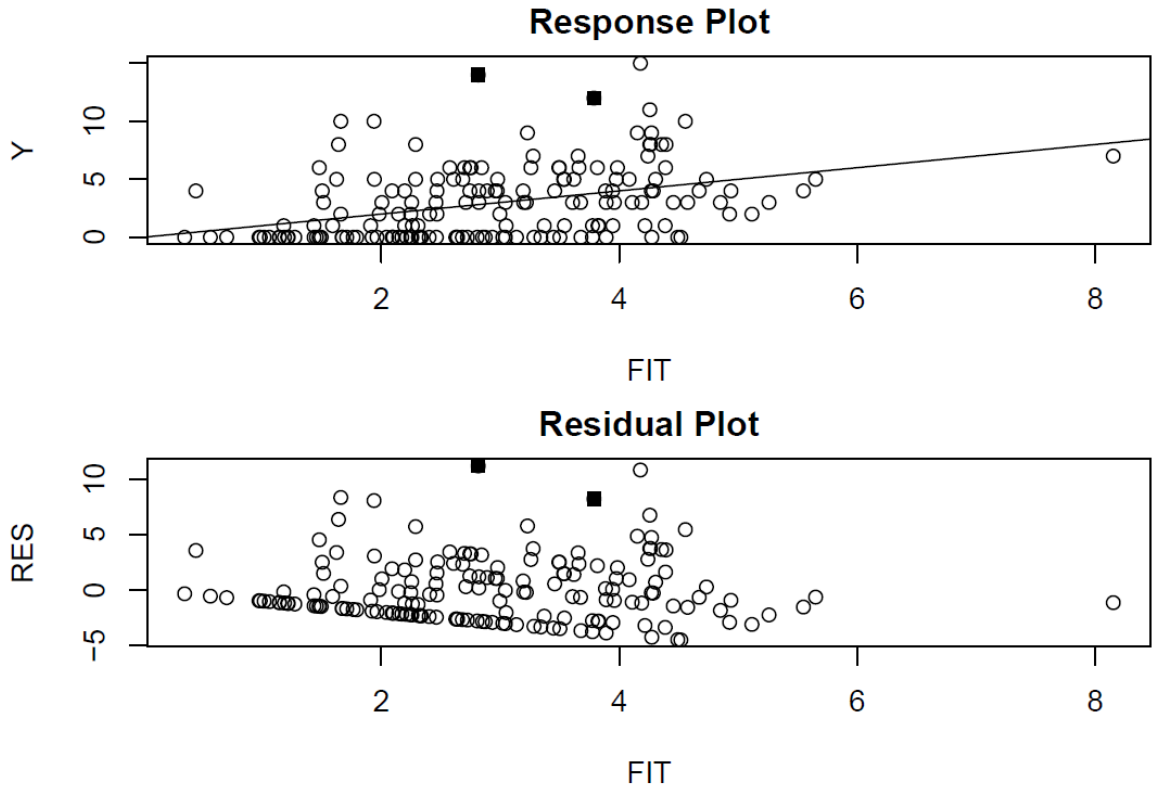

Figure 6.12. MLR plot for Crab data

CHAPTER 7

CONCLUSIONS

The response plot of the estimated sufficient predictor $\hat{\alpha} + \mathbf{x}^T\hat{\boldsymbol{\beta}}$ versus Y is useful for checking many regression models. See Olive (2013) for more on plots for such models, including a plot to detect overdispersion.

`software`

Plots and simulations were done in *R*. See R Core Team (2025). Programs are from the Olive (2025) collections of *R* functions *slpack.txt*, available from *source*("*http* : *//parker.ad.siu.edu/Olive/slpack.txt*"). The function `MLRplot` generates response and residual plots for multiple linear regression using one component partial least squares. Similarly, `prplot` makes the plots for Poisson regression based on one component partial least squares. Let $up = 1 - \delta/2$ be the correction factor used for the confidence intervals. The function `covxycis` obtains large sample $100(1-\delta)$ CI is $\hat{\eta}_i \pm t_{n-1,\,up}\, SE(\hat{\eta}_i)$ for both $\eta_i = Cov(x_i, Y)$ and $\eta_i = Cov(x_i, Z)$ where $i = 1, \ldots, p$.

For Table 5.1 and 5.2, the function `nbinroplssimzz` was used to create negative binomial regression data sets for finite κ, while the function `proplssimzz` was used to create the Poisson regression data sets corresponding to $\kappa = \infty$.

REFERENCES

[1] Abid, A.M., and Olive, D.J. (2025), "Some Simple High Dimensional One and Two Sample Tests," unpublished manuscript at (http://parker.ad.siu.edu/Olive/pphd1samp.pdf).

[2] Agresti, A. (2002), *Categorical Data Analysis*, 2nd ed., Wiley, Hoboken, NJ.

[3] 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.

[4] Brown, P.J. (1993), *Measurement, Regression, and Calibration*, Oxford University Press, New York, NY.

[5] Cameron, A.C., and Trivedi, P.K. (1998), *Regression Analysis of Count Data*, 1st and Cambridge University Press, Cambridge, UK.

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

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

[8] Chen, J., and Chen, Z. (2008), "Extended Bayesian Information Criterion for Model Selection with Large Model Spaces," *Biometrika*, 95, 759-771.

[9] Cook, R.D., and Forzani, L. (2024), *Partial Least Squares Regression*, Chapman and Hall/CRC, Boca Raton, FL.

[10] 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.

[11] Cook, R.D., and Weisberg, S. (1999), *Applied Regression Including Computing and Graphics,* Wiley, New York, NY.

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

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

[14] Friedman, J., Hastie, T., Hoefling, H., and Tibshirani, R. (2007), "Pathwise Coordinate Optimization," *Annals of Applied Statistics*, 1, 302-332.

[15] Friedman, J., Hastie, T., and Tibshirani, R. (2010), "Regularization Paths for Generalized Linear Models via Coordinate Descent," *Journal of Statistical Software*, 33, 1-22.

[16] Helland, I.S. (1990), "Partial Least Squares Regression and Statistical Models," *Scandanavian Journal of Statistics*, 17, 97-114.

[17] Helland, I.S. and Almøy, T. (1994), "Comparison of Prediction Methods When Only a Few Components Are Relevant," *Journal of the American Statistical Association,* 89, 583-591.

[18] Hilbe, J.M. (2011), *Negative Binomial Regression*, Cambridge University Press, 2nd ed., Cambridge, UK.

[19] Li, K.C., and Duan, N. (1989), "Regression Analysis Under Link Violation," *The Annals of Statistics,* 17, 1009-1052.

[20] Mevik, B.–H., Wehrens, R., and Liland, K.H. (2015), *pls: Partial Least Squares and Principal Component Regression*, *R* package version 2.5-0, (https://CRAN.R-project.org/package=pls).

[21] Myers, R.H., Montgomery, D.C., and Vining, G.G. (2002), *Generalized Linear Models with Applications in Engineering and the Sciences*, Wiley, New York, NY.

[22] Naik, P. and Tsai, C.L. (2000), "Partial Least Squares Estimator for Single Index Models," *Journal of the Royal Statistical Society, B,* 62, 763-771.

[23] Nelder, J.A., and Wedderburn, R.W.M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society, A*, 135, 370-384.

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

[25] Olive, D.J. (2017), *Linear Regression*, Springer, New York, NY.

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

[27] Olive, D.J., Alshammari, A.A., Pathiranage, K.G., and Hettige, L.A.W. (2024), "Testing with

the One Component Partial Least Squares and the Marginal Maximum Likelihood Estimators," is at (http://parker.ad.siu.edu/Olive/pphdwls.pdf).

[28] Olive, D.J., and Hawkins, D.M. (2005), "Variable Selection for 1D Regression Models," *Technometrics*, 47, 43-50.

[29] 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: Theory and Methods*, 54, 130-145.

[30] Pelawa Watagoda, L.C.R., and Olive, D.J. (2021), "Comparing Six Shrinkage Estimators with Large Sample Theory and Asymptotically Optimal Prediction Intervals," *Statistical Papers*, 62, 2407-2431.

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

[32] R Core Team (2025), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

[33] 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.

[34] 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.

[35] Simonoff, J.S. (2003), *Analyzing Categorical Data*, Springer, New York, NY.

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

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

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

[39] Zhao, A., Li, C., Li, R., and Zhang, Z. (2024), "Testing High-Dimensional Regression Coef-

ficients in Linear Models," *The Annals of Statistics*, 52, 2034-2058.

[40] 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.

[41] Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M. (2009), *Mixed Effects Models and Extensions in Ecology with R,* Springer, New York, NY.

VITA

Graduate School
Southern Illinois University Carbondale

Paul Amarh Quaye
paul.quaye@siu.edu / latticequaye@gmail.com

Kwame Nkrumah University of Science and Technology, Kumasi- Ghana
Bachelor of Science, Mathematics, June 2015

Southern Illinois University Edwardsville
Master of Science, Mathematics, May 2021

Dissertation Paper Title:

Testing Poisson Regression and Related Models with the One Component Partial Least Squares Estimator.

Major Professor: Prof. David J. Olive

Publications:

Malinzi, J., & Quaye, P. A. (2018). Exact solutions of non-linear evolution models in physics and biosciences using the hyperbolic tangent method. Mathematical and Computational Applications, 23(3), 35.

Quaye, P. (2021). Quantifying Statistical Evidence: P-Values, the Likelihood Principle, and Bayes Factors. Southern Illinois University at Edwardsville.