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

David J. Olive and Paul A. Quaye \* Southern Illinois University

July 4, 2025

#### Abstract

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.

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

### **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  $Y \perp \boldsymbol{x} \mid \boldsymbol{x}^T \boldsymbol{\beta}$ . 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  $Y \sim \text{Poisson}(e^{\text{SP}})$ . 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  $V(Y) = \mu + \mu^2 / \kappa$ .

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

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

<sup>\*</sup>David J. Olive is Professor, School of Mathematical & Statistical Sciences, Southern Illinois University, Carbondale, IL 62901, USA.

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  $\mu(SP) = \exp(SP)$ ,

$$E(Y|SP) = \frac{\exp(SP)}{1 - \exp(-\exp(SP))}$$
, 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  $n \ge 10p$ . 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 rgression, 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

$$\boldsymbol{x}^{T}\boldsymbol{\beta} = \boldsymbol{x}_{S}^{T}\boldsymbol{\beta}_{S} + \boldsymbol{x}_{E}^{T}\boldsymbol{\beta}_{E} = \boldsymbol{x}_{S}^{T}\boldsymbol{\beta}_{S}$$
(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) 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  $\boldsymbol{\beta}_{O} = \mathbf{0}$  if  $S \subseteq I$ . The model using  $\boldsymbol{x}^{T}\boldsymbol{\beta}$  is the full model.

To clarify notation, suppose p = 3, a constant  $\alpha$  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  $\hat{\beta}_{VS}$  is complicated. See Pelawa Watagoda and Olive (2021) for multiple linear regression, and Rathnayake and Olive (2023) for models such as generalized linear models. For fixed p, these two papers showed that  $\hat{\beta}_{VS}$ 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:

$$Z_i = \alpha_Z + \boldsymbol{x}_i^T \boldsymbol{\beta}_Z + e_i \tag{2}$$

where the  $e_i$  are independent with expected value  $E(Z_i) = 0$  and variance  $V(Z_i) = \sigma_i^2$ . 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  $Y|\mathbf{x}^T\boldsymbol{\beta} = \alpha + \mathbf{x}^T\boldsymbol{\beta} + e$  or  $Y_i = \alpha + \mathbf{x}_i^T\boldsymbol{\beta} + e_i$  or

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

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\phi} + \boldsymbol{e},\tag{4}$$

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

For a multiple linear regression model with heterogeneity, assume model (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  $\sigma_i^2$  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  $\text{Cov}(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}} = E[(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{x} - E(\boldsymbol{x}))^T] = E(\boldsymbol{x}\boldsymbol{x}^T) - E(\boldsymbol{x})E(\boldsymbol{x}^T) \text{ and } \boldsymbol{\eta} = \text{Cov}(\boldsymbol{x}, Y) = \boldsymbol{\Sigma}_{\boldsymbol{x}Y} = E[(\boldsymbol{x} - E(\boldsymbol{x})(Y - E(Y))] = E(\boldsymbol{x}Y) - E(\boldsymbol{x})E(Y) = E[(\boldsymbol{x} - E(\boldsymbol{x}))Y] = E[\boldsymbol{x}(Y - E(Y))].$ Let

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

and

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

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

$$\hat{oldsymbol{eta}}_{OLS} = ilde{\Sigma}_{oldsymbol{x}}^{-1} ilde{\Sigma}_{oldsymbol{x}Y} = \hat{\Sigma}_{oldsymbol{x}}^{-1} \hat{\Sigma}_{oldsymbol{x}Y} = \hat{\Sigma}_{oldsymbol{x}}^{-1} \hat{oldsymbol{\eta}}.$$

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

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

$$\lambda = \frac{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^{T} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}}{\boldsymbol{\Sigma}_{\boldsymbol{x}Y}^{T} \boldsymbol{\Sigma}_{\boldsymbol{x}} \boldsymbol{\Sigma}_{\boldsymbol{x}Y}} \quad \text{and} \quad \hat{\lambda} = \frac{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^{T} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}}{\hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}^{T} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{x}Y}}$$
(5)

for  $\Sigma_{\boldsymbol{x}Y} \neq \mathbf{0}$ . If  $\Sigma_{\boldsymbol{x}Y} = \mathbf{0}$ , then  $\beta_{OPLS} = \mathbf{0}$ . 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  $\hat{\eta}_{OPLS} = \hat{\Sigma}_{\boldsymbol{x}Y}$  and OPLS under milder regularity conditions than those in the previous literature, where  $\eta_{OPLS} = \Sigma_{\boldsymbol{x}Y}$ . 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 (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  $\hat{\beta}_{MMLE} = (\hat{\beta}_{1,M}, ..., \hat{\beta}_{p,M})^T$ .

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

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

If the  $t_i$  are the predictors that are scaled or standardized to have unit sample variances, then

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

where (t, Y) denotes that Y was regressed on t, and 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  $x_i$  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,  $\beta_{OLS} = c\beta$ . Hence  $\Sigma_{xY} = c\Sigma_x\beta$ . Thus  $\Sigma_{xY} = d\beta$  if  $\Sigma_x = \tau^2 I_p$  for some constant  $\tau^2 > 0$ . If  $\beta = \beta_{OLS}$  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  $\alpha + \beta^T x$ . Chang and Olive (2010) considered OLS tests for these models. Simulation with  $\hat{\Sigma}_{\boldsymbol{x}Y}$  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 (2024) 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  $\beta_i$  or  $\beta_i$ ) where  $n \ge Ja$  with  $J \ge 10$ . 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.

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

#### Large Sample Theory 2

This section reviews the Olive and Zhang (2025) large sample theory for  $\hat{\eta}_{OPLS} = \hat{\Sigma}_{xY}$ 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  $H_0$ :  $\beta_i - \beta_i = 0$ . 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  $\Sigma^*$  be  $\hat{\Sigma}$  or  $\tilde{\Sigma}$ . Then  $\Sigma^*_{\boldsymbol{w}} = \boldsymbol{A}_n \Sigma^*_{\boldsymbol{x}} \boldsymbol{A}_n^T$  and  $\Sigma^*_{\boldsymbol{w}Y} = \boldsymbol{A}_n \Sigma^*_{\boldsymbol{x}Y}$ . b) If  $\boldsymbol{A}_n$  is a constant matrix, then  $\Sigma_{\boldsymbol{w}} = \boldsymbol{A}_n \Sigma_{\boldsymbol{x}} \boldsymbol{A}_n^T$  and  $\Sigma_{\boldsymbol{w}Y} = \boldsymbol{A}_n \Sigma_{\boldsymbol{x}Y}$ .

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

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

Then the low order moments are needed for  $\Sigma_z$  to be a consistent estimator of  $\Sigma_w$ .

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

$$\sqrt{n}(\overline{\boldsymbol{w}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}), \ \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}), \tag{8}$$

and  $\sqrt{n}(\tilde{\boldsymbol{\eta}}_n - \boldsymbol{\eta}) \stackrel{D}{\rightarrow} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\boldsymbol{w}}).$ 

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

$$\sqrt{n}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_{OPLS} - \boldsymbol{\beta}_{OPLS}) \xrightarrow{D} N_k(\boldsymbol{0}, \lambda^2 \boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{w}} \boldsymbol{A}^T).$$
(9)

### 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  $H_0: A\beta_{OPLS} = \mathbf{0}$  versus  $H_1: A\beta_{OPLS} \neq \mathbf{0}$  is equivalent to testing  $H_0: A\eta = \mathbf{0}$  versus  $H_1: A\eta \neq \mathbf{0}$  where A is a  $k \times p$  constant matrix. Let  $\operatorname{Cov}(\hat{\Sigma}_{\boldsymbol{x}Y}) = \operatorname{Cov}(\hat{\eta}) = \boldsymbol{\Sigma}_{\boldsymbol{w}}$  be the asymptotic covariance matrix of  $\hat{\eta} = \hat{\Sigma}_{\boldsymbol{x}Y}$ . In high dimensions where n < 5p, we can't get a good nonsingular estimator of  $\operatorname{Cov}(\hat{\Sigma}_{\boldsymbol{x}Y})$ , but we can get good nonsingular estimators of  $\operatorname{Cov}(\hat{\Sigma}_{\boldsymbol{u}Y}) = \operatorname{Cov}((\hat{\eta}_{i1}, ..., \hat{\eta}_{ik})^T)$  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  $\sigma_{x_i,Y} = \operatorname{Cov}(x_i, Y)$ .

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  $\operatorname{Cov}(\boldsymbol{x}, Y)$  if  $\operatorname{Cov}(\boldsymbol{x}, v) + \operatorname{Cov}(\boldsymbol{u}, Y) + \operatorname{Cov}(\boldsymbol{u}, v) = \mathbf{0}$ , which occurs, for example, if  $\boldsymbol{x} \perp v$ ,  $\boldsymbol{u} \perp Y$ , and  $\boldsymbol{u} \perp v$ . 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  $\lambda$ . 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 (3). Assume that the cases  $(\boldsymbol{x}_i^T, Y_i)^T$  are iid with  $E(Y) = \mu_Y$ ,  $E(\boldsymbol{x}) = \boldsymbol{\mu}_{\boldsymbol{x}}$  and nonsingular  $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma}_{\boldsymbol{x}}$ . Let  $\boldsymbol{\beta} = \boldsymbol{\beta}_{OLS}$ . Then testing  $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$  versus  $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0$  is equivalent to testing  $H_0 : \boldsymbol{\mu} = \boldsymbol{0}$  versus  $H_1 : \boldsymbol{\mu} \neq \boldsymbol{0}$  with  $\boldsymbol{\mu} = E(\boldsymbol{w}_i) = \boldsymbol{\Sigma}_{\boldsymbol{x}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)$ where  $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \boldsymbol{\mu}_Y - (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})^T \boldsymbol{\beta}_0)$ , and a one sample test can be applied to  $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y} - (\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \boldsymbol{\beta}_0)$ .

This paper modifies the above test for  $\beta_0 = 0$ . The resulting test can be used for many regression models, not just multiple linear regression. Suppose  $\beta_D = D^{-1}\Sigma_{xY}$ where D is a  $p \times p$  positive definite matrix. Then  $\beta_D = 0$  if and only if  $\Sigma_{xY} = 0$ . Then  $D^{-1} = \lambda I$  for OPLS,  $D^{-1} = \Sigma_x^{-1}$  for OLS, and  $D^{-1} = [diag(\Sigma_x)]^{-1}$  for the MMLE. The k-component partial least squares estimator can be found by regressing Yon a constant and on  $W_i = \hat{\eta}_i^T x$  for i = 1, ..., k where  $\hat{\eta}_i = \hat{\Sigma}_x^{i-1} \hat{\Sigma}_{xY}$  for i = 1, ..., k. See Helland (1990). Hence  $\beta_{kPLS} = 0$  if  $\Sigma_{xY} = 0$ . Thus if the cases  $(x_i^T, Y_i)^T$  are iid, then using  $\beta_0 = 0$  gives tests for  $H_0 : \beta = 0, H_0 : \beta_{MMLE} = 0, H_0 : \Sigma_{xY} = 0$ ,  $H_0 : \beta_{OPLS} = 0$ , and  $H_0 : \beta_{kPLS} = 0$ . For multiple linear regression with heterogeneity, model (3) holds with  $E(e_i) = 0$  and  $V(e_i) = \sigma_i^2$ . Under mild conditions,  $\hat{\beta}_{OLS}$  is still a consistent estimator of  $\beta = \beta_{OLS} = \Sigma_x^{-1} \Sigma_{xY}$ . Hence the test can be used when the constant variance assumption is violated.

For a generalized linear model and several other regression models that depend on the predictors  $\boldsymbol{x}$  only through  $SP = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}$ , if  $\boldsymbol{\beta} = \boldsymbol{0}$ , then the  $Y_i$  are iid and do not depend on  $\boldsymbol{x}$ , and thus satisfy a multiple linear regression model with  $\boldsymbol{\beta}_{OLS} = \boldsymbol{0}$ . Typically, if  $\boldsymbol{\beta} \neq \boldsymbol{0}$ , then  $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} \neq 0$ . An exception is when there is a lot of symmetry which rarely occurs with real data. For example, suppose Y = m(SP) + e where the iid errors  $e_i \sim N(0, \sigma_1^2)$  are independent of the predictors,  $SP \sim N(0, \sigma_2^2)$ , and the function m is symmetric about 0, e.g.  $m(SP) = (SP)^2$ . Then  $\boldsymbol{\beta}_{OLS} = 0$  and  $\boldsymbol{\Sigma}_{\boldsymbol{x}Y} = 0$  even if  $\boldsymbol{\beta} \neq \boldsymbol{0}$ .

If  $\boldsymbol{\beta}_0 = \mathbf{0}$ , then  $\boldsymbol{w}_i = (\boldsymbol{x}_i - \boldsymbol{\mu}_{\boldsymbol{x}})(Y_i - \boldsymbol{\mu}_Y)$ , and  $E(\boldsymbol{w}_i) = E(\boldsymbol{u}_i) = E[\boldsymbol{x}_i(Y_i - \boldsymbol{\mu}_Y)] = \boldsymbol{\Sigma}_{\boldsymbol{x}Y}$ . Hence we replace  $\boldsymbol{v}_i = (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(Y_i - \overline{Y})$  by  $\boldsymbol{z}_i = \boldsymbol{x}_i(Y_i - \overline{Y})$  and apply a high dimensional one sample test on the  $\boldsymbol{z}_i$ . Then  $\boldsymbol{\mu}_{\boldsymbol{x}}$  does not need to be estimated by  $\overline{\boldsymbol{x}}$ .

Next, we review some results for the one sample test that will be used. Suppose  $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$  are iid random vectors with  $E(\boldsymbol{x}) = \boldsymbol{\mu}$  and covariance matrix  $Cov(\boldsymbol{x}) = \boldsymbol{\Sigma}$ . Then the test  $H_0: \boldsymbol{\mu} = \boldsymbol{0}$  versus  $H_1: \boldsymbol{\mu} \neq \boldsymbol{0}$  is equivalent to the test  $H_0: \boldsymbol{\mu}^T \boldsymbol{\mu} = 0$  versus  $H_1: \boldsymbol{\mu}^T \boldsymbol{\mu} \neq 0$ . A U-statistic for estimating  $\boldsymbol{\mu}^T \boldsymbol{\mu}$  is

$$T_n = T_n(\boldsymbol{x}) = \frac{1}{n(n-1)} \sum_{i \neq j} \boldsymbol{x}_i^T \boldsymbol{x}_j.$$
 (10)

Let  $V_0(T_n)$  be the variance of  $T_n$  when  $H_0: \boldsymbol{\mu} = \boldsymbol{0}$  is true. Let the variance  $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2$  for  $i \neq j$ . Abid and Olive (2025) give a straight forward proof that

$$V_0(T_n) = \frac{2\sigma_W^2}{n(n-1)}.$$

Chen and Qin (2010) proved that

$$V_0(T_n) = \frac{2}{n(n-1)} tr(\Sigma^2)$$

where tr() is the trace function. Thus  $V(\boldsymbol{x}_i^T \boldsymbol{x}_j) = \sigma_W^2 = tr(\boldsymbol{\Sigma}^2)$ . Srivastava and Du (2008), Bai and Saranadasa (1996), Chen and Qin (2010), Li (2023) and others proved that under mild regularity conditions when  $H_0$  is true,  $T_n/\sqrt{\hat{V}_0(T_n)} \xrightarrow{D} N(0, 1)$ . Under regularity conditions when  $H_0$  is true, Li (2023) proved that  $T_n/\sqrt{\hat{V}_0(T_n)} \xrightarrow{D} t_k$  as  $p \to \infty$  for fixed  $n \geq 3$  where k = 0.5n(n-1) - 1.

Two estimators of  $\sigma_W^2$  are simple to compute. Let  $W_{ij} = \boldsymbol{x}_i^T \boldsymbol{x}_j$  for  $i \neq j$ . Let  $s_n^2 = \hat{V}_0(T_n)$ . An estimator nearly the same as the one used by Li (2023) is

$$n(n-1)s_n^2 = \hat{\sigma}_W^2 = \frac{1}{n(n-1)} \sum_{i \neq j} (\boldsymbol{x}_i^T \boldsymbol{x}_j - T_n)^2 = \frac{1}{n(n-1)} \sum_{i \neq j} (W_{ij} - T_n)^2.$$

Ahlam and Olive (2025) proposed the following estimator. Let  $m = \text{floor}(n/2) = \lfloor n/2 \rfloor$  be the integer part of n/2. So floor(100/2) = floor(101/2) = 50. Let the iid random variables  $W_i = \boldsymbol{x}_{2i-1}^T \boldsymbol{x}_{2i}$  for i = 1, ..., m. Hence  $W_1, W_2, ..., W_m = \boldsymbol{x}_1^T \boldsymbol{x}_2, \boldsymbol{x}_3^T \boldsymbol{x}_4, ..., \boldsymbol{x}_{2m-1}^T \boldsymbol{x}_{2m}$ . Note that  $E(W_i) = \boldsymbol{\mu}^T \boldsymbol{\mu}$  and  $V(W_i) = \sigma_W^2$ . Let  $n(n-1)s_n^2 = S_W^2$  be the sample variance of the  $W_i$ .

## 3 Incorporating Information from Several Regression Estimators

The theory and tests from the previous section can be applied to model (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 < n-1. 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

This result suggests computing  $W_i = \hat{\eta}_i^T x$  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  $\hat{\eta}_1 = \hat{\Sigma}_{\boldsymbol{x}Z}$ ,  $\hat{\eta}_2 = \hat{\Sigma}_{\boldsymbol{x}} \hat{\Sigma}_{\boldsymbol{x}Z}$ ,  $\hat{\eta}_3 = \hat{\Sigma}_{\boldsymbol{x}}^2 \hat{\Sigma}_{\boldsymbol{x}Z}$ ,  $\hat{\eta}_4 = \hat{\beta}_L(\boldsymbol{x}, Z)$ = the lasso estimator from regressing Z on  $\boldsymbol{x}$ ,  $\hat{\eta}_5 = \hat{\beta}_{RR}(\boldsymbol{x}, Z)$ = the ridge regression estimator from regressing Z on  $\boldsymbol{x}$ ,  $\hat{\eta}_6 = \hat{\beta}_{LPR}(\boldsymbol{x}, Y)$ = the lasso Poisson regression estimator from regressing Y on  $\boldsymbol{x}$ . Let  $\boldsymbol{x}_I$  denote the set of variables selected using  $\hat{\eta}_4$ . Then  $\hat{\eta}_7 = \hat{\Sigma}_{\boldsymbol{x}_IZ}$ ,  $\hat{\eta}_8 = \hat{\Sigma}_{\boldsymbol{x}_I}\hat{\Sigma}_{\boldsymbol{x}_IZ}$ ,  $\hat{\eta}_9 = \hat{\Sigma}_{\boldsymbol{x}_I}^2\hat{\Sigma}_{\boldsymbol{x}_IZ}$ ,  $\hat{\eta}_{10} = \hat{\beta}_{RR}(\boldsymbol{x}_I, Z)$ = 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  $\hat{\eta}_6$ .

### 4 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  $ESP_i = \alpha + \boldsymbol{\beta}^T \boldsymbol{x}_i = 1 + 1x_{i,1} + \dots + 1x_{i,k}$  for  $i = 1, \dots, n$ . Hence  $\alpha = 1$  and  $\boldsymbol{\phi} = (\alpha, \boldsymbol{\beta}^T)^T = (1, \dots, 1, 0, \dots, 0)^T$  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  $\boldsymbol{\beta}_{PR}$  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  $\boldsymbol{\Sigma}_{\boldsymbol{x}Z} = (c_o, \dots, c_o, 0, \dots, 0)^T$  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  $\eta_i$ . 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-1confidence intervals. Then the minimum observed coverage was found. The percentage of the observed coverages that were  $\geq 0.9, 0.92, 0.93, 0.94$ , and 0.96 were also recorded. The test  $H_0 : (\eta_i, \eta_j)^T = (0, 0)^T$  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.

Tables 1 and illustrates Theorem 1a) where k = 1 and Table 1 replaces Y with Z. 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 1 correspond to Poisson regression. The confidence interval for  $\eta_1$  never contained 0, hence the minimum coverage was 0 with observed power = 1-0 = 1. The proportion of CIs that had coverage  $\geq 0.94$  was 0.9898 (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 0.8081 (80/99 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 0.9438. The minimum average CI length

$\kappa$	mincov	$\cos 90$	$\cos 92$	$\cos 93$	$\cos 94$	$\cos 96$	testcov
$\infty$	0.0000	0.9899	0.9899	0.9899	0.9899	0.8081	0.9438
len	0.4166	0.4187	0.4875				
0.5	0.0062	0.9899	0.9899	0.9899	0.9899	0.7576	0.9440
len	0.5050	0.5084	0.5686				
1	0.0000	0.9899	0.9899	0.9899	0.9899	0.7475	0.9410
len	0.4809	0.4834	0.5421				
10	0.0000	0.9899	0.9899	0.9899	0.9899	0.6970	0.9412
len	0.4258	0.4279	0.4929				
100	0.0000	0.9899	0.9899	0.9899	0.9899	0.6566	0.9464
len	0.4174	0.4195	0.4882				
1000	0.0000	0.9899	0.9899	0.9899	0.9899	0.7071	0.9430
len	0.4164	0.4181	0.4848				
10000	0.0000	0.9899	0.9899	0.9899	0.9899	0.9899	0.9450
len	0.4163	0.4190	0.4875				

Table 1: Cov(x,Z), n=100, p=100, k=1,  $\kappa$ =1,0.5,10,100,1000,10000

was 0.4166, the sample mean of the average CI lengths was 0.4187, and the maximum average length was 0.4875, corresponding to  $\eta_1$ . The second two lines and below for Table 1 were for the negative binomial regression with kappa =  $\kappa = 0.5, 1, 10, 100, 1000, 10000$ . The interpretation of Table 2 is similar, but Y is used instead of Z, resulting in longer lengths.

### 5 CONCLUSIONS

The response plot of the estimated sufficient predictor  $\hat{\alpha} + \boldsymbol{x}^T \hat{\boldsymbol{\beta}}$  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

The *R* software was used in the simulations. See R Core Team (2024). Programs are from the Olive (2025) collections of *R* functions *slpack.txt*, available from (http://parker.ad. siu.edu/Olive/slpack.txt). For Table 1, the function nbinroplssim was used to create negative binomial regression data sets for finite  $\kappa$ , while the function proplssim was used to create the Poisson regression data sets corresponding to  $\kappa = \infty$ .

#### References

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).

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

Bai, Z.D., and Saranadasa, H. (1996), "Effects of High Dimension: by an Example of a Two Sample Problem," *Statistica Sinica*, 6, 311-329.

Basa, J., Cook, R.D., Forzani, L., and Marcos, M. (2024), "Asymptotic Distribution

$\kappa$	mincov	$\cos 90$	$\cos 92$	$\cos 93$	$\cos 94$	$\cos 96$	testcov
$\infty$	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 2: Cov(x,Y), n=100, p=100, k=1,  $\kappa = 0.5, 1, 10, 100, 1000, 10000$ 

of One-Component Partial Least Squares Regression Estimators in High Dimensions," The Canadian Journal of Statistics, 52, 118-130.

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

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

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

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

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

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

Chen, S.X., and Qin, Y.L. (2010), "A Two Sample Test for High-dimensional Data with Applications to Gene-Set Testing," *The Annals of Statistics*, 38, 808-835.

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

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.

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

Fan, J., and Song, R. (2010), "Sure Independence Screening in Generalized Linear

Models with np-Dimensionality," The Annals of Statistics, 38, 3217-3841.

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

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

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

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

Li, J. (2023), "Finite Sample t-Tests for High-dimensional Means," Journal of Multivariate Analysis, 196, 105183.

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

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

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

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

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

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," *Communications in Statistics: Theory and Methods*, to appear.

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.

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.

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

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

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

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

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

Simonoff, J.S. (2003), Analyzing Categorical Data, Springer, New York, NY.

Srivastava, M.S., and Du, M. (2008), "A Test for the Mean Vector with Fewer Observations Than the Dimension," *Journal of Multivariate Analysis*, 99, 386-402. Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal* of the Royal Statistical Society, B, 58, 267-288.

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

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

Zhao, A., Li, C., Li, R., and Zhang, Z. (2024), "Testing High-Dimensional Regression Coefficients in Linear Models," *The Annals of Statistics*, 52, 2034-2058.

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.

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.