



Improved shrinkage estimators in zero-inflated negative binomial regression model

Zahra Zandi* , Hossein Bevrani , Reza Arabi Belaghi 

Department of Statistics, University of Tabriz, Tabriz, Iran

Abstract

Zero-inflated negative binomial model is an appropriate choice to model count response variables with excessive zeros and overdispersion simultaneously. This paper addressed parameter estimation in the zero-inflated negative binomial model when there are many predictors, so that some of them are inactive and have not influence on the response variable. We proposed parameter estimation based on the linear shrinkage, pretest, shrinkage pretest, Stein-type, and positive Stein-type estimators. We obtained the asymptotic distributional biases and risks of the suggested estimators theoretically. We also conducted a Monte Carlo simulation study to compare the performance of each estimator with the unrestricted estimator in terms of simulated relative efficiency. Based on the results, the performances of the proposed estimators were better than that of the unrestricted estimator. The suggested estimators were applied to the wildlife fish data to appraise their performance.

Mathematics Subject Classification (2020). 62J07, 62J12, 62F10

Keywords. Monte Carlo simulation, overdispersion, shrinkage estimators, zero-inflated negative binomial regression

1. Introduction

Many fields such as public health, insurance, and epidemiology may consist of count variables. In that case, Poisson and negative binomial models are two effective methods in order to explain the relationship between the response variable and a set of explanatory variables. Based on the Poisson regression model, the variance and mean of the response variable are equal. However, this restriction is violated in many applications because the data are often overdispersed, i.e. the variance of the response variable is significantly more than its mean. The negative binomial model is an appropriate choice for accounting overdispersion. However, when there are many zeros in the data, the zero-inflated negative binomial (ZINB) model (introduced by [11]) is useful for analyzing such data. The ZINB model is a mixture of two data generation groups. One is called "not at risk class" group that generates only zeros and the other group is called "at risk class" or "count component" that generates count data form an ordinary negative binomial distribution. The ZINB

*Corresponding Author.

Email addresses: zahrazandi163@yahoo.com (Z. Zandi), bevrani@gmail.com (H. Bevrani), rezaarabi11@gmail.com (R. Arabi Belaghi)

Received: 08.04.2021; Accepted: 11.09.2021

model has been studied by several authors. Interested readers are referred to [8, 23, 24, 27–29].

The main goal of this study is to estimate the regression parameters in the zero-inflated negative binomial model via the linear shrinkage, pretest, shrinkage pretest, Stein-type, and positive Stein-type under prior information that is incorporated in the model through a linear restriction on the parameters [2]. Based on the pretest procedure introduced by [9], any uncertain prior information can be validated through a pretest, and depending on such validation, the information may be incorporated into the model as a restriction sub-model. The linear shrinkage (LS) estimator proposed by [26] uses a linear function of the unrestricted and restricted estimators along with a shrinkage parameter. Ahmed [1] suggested a shrinkage pretest (SP) strategy that combines both the pretest and the linear shrinkage approaches. The shrinkage estimator or Stein-type estimator proposed by [25], takes an approach by shrinking the full model estimator to a sub-model estimator. For more information about these strategies see [2, 22]. The shrinkage estimators have been applied by several authors for improving and estimation of parameter in regression models. Some of them are [5–7, 15, 18–20].

In this study, we proposed the shrinkage strategies in the zero-inflated negative binomial model to improve estimation of parameters when the model is assumed to be sparse. We established the asymptotic distributional biases and risks of the estimators when the linear restriction may be wrong. The simulated relative efficiency of the proposed estimators was also obtained and compared with the unrestricted maximum likelihood estimator (MLE) through a Monte Carlo simulation and a real dataset.

The rest of this article is organized as follows. The zero-inflated negative binomial model, unrestricted, restricted, and shrinkage estimators are introduced in Section 2. The asymptotic distributional biases and risks of the proposed estimators are presented in Section 3. The details of Monte Carlo simulation study are described in Section 4. A real data example is given in Section 5. The conclusions are presented in Section 6.

2. Statistical model and improved estimators

Zero-inflated negative binomial model is suitable for analysing count data with excess zeros and overdispersion problems. Suppose that the dependent variable Y_i , $i = 1, 2, \dots, n$ is generated independently from a zero-inflated negative binomial distribution, such that Y_i has a degenerate distribution at zero with probability τ_i and follows a negative binomial distribution with mean μ_i and overdispersion parameter $\theta > 0$ with probability $1 - \tau_i$. The probability distribution of a ZINB random variable is given by

$$f(y_i | \tau_i, \mu_i) = \begin{cases} \tau_i + (1 - \tau_i) \left(\frac{1}{1 + \theta \mu_i} \right)^{\frac{1}{\theta}} & \text{if } y_i = 0 \\ (1 - \tau_i) \left(\frac{\Gamma(y_i + \frac{1}{\theta})}{\Gamma(\frac{1}{\theta}) \Gamma(y_i + 1)} \right) \left(\frac{\theta \mu_i}{1 + \theta \mu_i} \right)^{y_i} \left(\frac{1}{1 + \theta \mu_i} \right)^{\frac{1}{\theta}} & \text{if } y_i > 0, \end{cases} \quad (2.1)$$

where $\log(\mu_i) = \mathbf{x}_i' \boldsymbol{\beta}$ denotes the log link related to the count part and $\log\left(\frac{\tau_i}{1 - \tau_i}\right) = \mathbf{z}_i' \boldsymbol{\gamma}$ indicates the logit link related to the not at risk class, $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})'$ and $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{iq})'$ are vectors of predictors related to the count component and not at risk class, respectively. $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)'$ is the parameter vector of count component and $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_q)'$ is the parameter vector in not at risk class. We assume that $\boldsymbol{\delta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')$ is a $k \times 1$ vector of all unknown parameters, such that $k = p + q$. The mean and variance of Y_i are as follows:

$$\begin{aligned} E(Y_i | \mathbf{x}_i) &= \mu_i(1 - \tau_i), \\ \text{Var}(Y_i | \mathbf{x}_i) &= \mu_i(1 - \tau_i)(1 + \mu_i(\tau_i + \theta)), \end{aligned}$$

when $\theta \rightarrow 0$, the zero-inflated negative binomial distribution reduces to the zero-inflated Poisson distribution.

In order to apply shrinkage strategies, we study two models. One is the full model that includes all k parameters and we estimate them with the maximum likelihood method which is named unrestricted estimator. The other model is the candidate sub-model containing only active parameters, where δ satisfies the linear restriction

$$\mathbf{H}\delta = \mathbf{h}, \tag{2.2}$$

where \mathbf{H} is a $p_2 \times k$ matrix of full rank and \mathbf{h} is a $p_2 \times 1$ vector of known constants, here p_2 is the number of inactive (non significant) parameters in δ .

The log-likelihood function of ZINB model based on a random sample $\mathbf{y} = (y_1, y_2, \dots, y_n)$ is given by

$$\begin{aligned} l = l(\beta, \gamma) = l(\mu_i, \tau_i; y_i) &= \sum_{i=1}^n \left\{ I(y_i = 0) \ln \left(\frac{\tau_i}{1 - \tau_i} + \left(\frac{1}{1 + \theta \mu_i} \right)^{\frac{1}{\theta}} \right) \right\} \\ &+ \sum_{i=1}^n \left\{ I(y_i > 0) \left[y_i \ln(\mu_i) + y_i \ln(\theta) - \left(y_i + \frac{1}{\theta} \right) \ln(1 + \theta \mu_i) \right. \right. \\ &+ \left. \left. \ln \left(\Gamma(y_i + \frac{1}{\theta}) \right) - \ln \left(\Gamma(y_i + 1) \right) - \ln \left(\Gamma(\frac{1}{\theta}) \right) \right] \right\} \\ &+ \sum_{i=1}^n \ln(1 - \tau_i), \end{aligned} \tag{2.3}$$

where $I(y_i = 0)$ is an indicator function that equals to 1 if $y_i = 0$ and 0 otherwise. Also, $I(y_i > 0) = 1 - I(y_i = 0)$.

The MLE or unrestricted estimator of $\delta = (\beta', \gamma')'$ denoted by $\hat{\delta} = (\hat{\beta}', \hat{\gamma}')'$ obtains by solving the first order derivatives of log-likelihood function (2.3) with respect to β_j and γ_t that are given by

$$\begin{aligned} \frac{\partial l(\beta, \gamma)}{\partial \beta_j} &= - \sum_{i=1}^n I(y_i = 0) x_{ij} e^{\mathbf{x}'_i \beta} \left\{ \frac{\left(\frac{1}{1 + \theta e^{\mathbf{x}'_i \beta}} \right)^{\frac{1}{\theta} + 1}}{e^{\mathbf{z}'_i \gamma} + \left(\frac{1}{1 + \theta e^{\mathbf{x}'_i \beta}} \right)^{\frac{1}{\theta}}} \right\} \\ &+ \sum_{i=1}^n I(y_i > 0) x_{ij} \left\{ y_i - \theta \left(y_i + \frac{1}{\theta} \right) \left(\frac{e^{\mathbf{x}'_i \beta}}{1 + \theta e^{\mathbf{x}'_i \beta}} \right) \right\} = 0, \end{aligned} \tag{2.4}$$

$$\begin{aligned} \frac{\partial l(\beta, \gamma)}{\partial \gamma_t} &= \sum_{i=1}^n I(y_i = 0) z_{it} \left\{ \frac{e^{\mathbf{z}'_i \gamma}}{e^{\mathbf{z}'_i \gamma} + \left(\frac{1}{1 + \theta e^{\mathbf{x}'_i \beta}} \right)^{\frac{1}{\theta}}} \right\} \\ &- \sum_{i=1}^n I(y_i = 0) z_{it} \left(\frac{e^{\mathbf{z}'_i \gamma}}{1 + e^{\mathbf{z}'_i \gamma}} \right) - \sum_{i=1}^n I(y_i > 0) z_{it} \left(\frac{e^{\mathbf{z}'_i \gamma}}{1 + e^{\mathbf{z}'_i \gamma}} \right) = 0, \end{aligned} \tag{2.5}$$

where $j = 1, 2, \dots, p$ and $t = 1, 2, \dots, q$. It's clear that Equations (2.4) and (2.5) are nonlinear in β_j and γ_t therefore, the unrestricted estimator must be obtained using the numerical methods as Expectation Maximization (EM) and Newton-Raphson algorithms. Under certain regularity conditions, as $n \rightarrow \infty$, $\sqrt{n}(\hat{\delta} - \delta) \sim N_k(\mathbf{0}, \mathbf{I}^{-1})$, where $\mathbf{0}$ is a $k \times 1$ vector with zero elements and \mathbf{I} is the expected Fisher information matrix that can be partitioned as follows:

$$\mathbf{I} = \begin{bmatrix} I_{\beta, \beta} & I_{\beta, \gamma} \\ I_{\gamma, \beta} & I_{\gamma, \gamma} \end{bmatrix}, \tag{2.6}$$

where the elements of $I_{\beta,\beta}$, $I_{\beta,\gamma} = I'_{\gamma,\beta}$, and $I_{\gamma,\gamma}$ are defined as

$$I_{\beta,\beta} = -E\left(\frac{\partial^2 l}{\partial\beta\partial\beta'}\right), \quad I_{\beta,\gamma} = I'_{\gamma,\beta} = -E\left(\frac{\partial^2 l}{\partial\beta\partial\gamma'}\right), \quad I_{\gamma,\gamma} = -E\left(\frac{\partial^2 l}{\partial\gamma\partial\gamma'}\right),$$

where for $j, r = 1, 2, \dots, p$ and $t, s = 1, 2, \dots, q$,

$$\frac{\partial^2 l}{\partial\beta_j\partial\beta_r} = -\sum_{i=1}^n x_{ij}x_{ir} \frac{\mu_i}{(1+\theta\mu_i)^2} \left\{ I(y_i = 0) \left(\frac{(1-\tau_i)[1-\tau_i+\tau_i(1-\mu_i)(1+\theta\mu_i)^{\frac{1}{\theta}}]}{[1-\tau_i+\tau_i(1+\theta\mu_i)^{\frac{1}{\theta}}]^2} \right) + I(y_i > 0)(\theta y_i + 1) \right\},$$

$$\frac{\partial^2 l}{\partial\beta_j\partial\gamma_t} = \sum_{i=1}^n x_{ij}z_{it} \left\{ I(y_i = 0) \frac{\mu_i\tau_i(1-\tau_i)}{(1+\theta\mu_i)^{1-\frac{1}{\theta}}[1-\tau_i+\tau_i(1+\theta\mu_i)^{\frac{1}{\theta}}]^2} \right\},$$

$$\frac{\partial^2 l}{\partial\gamma_t\partial\gamma_s} = \sum_{i=1}^n z_{it}z_{is}\tau_i(1-\tau_i) \left\{ I(y_i = 0) \left(\frac{(1+\theta\mu_i)^{\frac{1}{\theta}}}{[1-\tau_i+\tau_i(1+\theta\mu_i)^{\frac{1}{\theta}}]^2} - 1 \right) - I(y_i > 0) \right\}.$$

We denote the restricted estimator of $\delta = (\beta', \gamma')'$ by $\tilde{\delta} = (\tilde{\beta}', \tilde{\gamma}')'$ that obtain by maximizing the log-likelihood function (2.3) under the linear restriction $\mathbf{H}\delta - \mathbf{h} = \mathbf{0}$. Based on [12], the restricted estimator is given by

$$\tilde{\delta} = \hat{\delta} - \mathbf{I}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{I}^{-1}\mathbf{H}')^{-1}(\mathbf{H}\hat{\delta} - \mathbf{h}). \tag{2.7}$$

The likelihood ratio test statistic \mathbf{T}_n for testing $H_0 : \mathbf{H}\delta = \mathbf{h}$ versus $H_1 : \mathbf{H}\delta \neq \mathbf{h}$ is defined as

$$\mathbf{T}_n = 2 \left(l(\hat{\delta}) - l(\tilde{\delta}) \right), \tag{2.8}$$

where $l(\hat{\delta})$ and $l(\tilde{\delta})$ are values of the log-likelihood function at the unrestricted and restricted estimators, respectively. Under H_0 , as $n \rightarrow \infty$, the test statistic \mathbf{T}_n is asymptotically distributed as $\chi^2(p_2)$ [3].

We now introduce the linear shrinkage, preliminary test, shrinkage pretest, Stein-type, and positive Stein-type estimators in ZINB model.

2.1. The linear shrinkage estimator

Based on the level of confidence subspace information $0 \leq \lambda \leq 1$, the linear shrinkage (LS) estimator $\hat{\delta}^{LS}$ of δ is defined as a linear function of $\hat{\delta}$ and $\tilde{\delta}$, as

$$\hat{\delta}^{LS} = \lambda \tilde{\delta} + (1 - \lambda) \hat{\delta}. \tag{2.9}$$

The optimum value of λ is chosen by minimizing the mean squared error of LS estimator [30]. Larger values of λ indicate more assurance in subspace information.

2.2. The preliminary test estimator

The preliminary test or pretest estimator $\hat{\delta}^{PT}$ of δ is a function of the unrestricted and restricted estimators and the test statistic \mathbf{T}_n as follows:

$$\hat{\delta}^{PT} = \hat{\delta} - (\hat{\delta} - \tilde{\delta}) I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha}), \tag{2.10}$$

where $I(\cdot)$ is an indicator function. If $\mathbf{T}_n \leq \mathbf{T}_{n,\alpha}$, then $\hat{\delta}^{PT} = \tilde{\delta}$, otherwise $\hat{\delta}^{PT} = \hat{\delta}$, and $\mathbf{T}_{n,\alpha}$ is the upper value of a $\chi^2(p_2)$ distribution. In other words, the test statistic \mathbf{T}_n is compared to $\chi^2_{1-\alpha}(p_2)$.

2.3. The shrinkage pretest estimator

Ahmed [1] proposed the shrinkage pretest (*SP*) estimator $\hat{\delta}^{SP}$ of δ by replacing $\tilde{\delta}$ with $\hat{\delta}^{LS}$ in Equation (2.10):

$$\hat{\delta}^{SP} = \hat{\delta} - (\hat{\delta} - \hat{\delta}^{LS}) I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha}) \tag{2.11}$$

The performance of the *SP* estimator is better than the pretest estimator when candidate sub-model is incorrect.

2.4. The Stein-type estimator

The Stein-type estimator $\hat{\delta}^S$ of δ is a function of the test statistic \mathbf{T}_n , the unrestricted and restricted estimators as follows:

$$\hat{\delta}^S = \hat{\delta} - \left(\frac{p_2 - 2}{\mathbf{T}_n}\right)(\hat{\delta} - \tilde{\delta}), \quad p_2 \geq 3 \tag{2.12}$$

This strategy shrinks the unrestricted estimator to a sub-model or restricted estimator.

2.5. The positive Stein-type estimator

The positive Stein-type estimator $\hat{\delta}^{S+}$ of δ controls the over-shrinking problem in Stein-type estimator is given by

$$\hat{\delta}^{S+} = \hat{\delta} - \left(\frac{p_2 - 2}{\mathbf{T}_n}\right)^+ (\hat{\delta} - \tilde{\delta}), \quad p_2 \geq 3, \tag{2.13}$$

where $z^+ = \max(0, z)$.

3. Asymptotic distribution bias and risk

In this section, following [14, 15], among others, we study the asymptotic distributional biases and risks of the estimators defined in Section 2 when the subspace information $\mathbf{H}\delta = \mathbf{h}$ is wrong. For fixed alternatives, the test statistics \mathbf{T}_n converges in probability to ∞ , as $n \rightarrow \infty$. So, the pretest, shrinkage pretest, Stein-type, and positive Stein-type estimators, which are dependent on \mathbf{T}_n , will be asymptotically converges to unrestricted estimator. Thus, their asymptotic risks yield to be the same [17]. Therefore, the asymptotic properties of the estimators are studied under the sequence of local alternatives that is given by

$$\mathcal{K}_{(n)} : \mathbf{H}\delta = \mathbf{h} + \frac{\omega}{\sqrt{n}}, \tag{3.1}$$

where $\omega = (\omega_1, \omega_2, \dots, \omega_{p_2})' \in R^{p_2}$ is a vector with real numbers. The vector $\frac{\omega}{\sqrt{n}}$ is the distance between the true model and the restricted model. To obtain the asymptotic risks (ADR) of the estimators, we define the following weighted quadratic loss function:

$$\mathcal{L}(\hat{\delta}^*, \delta; \mathbf{Q}) = \left(\sqrt{n}(\hat{\delta}^* - \delta)\right)' \mathbf{Q} \left(\sqrt{n}(\hat{\delta}^* - \delta)\right), \tag{3.2}$$

where $\hat{\delta}^*$ is any of the proposed estimators in the previous section and \mathbf{Q} is a weight matrix. When \mathbf{Q} is chosen as the identity matrix \mathbf{I} , the usual quadratic loss function is defined, which we use it in our simulation studies in next section. The cumulative distribution function of $\hat{\delta}^*$ under $\mathcal{K}_{(n)}$ is given by

$$G(\mathbf{x}) = \lim_{n \rightarrow \infty} P\left(\sqrt{n}(\hat{\delta}^* - \delta) \leq \frac{\mathbf{x}}{\mathcal{K}_{(n)}}\right), \tag{3.3}$$

where $G(\mathbf{x})$ is a non-degenerate distribution function. Now, we can define the asymptotic distributional risk by

$$ADR(\hat{\delta}^*; \mathbf{Q}) = \int \dots \int \mathbf{x}' \mathbf{Q} \mathbf{x} dG(\mathbf{x}) = trace(\mathbf{Q}\mathbf{V}), \tag{3.4}$$

where $\mathbf{V}(\hat{\boldsymbol{\delta}}^*)$ is the dispersion matrix for the distribution $G(\mathbf{x})$ and defined as

$$\mathbf{V}(\hat{\boldsymbol{\delta}}^*) = \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}}^* - \boldsymbol{\delta}) \sqrt{n}(\hat{\boldsymbol{\delta}}^* - \boldsymbol{\delta})'\right) = \int \dots \int \mathbf{x}'\mathbf{x} \, dG(\mathbf{x}). \tag{3.5}$$

The ADB of an estimator $\hat{\boldsymbol{\delta}}^*$ is defined as

$$ADB(\hat{\boldsymbol{\delta}}^*) = \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}}^* - \boldsymbol{\delta})\right) = \int \dots \int \mathbf{x} \, dG(\mathbf{x}). \tag{3.6}$$

We now present the following lemmas which are useful to obtain the ADB and ADR of all estimators.

Lemma 3.1. *Under the sequence of local alternatives $\mathcal{K}_{(n)}$ and the regularity conditions, as $n \rightarrow \infty$*

$$\begin{aligned} \mathbf{Z}_n &= \sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1}), \\ \mathbf{R}_n &= \sqrt{n}(\mathbf{H}\hat{\boldsymbol{\delta}} - \mathbf{h}) \xrightarrow{D} \mathbf{R} \sim \mathcal{N}_k(\boldsymbol{\omega}, \mathbf{H}\mathbf{I}^{-1}\mathbf{H}'), \\ \mathbf{V}_n &= \sqrt{n}(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}) \xrightarrow{D} \mathbf{V} \sim \mathcal{N}_k(-\mathcal{J}\boldsymbol{\omega}, \mathbf{I}^{-1} - \mathcal{J}\mathbf{H}\mathbf{I}^{-1}), \\ \mathbf{B}_n &= \sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) \xrightarrow{D} \mathbf{B} \sim \mathcal{N}_k(\mathcal{J}\boldsymbol{\omega}, \mathcal{J}\mathbf{H}\mathbf{I}^{-1}), \\ \begin{pmatrix} \mathbf{Z}_n \\ \mathbf{B}_n \end{pmatrix} &\xrightarrow{D} \begin{pmatrix} \mathbf{Z} \\ \mathbf{B} \end{pmatrix} \sim \mathcal{N}_{2k} \left[\begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix}, \begin{pmatrix} \mathbf{I}^{-1} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} \\ \mathcal{J}\mathbf{H}\mathbf{I}^{-1} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} \end{pmatrix} \right], \\ \begin{pmatrix} \mathbf{V}_n \\ \mathbf{B}_n \end{pmatrix} &\xrightarrow{D} \begin{pmatrix} \mathbf{V} \\ \mathbf{B} \end{pmatrix} \sim \mathcal{N}_{2k} \left[\begin{pmatrix} -\mathcal{J}\boldsymbol{\omega} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix}, \begin{pmatrix} \mathbf{I}^{-1} - \mathcal{J}\mathbf{H}\mathbf{I}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} \end{pmatrix} \right], \end{aligned}$$

where $\mathcal{J} = \mathbf{I}^{-1}\mathbf{H}'(\mathbf{H}\mathbf{I}^{-1}\mathbf{H}')^{-1}$.

Proof. See Appendix 1. □

Lemma 3.2. *Under the sequence of local alternatives $\mathcal{K}_{(n)}$ and the regularity conditions, as $n \rightarrow \infty$, the test statistic \mathbf{T}_n converges to $\chi_{p_2}^2(\Delta^*)$ with p_2 degrees of freedom and non-centrality parameter $\Delta^* = \boldsymbol{\omega}'(\mathbf{H}\mathbf{I}^{-1}\mathbf{H}')^{-1}\boldsymbol{\omega}$.*

Proof. See [10]. □

Lemma 3.3. *Let \mathbf{y} be a p_2 -dimensional random vector distributed as $\mathcal{N}_{p_2}(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)$. Then, for any measurable function φ , we have*

$$E[\mathbf{y}\varphi(\mathbf{y}'\mathbf{y})] = \boldsymbol{\mu}_y E[\varphi(\chi_{p_2+2}^2(\Delta^*))], \tag{3.7}$$

$$E[\mathbf{y}'\mathbf{y}\varphi(\mathbf{y}'\mathbf{y})] = \boldsymbol{\Sigma}_y E[\varphi(\chi_{p_2+2}^2(\Delta^*))] + \boldsymbol{\mu}'_y \boldsymbol{\mu}_y E[\varphi(\chi_{p_2+4}^2(\Delta^*))], \tag{3.8}$$

where Δ^* is the non-centrality parameter.

Proof. See [16]. □

Now, we present the asymptotic properties of the proposed estimators in the following theorems.

Theorem 3.4. Under the sequence of local alternatives $\mathcal{K}_{(n)}$ and the regularity conditions, as $n \rightarrow \infty$, the ADBs of the estimators are given as

$$\begin{aligned} ADB(\hat{\delta}) &= \mathbf{0}, \\ ADB(\tilde{\delta}) &= -\mathbf{J}\omega, \\ ADB(\hat{\delta}^{LS}) &= -\lambda \mathbf{J}\omega, \\ ADB(\hat{\delta}^{PT}) &= -\mathbf{J}\omega \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*), \\ ADB(\hat{\delta}^{SP}) &= -\lambda \mathbf{J}\omega \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*), \\ ADB(\hat{\delta}^S) &= -(p_2 - 2) \mathbf{J}\omega E\left(\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right), \\ ADB(\hat{\delta}^{S+}) &= ADB(\hat{\delta}^S) - \mathbf{J}\omega \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) \\ &\quad + (p_2 - 2) \mathbf{J}\omega E\left(\frac{I(\chi_{p_2+2}^2(\Delta^*) < p_2 - 2)}{\chi_{p_2+2}^2(\Delta^*)}\right). \end{aligned}$$

where $\Psi_v(\cdot; \Delta^*)$ is the cumulative distribution function of the $\chi_v^2(\Delta^*)$ distribution.

Proof. See Appendix 2. □

Theorem 3.5. Under the local alternatives $\mathcal{K}_{(n)}$ and the regularity conditions, as $n \rightarrow \infty$, the ADRs of the estimators are

$$\begin{aligned} ADR(\hat{\delta}; \mathbf{Q}) &= \text{trace}(\mathbf{Q}\mathbf{I}^{-1}), \\ ADR(\tilde{\delta}; \mathbf{Q}) &= ADR(\hat{\delta}; \mathbf{Q}) - \text{trace}(\mathbf{Q}\mathbf{J}\mathbf{H}\mathbf{I}^{-1}) + \Delta_Q^*, \\ ADR(\hat{\delta}^{LS}; \mathbf{Q}) &= ADR(\hat{\delta}; \mathbf{Q}) - \lambda(2 - \lambda) \text{trace}(\mathbf{Q}\mathbf{J}\mathbf{H}\mathbf{I}^{-1}) + \lambda^2 \Delta_Q^*, \\ ADR(\hat{\delta}^{PT}; \mathbf{Q}) &= ADR(\hat{\delta}; \mathbf{Q}) - \text{trace}(\mathbf{Q}\mathbf{J}\mathbf{H}\mathbf{I}^{-1}) \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) \\ &\quad + \Delta_Q^* [2\Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) - \Psi_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^*)], \\ ADR(\hat{\delta}^{SP}; \mathbf{Q}) &= ADR(\hat{\delta}; \mathbf{Q}) - \lambda(2 - \lambda) \text{trace}(\mathbf{Q}\mathbf{J}\mathbf{H}\mathbf{I}^{-1}) \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) \\ &\quad + \Delta_Q^* [2\lambda \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) - \lambda(2 - \lambda) \Psi_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^*)], \\ ADR(\hat{\delta}^S; \mathbf{Q}) &= ADR(\hat{\delta}; \mathbf{Q}) - 2(p_2 - 2) \text{trace}(\mathbf{Q}\mathbf{J}\mathbf{H}\mathbf{I}^{-1}) \left\{ E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] \right. \\ &\quad \left. - (p_2 - 2) E\left[\frac{1}{(\chi_{p_2+2}^2(\Delta^*))^2}\right] \right\} + (p_2 - 2) \Delta_Q^* \left\{ 2 E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] \right. \\ &\quad \left. - 2 E\left[\frac{1}{(\chi_{p_2+2}^2(\Delta^*))^2}\right] + (p_2 - 2) E\left[\frac{1}{(\chi_{p_2+4}^2(\Delta^*))^2}\right] \right\}, \\ ADR(\hat{\delta}^{S+}; \mathbf{Q}) &= ADR(\hat{\delta}^S; \mathbf{Q}) - \Delta_Q^* E\left(\left(1 - \frac{p_2 - 2}{\chi_{p_2+4}^2(\Delta^*)}\right)^2 I(\chi_{p_2+4}^2(\Delta^*) < p_2 - 2)\right) \\ &\quad - \text{trace}(\mathbf{Q}\mathbf{J}\mathbf{H}\mathbf{I}^{-1}) E\left(\left(1 - \frac{p_2 - 2}{\chi_{p_2+2}^2(\Delta^*)}\right)^2 I(\chi_{p_2+4}^2(\Delta^*) < p_2 - 2)\right) \\ &\quad + 2 \Delta_Q^* E\left(\left(1 - \frac{p_2 - 2}{\chi_{p_2+4}^2(\Delta^*)}\right) I(\chi_{p_2+4}^2(\Delta^*) < p_2 - 2)\right). \end{aligned}$$

where $\Delta_Q^* = \omega'(\mathbf{J}'\mathbf{Q}\mathbf{J})\omega$ and $I(\cdot)$ is an indicator function.

Proof. See Appendix 3. □

By comparing the ADRs of the estimators, we see that when $\Delta^* = 0$ or is near zero, the risk of the restricted estimator becomes unbounded, i.e. $ADR(\tilde{\delta}) < ADR(\hat{\delta})$. The linear shrinkage estimator performs better than the unrestricted estimator at and near $\Delta^* = 0$. The risk of the pretest estimator is smaller than that of the shrinkage pretest estimator at $\Delta^* = 0$. As Δ^* increases, the shrinkage pretest estimator dominates the pretest estimator. By comparing the risks of the Stein-type, positive Stein-type, and unrestricted estimators, it's clear that $ADR(\hat{\delta}^{S^+}) \leq ADR(\hat{\delta}^S) \leq ADR(\hat{\delta})$ for all $\Delta^* \geq 0$.

4. A Monte Carlo simulation experiment

Based on many studies in the field of shrinkage strategies, such as [17–19], we assess the performance of the suggested estimators with respect to the unrestricted estimator via Monte Carlo simulation study with statistical software *R* to confirm the theoretical results. To obtain the unrestricted estimator of parameters, we use *zeroinfl* function in *pscl* package. We generate the response variable form ZINB model with sample size of $n = 300$ using the following mixture of two models

$$\log(\mu_i) = \mathbf{x}'_i \boldsymbol{\beta}, \quad \log\left(\frac{\tau_i}{1 - \tau_i}\right) = \mathbf{z}'_i \boldsymbol{\gamma}, \quad i = 1, 2, \dots, n,$$

where $\mu_i = e^{\mathbf{x}'_i \boldsymbol{\beta}}$, $\tau_i = \frac{e^{\mathbf{z}'_i \boldsymbol{\gamma}}}{1 + e^{\mathbf{z}'_i \boldsymbol{\gamma}}}$, the \mathbf{x}_i are generated from $N_p(\mathbf{0}, \mathbf{S})$ with covariances $s_{ts} = \varphi^{|t-s|}$ for $t, s = 1, 2, \dots, p$, $p = 6, 8$ and the correlation $\varphi = 0.5$. Also, \mathbf{z}_i is generated from uniform distribution $U(0, 1)$. We consider the number of predictors $k = p + q = 9, 12, 15$ with $(p, q) = (6, 3), (6, 6), (8, 7)$. We select the level of significance $\alpha = 0.05, 0.50, 0.10$ for the pretest and shrinkage pretest methods and the level of confidence in the prior information for the LS and SP estimators is set to $\lambda = 0.25, 0.50, 0.75$. We consider the parameter vector of $\boldsymbol{\delta}$ as

$$\boldsymbol{\delta}_{k \times 1} = (\boldsymbol{\beta}'_{p \times 1}, \boldsymbol{\gamma}'_{q \times 1})' = (\beta_1, \beta_2, \dots, \beta_p, \gamma_1, \gamma_2, \dots, \gamma_q)'$$

To check the performance of the proposed estimators, we consider two restrictions on $\boldsymbol{\delta}$ as follows:

- **Case (i):** When $(p, q) = (6, 3), (6, 6)$, we assume that $\underbrace{\beta_p, \gamma_2, \gamma_3, \dots, \gamma_q}_{p_2}$ are the inactive parameters. Thus, $p_2 = 3, 6$ and the \mathbf{H} matrix and \mathbf{h} vector in the linear restriction (2.2) are as

$$\mathbf{H} = [\mathbf{0}_{p_2 \times (p-1)}, \mathbf{b}^1_{p_2 \times 1}, \mathbf{0}_{p_2 \times 1}, \mathbf{b}^2_{p_2 \times (k-(p+1))}], \quad \mathbf{h} = \mathbf{0}_{p_2 \times 1},$$

where $\mathbf{0}_{s \times t}$ is a matrix (vector) with zero elements and \mathbf{b}^1 and \mathbf{b}^2 are vector and matrix with one and zero elements. We set $\boldsymbol{\eta}_1 = (\beta_p, \gamma_2, \gamma_3, \gamma_q)'$ as the inactive parameters vector.

- **Case (ii):** When $(p, q) = (8, 7)$, we set $\underbrace{\beta_6, \beta_7, \beta_8, \gamma_2, \gamma_3, \dots, \gamma_7}_{p_2}$ as non significant parameters. In this case, $p_2 = 9$ and the \mathbf{H} and \mathbf{h} are as follows:

$$\mathbf{H}_{9 \times 15} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{h}_{9 \times 1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We set $\boldsymbol{\eta}_2 = (\beta_6, \beta_7, \beta_8, \gamma_2, \gamma_3, \dots, \gamma_7)'$ as the inactive parameters vector.

We test the following hypothesis:

$$\begin{aligned} \text{Case}(i) : H_0 : \boldsymbol{\eta}_1 = \mathbf{0} \text{ versus } H_1 : \boldsymbol{\eta}_1 \neq \mathbf{0} \\ \text{Case}(ii) : H_0 : \boldsymbol{\eta}_2 = \mathbf{0} \text{ versus } H_1 : \boldsymbol{\eta}_2 \neq \mathbf{0} \end{aligned}$$

We next define the parameter Δ as follows:

$$\Delta = \|\boldsymbol{\delta} - \boldsymbol{\delta}^{true}\|^2 = \sum_{r=1}^k (\delta_r - \delta_r^{true})^2, \tag{4.1}$$

where $\|\cdot\|$ is the Euclidean norm, $\boldsymbol{\delta}$ is the parameter vector in the simulated model, and $\boldsymbol{\delta}^{true}$ is the true parameter. We set $\Delta \in [0, 1]$ in our simulations. Samples are generated using the overdispersion parameter $\theta = 1.5$, and the true parameter as

$$\begin{aligned} \text{Case}(i) : \boldsymbol{\delta}^{true} &= (0.2, 0.5, 0.05, -0.15, 1.2, 0, 0.3, \underbrace{0, 0, \dots, 0}_{p_2-1})', p_2 = 3, 6 \\ \text{Case}(ii) : \boldsymbol{\delta}^{true} &= (0.2, 0.5, 0.05, -0.15, 1.2, 0, 0, 0, 0.3, \underbrace{0, 0, \dots, 0}_{p_2-3})', p_2 = 9 \end{aligned}$$

Based on the definition of Δ , when candidate sub-model is correct, i.e. H_0 is true, $\boldsymbol{\delta} = \boldsymbol{\delta}^{true}$ and $\Delta = 0$. When candidate sub-model is incorrect ($\Delta > 0$), the elements of $\boldsymbol{\delta}$ are as

$$\begin{aligned} \text{Case}(i) : \boldsymbol{\delta} &= (0.2, 0.5, 0.05, -0.15, 1.2, c, 0.3, \underbrace{0, 0, \dots, 0}_{p_2-1})', p_2 = 3, 6 \\ \text{Case}(ii) : \boldsymbol{\delta} &= (0.2, 0.5, 0.05, -0.15, 1.2, c, 0, 0, 0.3, \underbrace{0, 0, \dots, 0}_{p_2-3})', p_2 = 9 \end{aligned}$$

where c is a scalar with various values. We set $c = \{0.0, 0.44, 0.63, 0.77, 0.89, 1.0\}$, thus $\Delta = c^2 = \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0\}$. The number of replications is set to 1000 for all cases. The R codes related to this paper can be provided by the author upon request by email at (zahrazandi163@yahoo.com). Based on the simulated data, we estimate the simulated mean squared errors (*SMSE*) of all the estimators defined in Section 2 as follows:

$$SMSE(\hat{\boldsymbol{\delta}}^*) = \frac{1}{1000} \sum_{r=1}^{1000} (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}^*)_r' (\boldsymbol{\delta} - \hat{\boldsymbol{\delta}}^*)_r$$

To compare the performance of the estimators, we use the simulated relative efficiency (*SRE*) criterion that is defined as follows:

$$SRE(\hat{\boldsymbol{\delta}}, \hat{\boldsymbol{\delta}}^*) = \frac{SMSE(\hat{\boldsymbol{\delta}})}{SMSE(\hat{\boldsymbol{\delta}}^*)}, \tag{4.2}$$

where $\hat{\boldsymbol{\delta}}$ is the unrestricted estimator. A value of SRE greater than one indicates that $\hat{\boldsymbol{\delta}}^*$ performs better than $\hat{\boldsymbol{\delta}}$.

4.1. The results of simulation

The simulated relative efficiencies (*SREs*) for all the proposed estimators are reported in Table 1 for $\lambda = 0.25, 0.50, \text{ and } 0.75$ when $\Delta = 0$. The *SREs* for all estimators are greater than one and the performance of the restricted estimator is the best. The pretest estimator performs better than the shrinkage pretest estimator at $\Delta = 0$. The simulation results in Tables 2-4 and Figure 1 reveal that when Δ moves away from zero, the *SREs* of all estimators decline sharply. The *SRE* of the pretest and shrinkage pretest estimators is inversely related to α . When α decreases, the *SRE* of the PT and SP increases, and vice versa. The shrinkage pretest estimator also depends on λ , such that its *SRE* increases when λ increases. When λ increases, the simulated relative efficiency of the linear shrinkage estimator increases. The positive Stein-type estimator is better than

the Stein-type estimator at all parameter space. The SRE of all estimators increases when p_2 increases.

Table 1. The simulated relative efficiencies of $\tilde{\delta}, \hat{\delta}^{LS}, \hat{\delta}^{PT}, \hat{\delta}^{SP}, \hat{\delta}^S$, and $\hat{\delta}^{S+}$ with respect to $\hat{\delta}$ at $\Delta = 0$.

| Estimator | | | Number of inactive predictors p_2 . | | |
|-----------|------------------|-----------------|---------------------------------------|--------|--------|
| | | | 3 | 6 | 9 |
| RE | | | 4.304 | 11.114 | 12.613 |
| LS | $\lambda = 0.25$ | | 1.522 | 1.668 | 1.682 |
| | $\lambda = 0.50$ | | 2.411 | 3.183 | 3.267 |
| | $\lambda = 0.75$ | | 3.660 | 6.928 | 7.444 |
| PT | | $\alpha = 0.01$ | 4.219 | 10.516 | 12.609 |
| | | $\alpha = 0.05$ | 3.923 | 9.166 | 11.638 |
| | | $\alpha = 0.10$ | 3.128 | 6.531 | 9.181 |
| SP | $\lambda = 0.25$ | $\alpha = 0.01$ | 1.517 | 1.662 | 1.682 |
| | | $\alpha = 0.05$ | 1.494 | 1.645 | 1.672 |
| | | $\alpha = 0.10$ | 1.437 | 1.594 | 1.645 |
| | $\lambda = 0.50$ | $\alpha = 0.01$ | 2.390 | 3.145 | 3.266 |
| | | $\alpha = 0.05$ | 2.313 | 3.042 | 3.212 |
| | | $\alpha = 0.10$ | 2.077 | 2.763 | 3.044 |
| | $\lambda = 0.75$ | $\alpha = 0.01$ | 3.601 | 6.704 | 7.443 |
| | | $\alpha = 0.05$ | 3.392 | 6.155 | 7.108 |
| | | $\alpha = 0.10$ | 2.810 | 4.905 | 6.163 |
| S | | | 1.274 | 2.413 | 2.064 |
| S^+ | | | 1.840 | 4.881 | 8.456 |

Table 2. The simulated relative efficiencies of $\tilde{\delta}, \hat{\delta}^{LS}, \hat{\delta}^{PT}, \hat{\delta}^{SP}, \hat{\delta}^S$, and $\hat{\delta}^{S+}$ with respect to $\hat{\delta}$ for $\lambda = 0.25$.

| p_2 | Δ | RE | LS | PT | | | SP | | | S | S^+ |
|-------|----------|--------|-------|----------|--------|-------|----------|-------|-------|-------|-------|
| | | | | α | | | α | | | | |
| | | | | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | | |
| 3 | 0.0 | 4.304 | 1.522 | 4.219 | 3.923 | 3.128 | 1.517 | 1.499 | 1.437 | 1.274 | 1.840 |
| | 0.2 | 3.434 | 1.510 | 3.334 | 2.894 | 2.412 | 1.501 | 1.449 | 1.390 | 1.205 | 1.626 |
| | 0.4 | 2.172 | 1.483 | 2.007 | 1.612 | 1.356 | 1.441 | 1.319 | 1.215 | 1.319 | 1.338 |
| | 0.6 | 1.351 | 1.442 | 1.142 | 1.017 | 1.000 | 1.268 | 1.107 | 1.056 | 1.188 | 1.188 |
| | 0.8 | 0.885 | 1.391 | 0.872 | 0.927 | 0.963 | 1.074 | 1.007 | 1.002 | 1.115 | 1.115 |
| | 1.0 | 0.613 | 1.330 | 0.899 | 0.979 | 0.990 | 1.001 | 0.889 | 0.999 | 1.077 | 1.077 |
| 6 | 0.0 | 11.114 | 1.668 | 10.156 | 9.166 | 6.531 | 1.662 | 1.645 | 1.594 | 2.413 | 4.881 |
| | 0.2 | 8.713 | 1.663 | 8.341 | 6.499 | 4.396 | 1.657 | 1.616 | 1.535 | 2.561 | 3.926 |
| | 0.4 | 5.388 | 1.650 | 4.755 | 3.134 | 2.404 | 1.419 | 1.498 | 1.401 | 2.386 | 2.653 |
| | 0.6 | 3.306 | 1.629 | 2.292 | 1.543 | 1.292 | 1.470 | 1.262 | 1.161 | 1.904 | 1.923 |
| | 0.8 | 2.149 | 1.602 | 1.259 | 1.067 | 1.026 | 1.210 | 1.071 | 1.036 | 1.561 | 1.561 |
| | 1.0 | 1.483 | 1.568 | 1.000 | 0.990 | 0.993 | 1.051 | 1.010 | 1.003 | 1.371 | 1.371 |
| 9 | 0.0 | 12.613 | 1.682 | 12.609 | 11.638 | 9.181 | 1.682 | 1.672 | 1.645 | 2.064 | 8.456 |
| | 0.2 | 10.153 | 1.678 | 10.150 | 9.485 | 6.846 | 1.678 | 1.654 | 1.620 | 2.677 | 6.752 |
| | 0.4 | 6.472 | 1.668 | 6.327 | 4.782 | 4.005 | 1.663 | 1.601 | 1.556 | 3.248 | 4.260 |
| | 0.6 | 4.043 | 1.651 | 3.375 | 2.270 | 1.771 | 1.589 | 1.428 | 1.310 | 2.690 | 2.794 |
| | 0.8 | 2.653 | 1.629 | 1.742 | 1.274 | 1.134 | 1.382 | 1.181 | 1.099 | 2.068 | 2.072 |
| | 1.0 | 1.840 | 1.602 | 1.110 | 1.021 | 1.009 | 1.137 | 1.039 | 1.019 | 1.698 | 1.698 |

Table 3. The simulated relative efficiencies of $\tilde{\delta}, \hat{\delta}^{LS}, \hat{\delta}^{PT}, \hat{\delta}^{SP}, \hat{\delta}^S$, and $\hat{\delta}^{S+}$ with respect to $\hat{\delta}$ for $\lambda = 0.50$.

| p_2 | Δ | RE | LS | PT | | | SP | | | S | S^+ |
|-------|----------|--------|-------|----------|--------|-------|----------|-------|-------|-------|-------|
| | | | | α | | | α | | | | |
| | | | | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | | |
| 3 | 0.0 | 4.304 | 2.411 | 4.219 | 3.923 | 3.128 | 2.390 | 2.313 | 2.077 | 1.274 | 1.840 |
| | 0.2 | 3.434 | 2.317 | 3.334 | 2.849 | 2.412 | 2.282 | 2.090 | 1.893 | 1.205 | 1.626 |
| | 0.4 | 2.172 | 2.102 | 2.007 | 1.612 | 1.356 | 1.968 | 1.628 | 1.388 | 1.319 | 1.338 |
| | 0.6 | 1.351 | 1.826 | 1.142 | 1.017 | 1.000 | 1.434 | 1.152 | 1.076 | 1.188 | 1.188 |
| | 0.8 | 0.885 | 1.545 | 0.872 | 0.927 | 0.963 | 1.072 | 0.996 | 0.997 | 1.115 | 1.115 |
| | 1.0 | 0.613 | 1.292 | 0.899 | 0.979 | 0.990 | 0.983 | 0.995 | 0.997 | 1.077 | 1.077 |
| 6 | 0.0 | 11.114 | 3.183 | 10.516 | 9.166 | 6.531 | 3.145 | 3.042 | 2.763 | 2.413 | 4.881 |
| | 0.2 | 8.713 | 3.117 | 8.341 | 6.499 | 4.396 | 3.080 | 2.851 | 2.460 | 2.561 | 3.926 |
| | 0.4 | 5.388 | 2.949 | 4.755 | 3.134 | 2.404 | 2.790 | 2.255 | 1.919 | 2.386 | 2.653 |
| | 0.6 | 3.306 | 2.712 | 2.292 | 1.543 | 1.292 | 2.084 | 1.505 | 1.288 | 1.904 | 1.923 |
| | 0.8 | 2.149 | 2.439 | 1.259 | 1.067 | 1.026 | 1.362 | 1.110 | 1.053 | 1.561 | 1.561 |
| | 1.0 | 1.483 | 2.161 | 1.000 | 0.990 | 0.993 | 1.069 | 1.012 | 1.003 | 1.371 | 1.371 |
| 9 | 0.0 | 12.613 | 3.267 | 12.609 | 11.638 | 9.181 | 3.266 | 3.212 | 3.044 | 2.064 | 8.456 |
| | 0.2 | 10.153 | 3.215 | 10.150 | 9.485 | 6.846 | 3.214 | 3.069 | 2.880 | 2.677 | 6.752 |
| | 0.4 | 6.472 | 3.075 | 6.327 | 4.782 | 4.005 | 3.048 | 2.720 | 2.508 | 3.248 | 4.260 |
| | 0.6 | 4.043 | 2.869 | 3.375 | 2.270 | 1.771 | 2.574 | 1.971 | 1.634 | 2.690 | 2.794 |
| | 0.8 | 2.653 | 2.623 | 1.742 | 1.274 | 1.134 | 1.785 | 1.315 | 1.163 | 2.068 | 2.072 |
| | 1.0 | 1.840 | 2.364 | 1.110 | 1.021 | 1.009 | 1.215 | 1.056 | 1.027 | 1.698 | 1.698 |

Table 4. The simulated relative efficiencies of $\tilde{\delta}, \hat{\delta}^{LS}, \hat{\delta}^{PT}, \hat{\delta}^{SP}, \hat{\delta}^S$, and $\hat{\delta}^{S+}$ with respect to $\hat{\delta}$ for $\lambda = 0.75$.

| p_2 | Δ | RE | LS | PT | | | SP | | | S | S^+ |
|-------|----------|--------|-------|----------|--------|-------|----------|-------|-------|-------|-------|
| | | | | α | | | α | | | | |
| | | | | 0.01 | 0.05 | 0.10 | 0.01 | 0.05 | 0.10 | | |
| 3 | 0.0 | 4.304 | 3.660 | 4.219 | 3.923 | 3.128 | 3.601 | 3.392 | 2.810 | 1.274 | 1.840 |
| | 0.2 | 3.434 | 3.248 | 3.334 | 2.849 | 2.412 | 3.163 | 2.736 | 2.344 | 1.205 | 1.626 |
| | 0.4 | 2.172 | 2.472 | 2.007 | 1.612 | 1.356 | 2.257 | 1.759 | 1.446 | 1.319 | 1.338 |
| | 0.6 | 1.351 | 1.775 | 1.142 | 1.017 | 1.000 | 1.377 | 1.118 | 1.056 | 1.188 | 1.188 |
| | 0.8 | 0.885 | 1.275 | 0.872 | 0.927 | 0.963 | 0.995 | 0.969 | 0.983 | 1.115 | 1.115 |
| | 1.0 | 0.613 | 0.937 | 0.899 | 0.979 | 0.990 | 0.974 | 0.988 | 0.994 | 1.077 | 1.077 |
| 6 | 0.0 | 11.114 | 6.928 | 10.516 | 9.166 | 6.531 | 6.704 | 6.155 | 4.904 | 2.413 | 4.881 |
| | 0.2 | 8.713 | 6.303 | 8.341 | 6.499 | 4.396 | 6.117 | 5.109 | 3.771 | 2.561 | 3.926 |
| | 0.4 | 5.388 | 5.028 | 4.755 | 3.134 | 2.404 | 4.490 | 3.056 | 2.371 | 2.386 | 2.653 |
| | 0.6 | 3.306 | 3.774 | 2.292 | 1.543 | 1.292 | 2.509 | 1.622 | 1.336 | 1.904 | 1.923 |
| | 0.8 | 2.149 | 2.801 | 1.259 | 1.067 | 1.026 | 1.382 | 1.108 | 1.049 | 1.561 | 1.561 |
| | 1.0 | 1.483 | 2.106 | 1.000 | 0.990 | 0.993 | 1.051 | 1.005 | 1.000 | 1.371 | 1.371 |
| 9 | 0.0 | 12.613 | 7.444 | 12.609 | 11.638 | 9.181 | 7.443 | 7.108 | 6.163 | 2.064 | 8.456 |
| | 0.2 | 10.153 | 6.885 | 10.150 | 8.485 | 6.846 | 6.884 | 6.113 | 5.259 | 2.677 | 6.752 |
| | 0.4 | 6.472 | 5.654 | 6.327 | 4.782 | 4.005 | 5.545 | 4.364 | 3.735 | 3.248 | 4.260 |
| | 0.6 | 4.043 | 4.363 | 3.375 | 2.270 | 1.771 | 3.605 | 2.377 | 1.832 | 2.690 | 2.794 |
| | 0.8 | 2.653 | 3.308 | 1.742 | 1.274 | 1.134 | 1.959 | 1.352 | 1.176 | 2.068 | 2.072 |
| | 1.0 | 1.840 | 2.525 | 1.110 | 1.021 | 1.009 | 1.205 | 1.050 | 1.024 | 1.698 | 1.698 |

5. Wildlife fish data

We now apply the wildlife fish dataset which is analyzed by [4, 21]. The dataset consists of 250 groups that went to a state park to caught fish. The dependent and predictor variables are described in Table 5. The histogram of the dependent variable in Figure 2 confirms that there are excessive zeros in the data (about 50%), thus the dependent variable is inflated by zeros.

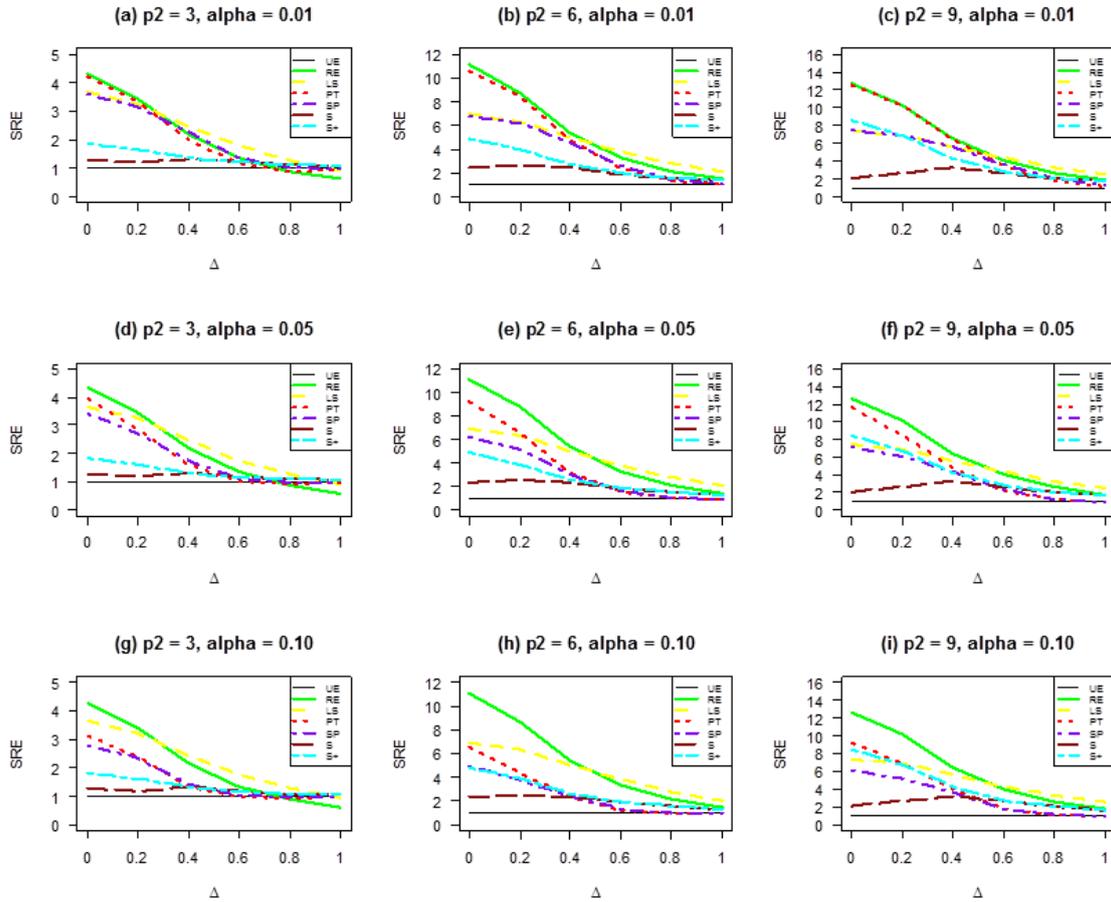


Figure 1. The simulated relative efficiencies of $\tilde{\delta}, \hat{\delta}^{LS}, \hat{\delta}^{PT}, \hat{\delta}^{SP}, \hat{\delta}^S,$ and $\hat{\delta}^{S^+}$ with respect to $\hat{\delta}$. Here, $p_2 = 3, 6, 9, \lambda = 0.75$.

Table 5. The dependent variable and predictors in the real data.

| Variable | Description |
|---------------------------|---|
| Dependent variable | |
| count | the number of fish that were caught |
| Predictors | |
| nofish $X_1 (Z_1)$ | whether the trip was not just for fishing? 0 if no and 1 if yes |
| livebait $X_2 (Z_2)$ | whether live bait was used or not? 0 if no and 1 if yes |
| camper $X_3 (Z_3)$ | whether or not they brought a camper? |
| persons $X_4 (Z_4)$ | how many total persons on the trip? |
| child $X_5 (Z_5)$ | how many children present? |

Following [13], the standard fit test for ZINB model is the Vuong test. By using it at 95% confidence level for our data, the value of Vuong statistic is $V = -5.13$. Since $V < -1.96$ indicates that ZINB is the preferred model. Following [15], all predictors are included in both count component and not at risk class of the ZINB model, so $p = 5, q = 5,$ and $k = p + q = 10$. To specify the active and inactive covariates, we have applied variable selection methods based on AIC and BIC . These criteria show that the coefficients of **nofish** (β_1), **persons** (β_4), and **child** (β_5) are the significant parameters for the count part and the coefficient of **child** (γ_5) is the significant parameter for the not at risk class.

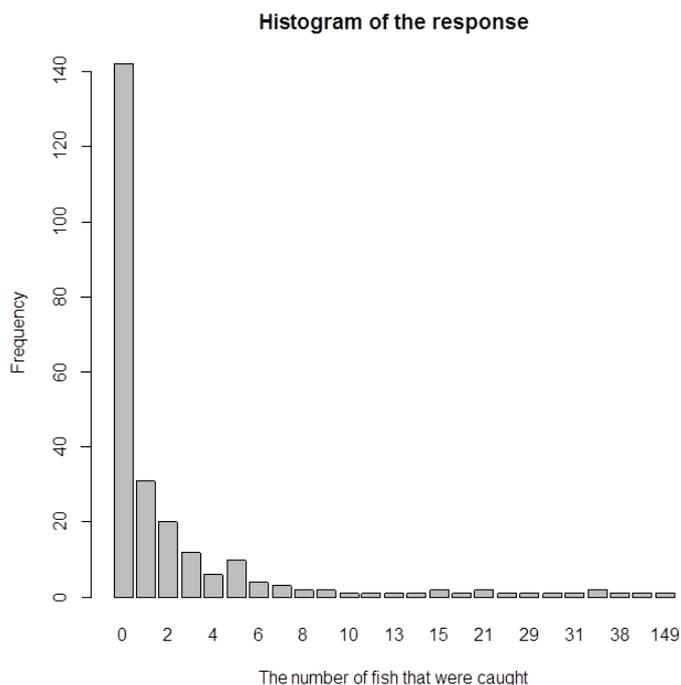


Figure 2. The Histogram of the dependent variable in the real data.

Thus, the linear restriction $\mathbf{H}\boldsymbol{\delta} = \mathbf{h}$ is as follows:

$$\mathbf{H} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Table 6. Estimates, standard errors (in parentheses), and the SRE of the significant parameters in the wildlife fish dataset with respect to the unrestricted estimator when $\Delta = 0$, $\alpha = 0.05$, and $\lambda = 0.75$.

| Estimator | $\hat{\beta}_1$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\gamma}_5$ | SRE |
|----------------------|-------------------|------------------|-------------------|------------------|--------|
| <i>UE</i> | -0.104 (0.339) | 0.075 (0.229) | -0.119 (0.372) | 0.798 (5.091) | 1000 |
| <i>RE</i> | -0.100 (0.325) | 0.075 (0.226) | -0.113 (0.355) | 0.048 (0.176) | 14.241 |
| <i>LS</i> | -0.101 (0.328) | 0.075 (0.226) | -0.115 (0.359) | 0.235 (1.327) | 8.094 |
| <i>PT</i> | -0.103 (0.332) | 0.073 (0.221) | -0.115 (0.362) | 0.421 (4.001) | 2.115 |
| <i>SP</i> | -0.103 (0.334) | 0.073 (0.223) | -0.116 (0.365) | 0.515 (4.080) | 1.987 |
| <i>S</i> | -0.102 (0.334) | 0.076 (0.241) | -0.119 (0.378) | 0.436 (4.261) | 1.689 |
| <i>S⁺</i> | -0.103 (0.334) | 0.077 (0.242) | -0.120 (0.380) | 0.543 (3.804) | 2.133 |

Thus, $\beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3, \gamma_4$ are equal to zero and $\beta_1, \beta_4, \beta_5, \gamma_5$ are the significant parameters. We have estimated the significant parameters based on the shrinkage estimators and compare their performance with the unrestricted estimator. Therefore, we have chosen $m = 150$ observations with replacement at 100 times from the original data set using bootstrap sampling to compute the relative efficiencies, estimation of the active estimators, and their standard errors. The results are reported in Table 6 for $\lambda = 0.75$ and $\alpha = 0.05$ at $\Delta = 0$.

6. Conclusion

In this paper, we considered the estimation of parameters in the zero-inflated negative binomial regression model based on the shrinkage estimators and we compared their performance to the unrestricted estimator. We computed the properties of the suggested estimators based on simulated relative efficiency using Monte Carlo experiments in R program.

Based on the simulation results, the performance of the restricted estimator was the best when candidate sub model was correct. The linear shrinkage estimator had much higher relative efficiency when λ approaches to one. The pretest estimator outperformed the shrinkage pretest estimator when subspace information was correct or nearly correct, while the shrinkage pretest estimator dominated the pretest estimator when Δ moved away from zero. The estimators based on the Stein-type and positive Stein-type strategies outperformed the unrestricted estimator. The performance of the positive Stein-type shrinkage estimator was better than that of the Stein-type estimator in all parts of the parameter space.

Acknowledgment. The authors are thankful to the reviewers for the insightful comments and suggestions that have resulted in a much improved version of this paper.

References

- [1] S.E. Ahmed, *Shrinkage preliminary test estimation in multivariate normal distributions*, J. Stat. Comput. Simul. **43** (3-4), 177-195, 1992.
- [2] S.E. Ahmed, *Penalty, Shrinkage and Pretest Strategies: Variable Selection and Estimation*, Springer, 2014.
- [3] J. Aitchison and S.D. Silvey, *Maximum-likelihood estimation of parameters subject to restraints*, Ann. Math. Stat. **29** (3), 813-828, 1958.
- [4] Y. Al-Taweel and Z. Algamal, *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Period. Eng. Nat. Sci. **8** (1), 248-255, 2020.
- [5] R. Arabi Belaghi, M. Arashi and S.M.M. Tabatabaey, *Improved estimators of the distribution function based on lower record values*, Statist. Papers **56** (2), 453-477, 2015.
- [6] M. Arashi, *Preliminary test and Stein estimations in simultaneous linear equations*, Linear Algebra Appl. **436** (5), 1195-1211, 2012.
- [7] M. Arashi, B.G. Kibria, M. Norouzrad and S. Nadarajah, *Improved preliminary test and Stein-rule Liu estimators for the ill-conditioned elliptical linear regression model*, J. Multivariate Anal. **126**, 53-74, 2014.
- [8] C.C. Astuti and A.D. Mulyanto, *Estimation parameters and modelling zero inflated negative binomial*, CAUCHY: Journal Matematika Murni dan Aplikasi **4** (3), 115-119, 2016.
- [9] T.A. Bancroft, *On biases in estimation due to the use of preliminary tests of significance*, Ann. Math. Stat. **15** (2), 190-204, 1944.

- [10] R.R. Davidson and W.E. Lever, *The limiting distribution of the likelihood ratio statistic under a class of local alternatives*, Sankhya A **32** (2), 209-224, 1970.
- [11] W.H. Greene, *Accounting for excess zeros and sample selection in Poisson and negative binomial regression models*, Working Paper 94-10, New York University, New York, 1994.
- [12] C.C. Heyde, *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*, Springer Science and Business Media, 2008.
- [13] J.M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, 2011.
- [14] S. Hossain, S.E. Ahmed and K.A. Doksum, *Shrinkage, pretest, and penalty estimators in generalized linear models*, Stat. Methodol. **24**, 52-68, 2015.
- [15] S. Hossain and H.A. Howlader, *Estimation techniques for regression model with zero-inflated Poisson data*, International Journal of Statistics and Probability **4** (4), 64-76, 2015.
- [16] G.G. Judge and M.E. Bock, *The Statistical Implication of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland, 1978.
- [17] S. Lisawadi, S.E. Ahmed and O. Reangsephet, *Post estimation and prediction strategies in negative binomial regression model*, Int. J. Simul. Model., Doi:10.1080/02286203.2020.1792601, 2020.
- [18] S. Lisawadi, M. Kashif Ali Shah and S.E. Ahmed, *Model selection and post estimation based on a pretest for logistic regression models*, J. Stat. Comput. Simul. **86** (17), 3495-3511, 2016.
- [19] O. Reangsephet, S. Lisawadi and S.E. Ahmed, *Improving estimation of regression parameters in negative binomial regression model*, in: Proceedings of International Conference on Management Science and Engineering Management, Springer, 265-275, 2018.
- [20] O. Reangsephet, S. Lisawadi and S.E. Ahmed, *Adaptive estimation strategies in gamma regression model*, J. Stat. Theory Pract. **14** (1), 1-27, 2020.
- [21] S.E. Saffari and R. Adnan, *Parameter estimation on zero-inflated negative binomial regression with right truncated data*, Sains Malaysiana **41**, 1483-1487, 2012.
- [22] A.M.E. Saleh, *Theory of Preliminary Test and Stein-Type Estimation with Applications*, John Wiley and Sons, 2006.
- [23] M.L. Sheu, T.W. Hu, T.E. Keeler, M. Ong and H.Y. Sung, *The effect of a major cigarette price change on smoking behavior in California: a zero-inflated negative binomial model*, Health Econ. **13** (8), 781-791, 2004.
- [24] S. So, D.H. Lee and B.C. Jung, *An alternative bivariate zero-inflated negative binomial regression model using a copula*, Econom. Lett. **113** (2), 183-185, 2011.
- [25] C. Stein, *The admissibility of Hotelling's T^2 -test*, Ann. Math. Stat. **27** (3), 616-623, 1956.
- [26] J.R. Thompson, *Some shrinkage techniques for estimating the mean*, J. Amer. Statist. Assoc. **63** (321), 113-122, 1968.
- [27] P. Wang, *A bivariate zero-inflated negative binomial regression model for count data with excess zeros*, Econom. Lett. **78** (3), 373-378, 2003.
- [28] P. Wang and J.D. Alba, *A zero-inflated negative binomial regression model with hidden Markov chain*, Econom. Lett. **92** (2), 209-213, 2006.
- [29] B. Yuzbasi and Y. Asar, *Ridge type estimation in the zero-inflated negative binomial regression*, in Econometrics: Methods and Applications, 93-104, 2018.
- [30] Z. Zandi, H. Bevrani and R. Arabi Belaghi, *Using shrinkage strategies to estimate fixed effects in zero-inflated negative binomial mixed model*, Comm. Statist. Simulation Comput., Doi:10.1080/03610918.2021.1928704, 2021.

Appendix 1

Proof of Lemma 3.1. Based on the sequence of local alternatives in Equation (3.1), we show \mathbf{R}_n is asymptotically normally distributed with mean and variance as follows:

$$\begin{aligned}\mathbf{R}_n &= \sqrt{n}(\mathbf{H}\hat{\boldsymbol{\delta}} - \mathbf{h}) \\ &= \sqrt{n}(\mathbf{H}\hat{\boldsymbol{\delta}} - \mathbf{H}\boldsymbol{\delta} + \mathbf{H}\boldsymbol{\delta} - \mathbf{h}) \\ &= \mathbf{H}\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) + \sqrt{n}(\mathbf{H}\boldsymbol{\delta} - \mathbf{h}) \\ &= \mathbf{H}\mathbf{Z}_n + \boldsymbol{\omega},\end{aligned}$$

as $n \rightarrow \infty$, $\mathbf{Z}_n \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1})$. So, $\mathbf{R}_n \xrightarrow{D} \mathbf{R} \sim \mathcal{N}_k(E(\mathbf{R}), Var(\mathbf{R}))$, where

$$\begin{aligned}E(\mathbf{R}) &= \mathbf{H}E(\mathbf{Z}) + \boldsymbol{\omega} \\ &= \boldsymbol{\omega}, \\ Var(\mathbf{R}) &= \mathbf{H}Var(\mathbf{Z})\mathbf{H}' \\ &= \mathbf{H}\mathbf{I}^{-1}\mathbf{H}'.\end{aligned}$$

Then

$$\begin{aligned}\mathbf{V}_n &= \sqrt{n}(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}) \\ &= \sqrt{n}(\hat{\boldsymbol{\delta}} - \mathcal{J}(\mathbf{H}\hat{\boldsymbol{\delta}} - \mathbf{h}) - \boldsymbol{\delta}) \\ &= \sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - \mathcal{J}\sqrt{n}(\mathbf{H}\hat{\boldsymbol{\delta}} - \mathbf{h}) \\ &= \mathbf{Z}_n - \mathcal{J}\mathbf{R}_n \\ &= \mathbf{Z}_n - \mathcal{J}(\mathbf{H}\mathbf{Z}_n + \boldsymbol{\omega}) \\ &= (\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{Z}_n - \mathcal{J}\boldsymbol{\omega},\end{aligned}$$

where \mathbf{I}_k is the identity matrix of order $k \times k$. As $n \rightarrow \infty$, $\mathbf{Z}_n \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1})$. Thus, $\mathbf{V}_n \xrightarrow{D} \mathbf{V} \sim \mathcal{N}_k(E(\mathbf{V}), Var(\mathbf{V}))$ where

$$\begin{aligned}E(\mathbf{V}) &= E((\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{Z} - \mathcal{J}\boldsymbol{\omega}) \\ &= (\mathbf{I}_k - \mathcal{J}\mathbf{H})E(\mathbf{Z}) - \mathcal{J}\boldsymbol{\omega} \\ &= -\mathcal{J}\boldsymbol{\omega}, \\ Var(\mathbf{V}) &= Var((\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{Z} - \mathcal{J}\boldsymbol{\omega}) \\ &= Var((\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{Z}) \\ &= (\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{I}^{-1}(\mathbf{I}_k - \mathcal{J}\mathbf{H})' \\ &= \mathbf{I}^{-1} - \mathcal{J}\mathbf{H}\mathbf{I}^{-1}.\end{aligned}$$

Now, we write

$$\begin{aligned}\mathbf{B}_n &= \sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) \\ &= \sqrt{n}(\hat{\boldsymbol{\delta}} - [\hat{\boldsymbol{\delta}} - \mathcal{J}(\mathbf{H}\hat{\boldsymbol{\delta}} - \mathbf{h})]) \\ &= \mathcal{J}\mathbf{R}_n \\ &= \mathcal{J}(\mathbf{H}\mathbf{Z}_n + \boldsymbol{\omega}) \\ &= \mathcal{J}\mathbf{H}\mathbf{Z}_n + \mathcal{J}\boldsymbol{\omega},\end{aligned}$$

so as $n \rightarrow \infty$, $\mathbf{Z}_n \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1})$ and $\mathbf{B}_n \xrightarrow{D} \mathbf{B} \sim \mathcal{N}_k(E(\mathbf{B}), Var(\mathbf{B}))$, where

$$\begin{aligned}E(\mathbf{B}) &= E(\mathcal{J}\mathbf{H}\mathbf{Z} + \mathcal{J}\boldsymbol{\omega}) \\ &= \mathcal{J}\mathbf{H}E(\mathbf{Z}) + \mathcal{J}\boldsymbol{\omega} \\ &= \mathcal{J}\boldsymbol{\omega},\end{aligned}$$

$$\begin{aligned}
\text{Var}(\mathbf{B}) &= \text{Var}(\mathcal{J}\mathbf{H}\mathbf{Z} + \mathcal{J}\boldsymbol{\omega}) \\
&= (\mathcal{J}\mathbf{H}) \text{Var}(\mathbf{Z}) (\mathcal{J}\mathbf{H})' \\
&= \mathcal{J}\mathbf{H}\mathbf{I}^{-1} (\mathcal{J}\mathbf{H})' \\
&= \mathcal{J}\mathbf{H}\mathbf{I}^{-1}.
\end{aligned}$$

The joint distribution of \mathbf{Z}_n and \mathbf{B}_n is as follows:

$$\begin{pmatrix} \mathbf{Z}_n \\ \mathbf{B}_n \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k \mathbf{Z}_n + \mathbf{0} \\ \mathcal{J}\mathbf{H}\mathbf{Z}_n + \mathcal{J}\boldsymbol{\omega} \end{pmatrix} = \begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z}_n + \begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix},$$

so as $n \rightarrow \infty$, $\mathbf{Z}_n \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1})$ and $\begin{pmatrix} \mathbf{Z}_n \\ \mathbf{B}_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \mathbf{Z} \\ \mathbf{B} \end{pmatrix} \sim \mathcal{N}_{2k}(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1)$ where

$$\begin{aligned}
\boldsymbol{\mu}_1 &= E \left[\begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z} + \begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix} \right] \\
&= \begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} E(\mathbf{Z}) + \begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix}, \\
\boldsymbol{\sigma}_1 &= \text{Var} \left[\begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z} + \begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix} \right] \\
&= \text{Var} \left[\begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z} \right] \\
&= \begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} \text{Var}[\mathbf{Z}] \begin{pmatrix} \mathbf{I}_k & (\mathcal{J}\mathbf{H})' \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}_k \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{I}^{-1} \begin{pmatrix} \mathbf{I}_k & (\mathcal{J}\mathbf{H})' \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}^{-1} & \mathbf{I}^{-1} (\mathcal{J}\mathbf{H})' \\ \mathcal{J}\mathbf{H}\mathbf{I}^{-1} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} (\mathcal{J}\mathbf{H})' \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}^{-1} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} \\ \mathcal{J}\mathbf{H}\mathbf{I}^{-1} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} \end{pmatrix}.
\end{aligned}$$

In a similar way, we consider

$$\begin{aligned}
\begin{pmatrix} \mathbf{V}_n \\ \mathbf{B}_n \end{pmatrix} &= \begin{pmatrix} (\mathbf{I}_k - \mathcal{J}\mathbf{H}) \mathbf{Z}_n - \mathcal{J}\boldsymbol{\omega} \\ \mathcal{J}\mathbf{H}\mathbf{Z}_n + \mathcal{J}\boldsymbol{\omega} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{I}_k - \mathcal{J}\mathbf{H} \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z}_n + \begin{pmatrix} -\mathbf{I}_k \\ \mathbf{I}_k \end{pmatrix} \mathcal{J}\boldsymbol{\omega}.
\end{aligned}$$

Therefore, as $n \rightarrow \infty$, $\mathbf{Z}_n \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_k(\mathbf{0}, \mathbf{I}^{-1})$ and $\begin{pmatrix} \mathbf{V}_n \\ \mathbf{B}_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \mathbf{V} \\ \mathbf{B} \end{pmatrix} \sim \mathcal{N}_{2k}(\boldsymbol{\mu}_2, \boldsymbol{\sigma}_2)$ where

$$\begin{aligned}
\boldsymbol{\mu}_2 &= E \left[\begin{pmatrix} \mathbf{I}_k - \mathcal{J}\mathbf{H} \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z} + \begin{pmatrix} -\mathbf{I}_k \\ \mathbf{I}_k \end{pmatrix} \mathcal{J}\boldsymbol{\omega} \right] \\
&= \begin{pmatrix} \mathbf{I}_k - \mathcal{J}\mathbf{H} \\ \mathcal{J}\mathbf{H} \end{pmatrix} E(\mathbf{Z}) + \begin{pmatrix} -\mathbf{I}_k \\ \mathbf{I}_k \end{pmatrix} \mathcal{J}\boldsymbol{\omega} \\
&= \begin{pmatrix} -\mathcal{J}\boldsymbol{\omega} \\ \mathcal{J}\boldsymbol{\omega} \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
 \sigma_2 &= \text{Vae} \left[\begin{pmatrix} \mathbf{I}_k - \mathcal{J}\mathbf{H} \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{Z} + \begin{pmatrix} -\mathbf{I}_k \\ \mathbf{I}_k \end{pmatrix} \mathcal{J}\boldsymbol{\omega} \right] \\
 &= \begin{pmatrix} \mathbf{I}_k - \mathcal{J}\mathbf{H} \\ \mathcal{J}\mathbf{H} \end{pmatrix} \text{Var}(\mathbf{Z}) \begin{pmatrix} (\mathbf{I}_k - \mathcal{J}\mathbf{H})' & (\mathcal{J}\mathbf{H})' \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{I}_k - \mathcal{J}\mathbf{H} \\ \mathcal{J}\mathbf{H} \end{pmatrix} \mathbf{I}^{-1} \begin{pmatrix} (\mathbf{I}_k - \mathcal{J}\mathbf{H})' & (\mathcal{J}\mathbf{H})' \end{pmatrix} \\
 &= \begin{pmatrix} (\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{I}^{-1}(\mathbf{I}_k - \mathcal{J}\mathbf{H})' & (\mathbf{I}_k - \mathcal{J}\mathbf{H})\mathbf{I}^{-1}(\mathcal{J}\mathbf{H})' \\ \mathcal{J}\mathbf{H}\mathbf{I}^{-1}(\mathbf{I}_k - \mathcal{J}\mathbf{H})' & \mathcal{J}\mathbf{H}\mathbf{I}^{-1}(\mathcal{J}\mathbf{H})' \end{pmatrix} \\
 &= \begin{pmatrix} \mathbf{I}^{-1} - \mathcal{J}\mathbf{H}\mathbf{I}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathcal{J}\mathbf{H}\mathbf{I}^{-1} \end{pmatrix}.
 \end{aligned}$$

□

Appendix 2

Proof of Theorem 3.4. Here, we provide the proof of bias expressions. Based on Lemma 3.1 we have

$$\begin{aligned}
 ADB(\hat{\boldsymbol{\delta}}) &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})] = E(\mathbf{Z}) = \mathbf{0}, \\
 ADB(\tilde{\boldsymbol{\delta}}) &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta})] = E(\mathbf{V}) = -\mathcal{J}\boldsymbol{\omega}, \\
 ADB(\hat{\boldsymbol{\delta}}^{LS}) &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}}^{LS} - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\lambda \tilde{\boldsymbol{\delta}} + (1 - \lambda)\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - \lambda \sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})] \\
 &= E(\mathbf{Z}) - \lambda E(\mathbf{B}) \\
 &= -\lambda \mathcal{J}\boldsymbol{\omega}, \\
 ADB(\hat{\boldsymbol{\delta}}^{SP}) &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}}^{SP} - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}} - \lambda(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha}) - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) - \lambda \sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha})] \\
 &= E(\mathbf{Z}) - \lambda E(\mathbf{B}I(\chi_{p_2+2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)) \\
 &= -\lambda E(\mathbf{B}I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)),
 \end{aligned}$$

based on Equation (3.7), we can write

$$\begin{aligned}
 ADB(\hat{\boldsymbol{\delta}}^{SP}) &= -\lambda \mathcal{J}\boldsymbol{\omega} E[I(\chi_{p_2+2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)] \\
 &= -\lambda \mathcal{J}\boldsymbol{\omega} P(\chi_{p_2+2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2) \\
 &= -\lambda \mathcal{J}\boldsymbol{\omega} \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*).
 \end{aligned}$$

If $\lambda = 1$,

$$ADB(\hat{\boldsymbol{\delta}}^{PT}) = -\mathcal{J}\boldsymbol{\omega} \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*).$$

In a similar way, we can obtain

$$\begin{aligned}
 ADB(\hat{\boldsymbol{\delta}}^S) &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}}^S - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\tilde{\boldsymbol{\delta}} + (1 - (p_2 - 2) \mathbf{T}_n^{-1})(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} [E[\sqrt{n}(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}) + E[\sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})] - (p_2 - 2) E[\mathbf{T}_n^{-1} \sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})]] \\
 &= \lim_{n \rightarrow \infty} [E[\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})] - (p_2 - 2) E[\mathbf{T}_n^{-1} \sqrt{n}(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})]] \\
 &= E(\mathbf{Z}) - (p_2 - 2) E(\mathbf{T}_n^{-1} \mathbf{B}) \\
 &= -(p_2 - 2) \boldsymbol{\mathcal{J}} \boldsymbol{\omega} E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right], \\
 ADB(\hat{\boldsymbol{\delta}}^{S+}) &= \lim_{n \rightarrow \infty} E[\sqrt{n}(\hat{\boldsymbol{\delta}}^{S+} - \boldsymbol{\delta})] \\
 &= \lim_{n \rightarrow \infty} E[\sqrt{n}[\hat{\boldsymbol{\delta}}^S - \boldsymbol{\delta} - (1 - (p_2 - 2) \mathbf{T}_n^{-1})(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) I(\mathbf{T}_n < p_2 - 2)]] \\
 &= ADB(\hat{\boldsymbol{\delta}}^S) - \lim_{n \rightarrow \infty} E\sqrt{n}[(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}})(1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < p_2 - 2)] \\
 &= ADB(\hat{\boldsymbol{\delta}}^S) - E[\mathbf{B}(1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < p_2 - 2)] \\
 &= ADB(\hat{\boldsymbol{\delta}}^S) - E[\mathbf{B} I(\mathbf{T}_n < p_2 - 2)] + (p_2 - 2) E[\mathbf{B} \mathbf{T}_n^{-1} I(\mathbf{T}_n < p_2 - 2)] \\
 &= ADB(\hat{\boldsymbol{\delta}}^S) - \boldsymbol{\mathcal{J}} \boldsymbol{\omega} \boldsymbol{\Psi}_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) \\
 &\quad + (p_2 - 2) \boldsymbol{\mathcal{J}} \boldsymbol{\omega} E\left(\frac{I(\chi_{p_2+2}^2(\Delta^*) < p_2 - 2)}{\chi_{p_2+2}^2(\Delta^*)}\right).
 \end{aligned}$$

□

Appendix 3

Proof of Theorem 3.5. We first derive the asymptotic covariance of the estimators as defined in Equation (3.5)

$$\begin{aligned}
 \mathbf{V}(\hat{\boldsymbol{\delta}}) &= \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta}) \sqrt{n}(\hat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\right) \\
 &= \lim_{n \rightarrow \infty} E(\mathbf{Z}_n \mathbf{Z}_n') \\
 &= E(\mathbf{Z} \mathbf{Z}') \\
 &= Var(\mathbf{Z}) + E(\mathbf{Z}) E(\mathbf{Z}') \\
 &= \mathbf{I}^{-1}, \\
 \mathbf{V}(\tilde{\boldsymbol{\delta}}) &= \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta}) \sqrt{n}(\tilde{\boldsymbol{\delta}} - \boldsymbol{\delta})'\right) \\
 &= \lim_{n \rightarrow \infty} E(\mathbf{V}_n \mathbf{V}_n') \\
 &= E(\mathbf{V} \mathbf{V}'), \\
 &= Var(\mathbf{V}) + E(\mathbf{V}) E(\mathbf{V}') \\
 &= \mathbf{I}^{-1} - \boldsymbol{\mathcal{J}} \mathbf{H} \mathbf{I}^{-1} + (\boldsymbol{\mathcal{J}} \boldsymbol{\omega}) (\boldsymbol{\mathcal{J}} \boldsymbol{\omega})',
 \end{aligned}$$

$$\begin{aligned}
 \mathbf{V}(\hat{\boldsymbol{\delta}}^{LS}) &= \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}}^{LS} - \boldsymbol{\delta}) \sqrt{n}(\hat{\boldsymbol{\delta}}^{LS} - \boldsymbol{\delta})'\right) \\
 &= \lim_{n \rightarrow \infty} E[(\mathbf{Z}_n - \lambda \mathbf{B}_n)(\mathbf{Z}_n - \lambda \mathbf{B}_n)'] \\
 &= E[(\mathbf{Z} - \lambda \mathbf{B})(\mathbf{Z} - \lambda \mathbf{B})'] \\
 &= E[\mathbf{Z}\mathbf{Z}'] - 2\lambda E[\mathbf{Z}\mathbf{B}'] + \lambda^2 E[\mathbf{B}\mathbf{B}'] \\
 &= \mathbf{I}^{-1} - 2\lambda \underbrace{E[\mathbf{Z}\mathbf{B}']}_{e_1} + \lambda^2 [\boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1} + (\boldsymbol{\mathcal{J}}\boldsymbol{\omega})(\boldsymbol{\mathcal{J}}\boldsymbol{\omega})'].
 \end{aligned}$$

Using the conditional expectation of a multivariate normal distribution, e_1 becomes

$$\begin{aligned}
 e_1 &= E[\mathbf{Z}\mathbf{B}'] \\
 &= E(E[\mathbf{Z}\mathbf{B}' | \mathbf{B}]) \\
 &= E(\mathbf{B}' E[\mathbf{Z} | \mathbf{B}]) \\
 &= E(\{E[\mathbf{Z} + (\boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1})(\boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1})^{-1}(\mathbf{B} - \boldsymbol{\mathcal{J}}\boldsymbol{\omega})]\} \mathbf{B}') \\
 &= E((\mathbf{B} - \boldsymbol{\mathcal{J}}\boldsymbol{\omega}) \mathbf{B}') \\
 &= E(\mathbf{B}\mathbf{B}') - \boldsymbol{\mathcal{J}}\boldsymbol{\omega} E(\mathbf{B}') \\
 &= \boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1}
 \end{aligned}$$

Therefore,

$$\mathbf{V}(\hat{\boldsymbol{\delta}}^{LS}) = \mathbf{I}^{-1} - \lambda(2 - \lambda)\boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1} + \lambda^2(\boldsymbol{\mathcal{J}}\boldsymbol{\omega})(\boldsymbol{\mathcal{J}}\boldsymbol{\omega})'.$$

Next, we obtain $\mathbf{V}(\hat{\boldsymbol{\delta}}^{SP})$ as follows:

$$\begin{aligned}
 \mathbf{V}(\hat{\boldsymbol{\delta}}^{SP}) &= \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}}^{SP} - \boldsymbol{\delta}) \sqrt{n}(\hat{\boldsymbol{\delta}}^{SP} - \boldsymbol{\delta})'\right) \\
 &= \lim_{n \rightarrow \infty} E[\{\mathbf{Z}_n - \lambda \mathbf{B}_n I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha})\} \{\mathbf{Z}_n - \lambda \mathbf{B}_n I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha})\}'] \\
 &= \lim_{n \rightarrow \infty} E(\mathbf{Z}_n \mathbf{Z}_n') - 2\lambda \lim_{n \rightarrow \infty} E(\mathbf{Z}_n \mathbf{B}_n' I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha})) \\
 &\quad + \lambda^2 \lim_{n \rightarrow \infty} E(\mathbf{B}_n \mathbf{B}_n' I(\mathbf{T}_n \leq \mathbf{T}_{n,\alpha})) \\
 &= E(\mathbf{Z}\mathbf{Z}') - 2\lambda E(\mathbf{Z}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)) + \lambda^2 E(\mathbf{B}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)) \\
 &= \mathbf{I}^{-1} - 2\lambda \underbrace{E(\mathbf{Z}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2))}_{e_2} + \lambda^2 \underbrace{E(\mathbf{B}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2))}_{e_3},
 \end{aligned}$$

using Equation (3.8), we have

$$e_3 = \boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1} \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) + (\boldsymbol{\mathcal{J}}\boldsymbol{\omega})(\boldsymbol{\mathcal{J}}\boldsymbol{\omega})' \Psi_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^*),$$

and by using conditional expectation, e_2 becomes

$$\begin{aligned}
 e_2 &= E[\mathbf{Z}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)] \\
 &= E[E(\mathbf{Z}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2) | \mathbf{B})] \\
 &= E[E(\mathbf{Z} | \mathbf{B}) \mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)] \\
 &= E[\{E(\mathbf{Z}) + (\mathbf{B} - \boldsymbol{\mathcal{J}}\boldsymbol{\omega})\} \mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)] \\
 &= \underbrace{E[\mathbf{B}\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)]}_{e_3} - \boldsymbol{\mathcal{J}}\boldsymbol{\omega} E[\mathbf{B}' I(\chi_{p_2}^2(\Delta^*) \leq \chi_{p_2,\alpha}^2)] \\
 &= \boldsymbol{\mathcal{J}}\mathbf{H}\mathbf{I}^{-1} \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) - (\boldsymbol{\mathcal{J}}\boldsymbol{\omega})(\boldsymbol{\mathcal{J}}\boldsymbol{\omega})' [\Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) - \Psi_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^*)].
 \end{aligned}$$

Therefore,

$$\mathbf{V}(\hat{\boldsymbol{\delta}}^{SP}) = \mathbf{I}^{-1} - \lambda(2 - \lambda) \mathbf{J} \mathbf{H} \mathbf{I}^{-1} \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) + (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' [2\lambda \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) - \lambda(2 - \lambda) \Psi_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^*)].$$

For $\lambda = 1$, $\mathbf{V}(\hat{\boldsymbol{\delta}}^{PT})$ reduces to

$$\mathbf{V}(\hat{\boldsymbol{\delta}}^{PT}) = \mathbf{I}^{-1} - \mathbf{J} \mathbf{H} \mathbf{I}^{-1} \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) + (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' [2 \Psi_{p_2+2}(\chi_{p_2,\alpha}^2; \Delta^*) - \Psi_{p_2+4}(\chi_{p_2,\alpha}^2; \Delta^*)].$$

Now, we obtain $\mathbf{V}(\hat{\boldsymbol{\delta}}^S)$ as follows:

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\delta}}^S) &= \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}}^S - \boldsymbol{\delta}) \sqrt{n}(\hat{\boldsymbol{\delta}}^S - \boldsymbol{\delta})'\right) \\ &= \lim_{n \rightarrow \infty} E\left[\sqrt{n}(\tilde{\boldsymbol{\delta}} + (1 - (p_2 - 2)\mathbf{T}_n^{-1})(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) - \boldsymbol{\delta}) \right. \\ &\quad \left. \times \sqrt{n}(\tilde{\boldsymbol{\delta}} + (1 - (p_2 - 2)\mathbf{T}_n^{-1})(\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) - \boldsymbol{\delta})'\right] \\ &= \lim_{n \rightarrow \infty} E[(\mathbf{Z}_n - (p_2 - 2)\mathbf{T}_n^{-1}\mathbf{B}_n)(\mathbf{Z}_n - (p_2 - 2)\mathbf{T}_n^{-1}\mathbf{B}_n)'] \\ &= E[(\mathbf{Z} - (p_2 - 2)\mathbf{T}_n^{-1}\mathbf{B})(\mathbf{Z} - (p_2 - 2)\mathbf{T}_n^{-1}\mathbf{B})'] \\ &= E[(\mathbf{Z}\mathbf{Z}') - 2(p_2 - 2) \underbrace{E[\mathbf{B}\mathbf{Z}'\mathbf{T}_n^{-1}]}_{e_4} + (p_2 - 2)^2 \underbrace{E[\mathbf{B}\mathbf{B}'\mathbf{T}_n^{-2}]}_{e_5}], \end{aligned}$$

similar to e_1 , we can write e_4 as follows:

$$\begin{aligned} e_4 &= E[\mathbf{B}\mathbf{Z}'\mathbf{T}_n^{-1}] \\ &= E[\mathbf{B}\mathbf{B}'\mathbf{T}_n^{-1}] - \mathbf{J} \boldsymbol{\omega} E[\mathbf{B}\mathbf{T}_n^{-1}] \\ &= \mathbf{J} \mathbf{H} \mathbf{I}^{-1} E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] - (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' \left(E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] - E\left[\frac{1}{\chi_{p_2+4}^2(\Delta^*)}\right]\right), \end{aligned}$$

and by using Equation (3.8), e_5 becomes

$$\begin{aligned} e_5 &= E[\mathbf{B}\mathbf{B}'\mathbf{T}_n^{-2}] \\ &= \mathbf{J} \mathbf{H} \mathbf{I}^{-1} E\left[\frac{1}{(\chi_{p_2+2}^2(\Delta^*))^2}\right] + (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' E\left[\frac{1}{(\chi_{p_2+4}^2(\Delta^*))^2}\right]. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\delta}}^S) &= \mathbf{I}^{-1} - 2(p_2 - 2) \mathbf{J} \mathbf{H} \mathbf{I}^{-1} E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] \\ &\quad + 2(p_2 - 2) (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' \left\{E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] - E\left[\frac{1}{\chi_{p_2+4}^2(\Delta^*)}\right]\right\} \\ &\quad + (p_2 - 2)^2 \mathbf{J} \mathbf{H} \mathbf{I}^{-1} E\left[\frac{1}{(\chi_{p_2+2}^2(\Delta^*))^2}\right] + (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' E\left[\frac{1}{(\chi_{p_2+4}^2(\Delta^*))^2}\right] \\ &= \mathbf{I}^{-1} + (p_2 - 2) \mathbf{J} \mathbf{H} \mathbf{I}^{-1} \left\{(p_2 - 2) E\left[\frac{1}{(\chi_{p_2+2}^2(\Delta^*))^2}\right] - 2E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right]\right\} \\ &\quad + (p_2 - 2) (\mathbf{J} \boldsymbol{\omega}) (\mathbf{J} \boldsymbol{\omega})' \left\{-2E\left[\frac{1}{\chi_{p_2+4}^2(\Delta^*)}\right] + 2E\left[\frac{1}{\chi_{p_2+2}^2(\Delta^*)}\right] \right. \\ &\quad \left. + (p_2 - 2) E\left[\frac{1}{(\chi_{p_2+4}^2(\Delta^*))^2}\right]\right\}. \end{aligned}$$

Finally, we can write $\mathbf{V}(\hat{\boldsymbol{\delta}}^{S^+})$ as follows:

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\delta}}^{S^+}) &= \lim_{n \rightarrow \infty} E\left(\sqrt{n}(\hat{\boldsymbol{\delta}}^{S^+} - \boldsymbol{\delta}) \sqrt{n}(\hat{\boldsymbol{\delta}}^{S^+} - \boldsymbol{\delta})'\right) \\ &= \lim_{n \rightarrow \infty} E\left[\sqrt{n}\left(\hat{\boldsymbol{\delta}}^S - (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2)) (\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) - \boldsymbol{\delta}\right)\right. \\ &\quad \left. \times \sqrt{n}\left(\hat{\boldsymbol{\delta}}^S - (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2)) (\hat{\boldsymbol{\delta}} - \tilde{\boldsymbol{\delta}}) - \boldsymbol{\delta}\right)'\right] \\ &= \mathbf{V}(\hat{\boldsymbol{\delta}}^S) - 2E[\mathbf{B} \mathbf{V}' (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2))] \\ &\quad - 2E[\mathbf{B} \mathbf{B}' (1 - (p_2 - 2) \mathbf{T}_n^{-1})^2 I(\mathbf{T}_n < (p_2 - 2))] \\ &\quad + E[\mathbf{B} \mathbf{B}' (1 - (p_2 - 2) \mathbf{T}_n^{-1})^2 I(\mathbf{T}_n < (p_2 - 2))] \\ &= \mathbf{V}(\hat{\boldsymbol{\delta}}^S) - 2 \underbrace{E[\mathbf{B} \mathbf{V}' (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2))]}_{e_6} \\ &\quad - \underbrace{E[\mathbf{B} \mathbf{B}' (1 - (p_2 - 2) \mathbf{T}_n^{-1})^2 I(\mathbf{T}_n < (p_2 - 2))]}_{e_7}, \end{aligned}$$

now we obtain e_6

$$\begin{aligned} e_6 &= E[\mathbf{B} \mathbf{V}' (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2))] \\ &= E[\mathbf{B} E\{\mathbf{V}' (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2)) \mid \mathbf{B}\}] \\ &= E[\mathbf{B} E\{-\boldsymbol{\mathcal{J}} \boldsymbol{\omega} + \mathbf{0} \times (\boldsymbol{\mathcal{J}} \mathbf{H} \mathbf{I}^{-1})^{-1} (\mathbf{B} - \boldsymbol{\mathcal{J}} \boldsymbol{\omega})\}' \\ &\quad \times (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2))] \\ &= -E[\mathbf{B} \boldsymbol{\mathcal{J}} \boldsymbol{\omega} (1 - (p_2 - 2) \mathbf{T}_n^{-1}) I(\mathbf{T}_n < (p_2 - 2))] \\ &= -(\boldsymbol{\mathcal{J}} \boldsymbol{\omega}) (\boldsymbol{\mathcal{J}} \boldsymbol{\omega})' E\left[\left(1 - \frac{p_2 - 2}{\chi_{p_2+2}^2(\Delta^*)}\right) I(\chi_{p_2+2}^2(\Delta^*) < p_2 - 2)\right], \end{aligned}$$

and based on Equation (3.8), e_7 becomes

$$\begin{aligned} e_7 &= E[\mathbf{B} \mathbf{B}' (1 - (p_2 - 2) \mathbf{T}_n^{-1})^2 I(\mathbf{T}_n < (p_2 - 2))] \\ &= \boldsymbol{\mathcal{J}} \mathbf{H} \mathbf{I}^{-1} E\left[\left(1 - \frac{p_2 - 2}{\chi_{p_2+2}^2(\Delta^*)}\right)^2 I(\chi_{p_2+2}^2(\Delta^*) < p_2 - 2)\right] \\ &\quad + (\boldsymbol{\mathcal{J}} \boldsymbol{\omega}) (\boldsymbol{\mathcal{J}} \boldsymbol{\omega})' E\left[\left(1 - \frac{p_2 - 2}{\chi_{p_2+4}^2(\Delta^*)}\right)^2 I(\chi_{p_2+4}^2(\Delta^*) < p_2 - 2)\right]. \end{aligned}$$

Therefore, $\mathbf{V}(\hat{\boldsymbol{\delta}}^{S^+})$ becomes

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\delta}}^{S^+}) &= \mathbf{V}(\hat{\boldsymbol{\delta}}^S) + 2(\boldsymbol{\mathcal{J}} \boldsymbol{\omega}) (\boldsymbol{\mathcal{J}} \boldsymbol{\omega})' E\left[\left(1 - \frac{p_2 - 2}{\chi_{p_2+2}^2(\Delta^*)}\right) I(\chi_{p_2+2}^2(\Delta^*) < p_2 - 2)\right] \\ &\quad - (\boldsymbol{\mathcal{J}} \boldsymbol{\omega}) (\boldsymbol{\mathcal{J}} \boldsymbol{\omega})' E\left[\left(1 - \frac{p_2 - 2}{\chi_{p_2+4}^2(\Delta^*)}\right)^2 I(\chi_{p_2+4}^2(\Delta^*) < p_2 - 2)\right] \\ &\quad - \boldsymbol{\mathcal{J}} \mathbf{H} \mathbf{I}^{-1} E\left[\left(1 - \frac{p_2 - 2}{\chi_{p_2+2}^2(\Delta^*)}\right)^2 I(\chi_{p_2+2}^2(\Delta^*) < p_2 - 2)\right]. \end{aligned}$$

Now, the proof of Theorem 3.5 can be derived using the above results by following the definition of ADR. □