Journal of Data Science and Modeling, Vol. 1, No. 2, 99-124, June 2023

Research Manuscript

Estimation of Fixed Parameters in Negative Binomial Mixed Model Using Shrinkage Estimators

Zahra Zandi[1], Hossein Bevrani[2], Reza Arabi Belaghi[3] *

[1] Department of Statistics, University of Tabriz, Tabriz, Iran.

[2] Professor of Statistics, Department of Statistics, University of Tabriz, Tabriz, Iran.

[3] Assistant Professor, Department of Statistics, University of Tabriz, Tabriz, Iran.

Received: 29/10/2020 Accepted: 04/11/2021

Abstract: In this paper, we consider the problem of parameter estimation in negative binomial mixed model when it is suspected that some of the fixed parameters may be restricted to a subspace via linear shrinkage, preliminary test, shrinkage preliminary test, shrinkage, and positive shrinkage estimators along with the unrestricted maximum likelihood and restricted estimators. The random effects are considered as nuisance parameters. We conduct a Monte Carlo simulation study to evaluate the performance of each estimator in the sense of simulated relative efficiency. The results of the simulation study reveal that the proposed estimation strategies perform better than the maximum likelihood method. The proposed estimators are applied to a real dataset to appraise their performance.

Keywords: Longitudinal data; Monte Carlo simulation; Negative binomial mixed model; Over-dispersion; Shrinkage estimators

Mathematics Subject Classification (2010): 62J07, 62J05

*Corresponding author: rezaarabi11@gmail.com

1. Introduction

Longitudinal data arises when more than one response is measured on each subject in the study. These types of data are commonly used in many fields such as health research, economics, and biology. In this case, the outcomes are not independent, so the linear regression model cannot be used to analyze this data.

A generalized linear mixed model (GLMM) is an extension to the generalized linear model (GLM) in which the linear predictor contains random effects and the fixed effects, which is widely used to model correlated responses. The random effects are applied to capture the correlation within the observations. The GLMM is the conditional distribution of a response variable $\mathbf{y}$ given the $s \times 1$ vector of unobserved random effects $\mathbf{u}$ as $g(E[\mathbf{y} \mid \mathbf{u}]) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\,\mathbf{u}$. Here $\mathbf{X}$ and $\mathbf{Z}$ are $N \times p$ and $N \times s$ design matrices related to the fixed effect and the random effect parameters, respectively, and $g(t)$ is the link function. In Poisson and negative binomial model $g(t) = \ln(t)$. We refer to textbook of McCulloch et al. (2008) for more information about GLMMs. Poisson mixed model (PMM) is useful for analyzing correlated count data when the mean and variance of correlated responses are equal. But in practice, data are often over-dispersed, that is, the variance of data is greater than the average. In this case, parameter estimation can be seriously biased based on PMM. Hence, a negative binomial mixed model (NBMM) is appropriate for analyzing this data.

The main aim of this study is to estimate the fixed parameters in NBMM based on the linear shrinkage, preliminary test, shrinkage preliminary test, shrinkage, and positive shrinkage estimators under linear restriction on the fixed parameters when the random effects are considered as nuisance parameters. The linear restriction can be obtained from some uncertain prior information (UPI) or non-sampling information (NSI) about the parameters. Based on the information, some of the predictors may not have an influence on the interest response. Hence, we study two models: One is the unrestricted model that includes all p fixed parameters, and we estimate this model with an unrestricted estimator. The other model is the candidate restricted model where $\boldsymbol{\beta}$ satisfies the linear restriction

$$\mathbf{R}\,\boldsymbol{\beta} = \mathbf{r}$$

where $\mathbf{R}$ is a $q \times p$ matrix of rank $q \leq p$ and $\mathbf{r}$ is a $q \times 1$ vector of known constants, where q is the number of inactive fixed parameters in the unrestricted model. The shrinkage estimators are defined by using unrestricted and restricted estimators of parameters in the unrestricted and the restricted models, respectively. For detailed information about shrinkage strategies see Saleh (2006). Recently, Thomson andHossain (2018) proposed preliminary test, shrinkage, and positive shrinkage strategies to estimate the fixed parameters in a generalized linear mixed model. Hossain et al. (2018) proposed the shrinkage strategies in the linear mixed model. Also, Hossain et al. (2015) introduced the shrinkage and penalty estimators in a GLM. Many authors have applied shrinkage strategies in different regression models. Some of them are Roozbeh et al. (2020), Yuzbasi et al. (2020), Arashi and Roozbeh (2019), Saleh et al. (2019), Noori Asl et al. (2020), Zandi et al. (2021) and Hossain and Howlader (2015) among others.

In this paper, we develop the linear shrinkage, preliminary test, shrinkage preliminary test, shrinkage, and positive shrinkage estimation methods, and compare their performance with the maximum likelihood estimator for NBMM when some of the fixed covariates may be subject to a linear restriction.

The remainder of this article is organized as follows. The negative binomial mixed model and suggested estimators are introduced in Section 2. The asymptotic properties of the proposed estimators and their asymptotic distributional biases and risks are presented in Section 3. The results of a Monte Carlo simulation study are reported in Section 4. The proposed estimation methods are applied to the salamander's dataset in Section 5. Finally, conclusions are presented in Section 6.

2. Negative binomial mixed model

Suppose that we have a sample of N observations from n subjects. Let y_{ij} denote the response for the ith subject measured at the jth time, where $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, n_i$ and $N = \sum_{i=1}^{n} n_i$. Let $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ denotes the $n_i \times 1$ vector of response for the ith subject. Corresponding to each $\mathbf{y}_i$, let $\mathbf{X}_i = (\mathbf{x}_{i1}{}^T, \mathbf{x}_{i2}{}^T, \ldots, \mathbf{x}_{in_i}{}^T)^T$ and $\mathbf{Z}_i = (\mathbf{z}_{i1}{}^T, \mathbf{z}_{i2}{}^T, \ldots, \mathbf{z}_{in_i}{}^T)^T$ be the $n_i \times p$ and $n_i \times s$ design matrices related to the fixed effects and the random effects, respectively, where $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, \ldots, x_{ijp})^T$ and $\mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, \ldots, z_{ijs})^T$. Thus, the negative binomial mixed model on the ith subject is defined as

$$\ln(E(\mathbf{y}_i \mid \mathbf{u}_i)) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$ is the $p \times 1$ vector of unknown regression parameters for the fixed effects and $\mathbf{u}_i = (u_{i1}, u_{i2}, \ldots, u_{is})^T$ is the $s \times 1$ vector of random effects for the ith subject. Following Hossain et al. (2018) and Thomson andHossain (2018), we assume that $\mathbf{u}_i$ independently has a multivariate normal distribution $\mathcal{N}_s(\mathbf{0}, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = diag(\theta_1, \theta_2, \ldots, \theta_s)$ is the $s \times s$ variance-covariance matrix associated with the random effects. Suppose that conditional on $\mathbf{u}_i$, the elements of $\mathbf{y}_i$ are independent, and has a negative binomial distribution with the probability function as following:

$$f(\mathbf{y}_i \mid \mathbf{u}_i, \boldsymbol{\mu}_i) = \frac{\Gamma(\mathbf{y}_i + \frac{1}{\nu})}{\Gamma(\frac{1}{\nu})\,\Gamma(\mathbf{y}_i + 1)} \left(\frac{\nu\,\boldsymbol{\mu}_i}{1 + \nu\,\boldsymbol{\mu}_i}\right)^{\mathbf{y}_i} \left(\frac{1}{1 + \nu\,\boldsymbol{\mu}_i}\right)^{\frac{1}{\nu}}, \tag{2.2}$$

where $\Gamma(.)$ is the gamma function, $\nu > 0$ is the model heterogeneity or over-dispersion parameter, and $\boldsymbol{\mu}_i = E(\mathbf{y}_i \mid \mathbf{u}_i) = exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i)$ is the mean parameter.

The likelihood function of the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$ given the vector of responses $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)^T$ is

$$\begin{aligned} \mathbf{L}((\boldsymbol{\beta}, \boldsymbol{\theta}) \mid \mathbf{y}) &= \prod_{i=1}^{n} \int \prod_{j=1}^{n_i} f_{y_{ij}\mid\mathbf{u}_i}(y_{ij} \mid \mathbf{u}_i, \boldsymbol{\beta}) f_{\mathbf{u}_i}(\mathbf{u}_i \mid \boldsymbol{\theta})\, \mathbf{du}_i \\ &= \prod_{i=1}^{n} \int f_{\mathbf{y}_i\mid\mathbf{u}_i}(\mathbf{y}_i \mid \mathbf{u}_i, \boldsymbol{\beta}) f_{\mathbf{u}_i}(\mathbf{u}_i)\, \mathbf{du}_i. \end{aligned} \tag{2.3}$$

To obtain the unrestricted score equations, we use the log-likelihood function as

$$\begin{aligned} \mathcal{L}^*((\boldsymbol{\beta}, \boldsymbol{\theta}) \mid \mathbf{y}) &= \ln[\mathbf{L}((\boldsymbol{\beta}, \boldsymbol{\theta}) \mid \mathbf{y})] \\ &= \sum_{i=1}^{n} \ln\Big(\int f_{\mathbf{y}_i\mid\mathbf{u}_i}(\mathbf{y}_i \mid \mathbf{u}_i, \boldsymbol{\beta}) f_{\mathbf{u}_i}(\mathbf{u}_i \mid \boldsymbol{\theta})\, \mathbf{du}_i\Big) \end{aligned} \tag{2.4}$$

Based on Hossain et al. (2018) and Thomson andHossain (2018), we consider the random effects as nuisance parameters, and we assume that $\boldsymbol{\theta}$ is known, so that the only parameters that we estimate are the fixed effects, $\boldsymbol{\beta}$. Hence, the log-likelihood function becomes a function of $\boldsymbol{\beta}$ only, and the estimation procedure solves the corresponding score equation to estimate $\boldsymbol{\beta}$. We solve the corresponding score equation to obtain the unrestricted maximum likelihood estimator of $\boldsymbol{\beta}$ and is given by

$$\begin{aligned} \frac{\partial\,\mathcal{L}^*(\boldsymbol{\beta} \mid \mathbf{y})}{\partial\,\boldsymbol{\beta}} &= \sum_{i=1}^{n} E\Big(\frac{\partial\,\ln(f_{\mathbf{y}_i\mid\mathbf{u}_i}(\mathbf{y}_i \mid \mathbf{u}_i, \boldsymbol{\beta}))}{\partial\,\beta_t} \mid \mathbf{y}_i\Big) \\ &= \sum_{i=1}^{n} E\Bigg(\Big\{\mathbf{y}_i\,\mathbf{x}_{it} - (\mathbf{y}_i + \frac{1}{\nu})\Big(\frac{exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i)}{1 + \nu\,exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i)}\Big)\Big\} \mid \mathbf{y}_i\Bigg) \\ &= \mathbf{0}. \end{aligned} \tag{2.5}$$

Equation (2.5) has not closed form of $\boldsymbol{\beta}$. Therefore, we solve it via numerical optimization to obtain the maximum likelihood estimator (see McCulloch et al. (2008), chap. 14). We denote the unrestricted estimator (UE) of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}^{UE}$.

We assumed that $\boldsymbol{\theta}$ is known, so the log-likelihood function of the model (2.2) can be defined by

$$\mathcal{L}(\boldsymbol{\beta} \mid \mathbf{y}) = \sum_{i=1}^{n} \Big\{ y_i \ln(\nu) \;+\; \mathbf{y}_i \ln(\boldsymbol{\mu}_i) \;-\; (\mathbf{y}_i + \frac{1}{\nu}) \ln\big(1 + \nu\, \boldsymbol{\mu}_i\big) \\ \;+\; \ln\Big(\Gamma(\mathbf{y}_i + \frac{1}{\nu})\Big) \;-\; \ln\Big(\Gamma(\mathbf{y}_i + 1)\Big) \;-\; \ln(\Gamma(\frac{1}{\nu}))\Big\}, \quad (2.6)$$

We now obtain the observed Fisher information matrix as derived by McCulloch et al. (2008) as follows:

$$\mathbf{I} = \mathbf{I}(\boldsymbol{\beta}, \boldsymbol{\beta}) = -\frac{\partial^2 \mathcal{L}(\boldsymbol{\beta} \mid \mathbf{y})}{\partial \beta_t \, \partial \beta_k} \\ = \sum_{i=1}^{n} \Big(\nu\, (\mathbf{y}_i + \frac{1}{\nu})\, \mathbf{x}_{it}\, \mathbf{x}_{ik} \Big\{ \frac{exp(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i)}{[1 + \nu\, exp(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i)]^2} \Big\}\Big). \quad (2.7)$$

To obtain the restricted estimator (RE) of $\boldsymbol{\beta}$, we consider testing the general linear hypothesis

$$H_0 \,:\, \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad versus \quad H_1 \,:\, \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}. \quad (2.8)$$

The restricted estimator of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^{RE}$ can be obtained by maximizing the log-likelihood function (2.4) under the hypothesis H_0 in Equation (2.8). Based on Hossain et al. (2015), the restricted estimator has the following form:

$$\hat{\boldsymbol{\beta}}^{RE} = \hat{\boldsymbol{\beta}}^{UE} - \mathbf{I}^{-1}\mathbf{R}^T(\mathbf{R}\,\mathbf{I}^{-1}\mathbf{R}^T)^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}^{UE} - \mathbf{r}). \quad (2.9)$$

2.1 Likelihood ratio test

Based on the estimators $\hat{\boldsymbol{\beta}}^{UE}$ and $\hat{\boldsymbol{\beta}}^{RE}$, we define the likelihood ratio test statistic for testing $H_0 \,:\, \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ versus $H_1 \,:\, \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$ as:

$$\mathbf{D}_n = 2\,\{\mathcal{L}(\hat{\boldsymbol{\beta}}^{UE} \mid \mathbf{y}) - \mathcal{L}(\hat{\boldsymbol{\beta}}^{RE} \mid \mathbf{y})\}. \quad (2.10)$$

where $\mathcal{L}(\hat{\boldsymbol{\beta}}^{UE} \mid \mathbf{y})$ and $\mathcal{L}(\hat{\boldsymbol{\beta}}^{RE} \mid \mathbf{y})$ are values of the log-likelihood function (2.6) at the unrestricted and restricted estimators, respectively. Under the null hypothesis H_0 in Equation (2.8), the test statistic $\mathbf{D}_n$ asymptotically follows a χ^2-distribution with q degrees of freedom.

2.2 Improved estimators

We propose the improved estimators including linear shrinkage, preliminary test, shrinkage preliminary test, shrinkage, and positive shrinkage estimators in the following sub-sections.

2.2.1 The linear shrinkage estimator

The linear shrinkage (LS) estimator of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^{LS}$ as the linear combination of the unrestricted and restricted estimator

$$\hat{\boldsymbol{\beta}}^{LS} = \delta\,\hat{\boldsymbol{\beta}}^{RE} + (1-\delta)\,\hat{\boldsymbol{\beta}}^{UE}, \tag{2.11}$$

where δ denotes the shrinkage intensity or the level of confidence in the prior information. The optimal value of $0 \leq \delta \leq 1$ is obtained by minimizing the mean squared error of the LS estimator. When $\delta > 0$, the performance of the linear shrinkage estimator is better than the unrestricted estimator.

2.2.2 The preliminary test and shrinkage preliminary test estimators

The preliminary test estimator (PT) of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^{PT}$ is defined as:

$$\hat{\boldsymbol{\beta}}^{PT} = \hat{\boldsymbol{\beta}}^{UE} - (\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\,I(\mathbf{D}_n \leq \mathbf{D}_{n,\alpha}), \tag{2.12}$$

where $I(.)$ is an indicator function, $\mathbf{D}_{n,\alpha}$ is the α-level upper critical value of the test statistic $\mathbf{D}_n$ in Equation (2.10).
The shrinkage preliminary test estimator (SP) of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^{SP}$ can be defined by replacing the restricted estimator with the linear shrinkage estimator in Equation (2.12). Ahmed (1992) proposed this estimator as:

$$\hat{\boldsymbol{\beta}}^{SP} = \hat{\boldsymbol{\beta}}^{UE} - \delta\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\,I(\mathbf{D}_n \leq \mathbf{D}_{n,\alpha}). \tag{2.13}$$

2.2.3 The shrinkage and positive shrinkage estimators

The shrinkage estimator (S), $\hat{\boldsymbol{\beta}}^{S}$ of $\boldsymbol{\beta}$ combines the unrestricted and restricted estimators in an optimal way dominating the unrestricted estimator as follows:

$$\hat{\boldsymbol{\beta}}^{S} = \hat{\boldsymbol{\beta}}^{UE} - \Big(\frac{q-2}{\mathbf{D}_n}\Big)(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}), \qquad q \geq 3. \tag{2.14}$$

and positive shrinkage estimator of $\boldsymbol{\beta}$ denoted by $\hat{\boldsymbol{\beta}}^{S^+}$ has the following form:

$$\hat{\boldsymbol{\beta}}^{S^+} = \hat{\boldsymbol{\beta}}^{UE} - \Big(\frac{q-2}{\mathbf{D}_n}\Big)^{+}(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}), \qquad q \geq 3, \tag{2.15}$$

where $z^+ = \max(0, z)$. The positive shrinkage estimator adjustment controls for the over-shrinking problem in $\hat{\boldsymbol{\beta}}^{S}$.

3. Asymptotic results

In this section, we study the asymptotic properties, containing asymptotic distributional biases and risks of proposed estimators in previous section for NBMM. When the linear restriction $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ is wrong, we investigate the asymptotic properties of the estimators under the sequence of local alternatives defined by:

$$\mathcal{K}_{(n)} \; : \; \mathbf{R}\boldsymbol{\beta} = \mathbf{r} + \frac{\boldsymbol{\xi}}{\sqrt{n}}, \quad n = 1, 2, \ldots \tag{3.16}$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, ..., \xi_q)^T \in R^q$ is a $q \times 1$ fixed vector of real numbers. It is obvious that for $\boldsymbol{\xi} = \mathbf{0}$, we have $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ for all n, which is a special case of (3.16). To explore the properties of the estimators, we mention the regularity conditions as follows:

- The parameter space for $\boldsymbol{\beta}$ is compact. The score function (2.5) is continuous function of $\boldsymbol{\beta}$ for all $\mathbf{y}$ and measurable functions of $\mathbf{y}$.
- There exists unique MLE of $\boldsymbol{\beta}$ for $\mathcal{L}^*(\boldsymbol{\beta} \mid \mathbf{y})$. The moments of $\frac{\partial \mathcal{L}^*(\boldsymbol{\beta}|\mathbf{y})}{\partial \boldsymbol{\beta}}$ exist at least up to the third order.
- The design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ in negative binomial mixed model $\ln(E(\mathbf{y}_i \mid \mathbf{u}_i)) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i$ are of full rank and all of their elements are bounded by a single finite real number.

We now present the following lemma, which is helpful for deriving the asymptotic distributional bias (ADB) and the asymptotic distributional risk (ADR) of various estimators.

Lemma 3.1. *Under the above regularity conditions and the sequence of local alternatives in Equation (3.16), as $n \to \infty$*

$$\begin{aligned}
\mathbf{Z}_n &= \sqrt{n}(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}) \xrightarrow{D} \mathbf{Z} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{B}^{-1}),\\
\mathbf{K}_n &= \sqrt{n}(\mathbf{R}\hat{\boldsymbol{\beta}}^{RE} - \mathbf{r}) \xrightarrow{D} \mathbf{K} \sim \mathcal{N}_p(\boldsymbol{\xi}, \mathbf{R}\,\mathbf{B}^{-1}\mathbf{R}^T),\\
\mathbf{V}_n &= \sqrt{n}(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta}) \xrightarrow{D} \mathbf{V} \sim \mathcal{N}_p(-\mathcal{J}\boldsymbol{\xi}, \mathbf{B}^{-1} - \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1}),\\
\mathbf{W}_n &= \sqrt{n}(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) \xrightarrow{D} \mathbf{W} \sim \mathcal{N}_p(\mathcal{J}\boldsymbol{\xi}, \mathcal{J}\mathbf{R}\mathbf{B}^{-1}),
\end{aligned}$$

$$\begin{pmatrix} \mathbf{Z}_n \\ \mathbf{W}_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \mathbf{Z} \\ \mathbf{W} \end{pmatrix} \sim \mathcal{N}_{2p}\left[\begin{pmatrix} \mathbf{0} \\ \mathcal{J}\boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \mathbf{B}^{-1} & \mathcal{J}\mathbf{R}\mathbf{B}^{-1} \\ \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1} & \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1} \end{pmatrix} \right],$$

$$\begin{pmatrix} \mathbf{V}_n \\ \mathbf{W}_n \end{pmatrix} \xrightarrow{D} \begin{pmatrix} \mathbf{V} \\ \mathbf{W} \end{pmatrix} \sim \mathcal{N}_{2p}\left[\begin{pmatrix} -\mathcal{J}\boldsymbol{\xi} \\ \mathcal{J}\boldsymbol{\xi} \end{pmatrix}, \begin{pmatrix} \mathbf{B}^{-1} - \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1} \end{pmatrix} \right],$$

where $\boldsymbol{\mathcal{J}} = \mathbf{B}^{-1}\boldsymbol{R}^T(\boldsymbol{R}\mathbf{B}^{-1}\boldsymbol{R}^T)^{-1}$, $\mathbf{B} = \lim_{n\to\infty}\mathbf{I}(\boldsymbol{\beta},\boldsymbol{\beta})/n$ *converges in probability to a non-random* $p\times p$ *positive definite matrix, and* $\mathbf{I}(\boldsymbol{\beta},\boldsymbol{\beta})$ *is the observed Fisher information matrix in NBMM given in Equation (2.7).*

Using the above Lemma, we can present ADB and ADR results. In order to obtain the asymptotic risks (ADR) of the estimators, we define the following weighted quadratic loss function:

$$L(\hat{\boldsymbol{\beta}}^*,\boldsymbol{\beta};\mathbf{Q}) = \left(\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})\right)^T \mathbf{Q}\left(\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})\right), \tag{3.17}$$

where $\hat{\boldsymbol{\beta}}^*$ is any of the proposed estimators in the previous section and $\mathbf{Q}$ is a positive semi-definite weight matrix. When $\mathbf{Q}$ is chosen as the identity matrix $\mathbf{I}$, the usual quadratic loss function is defined, which we use in our simulation studies in the next section. The asymptotic distributional bias (ADB) of an estimator $\hat{\boldsymbol{\beta}}^*$ is defined as

$$ADB(\hat{\boldsymbol{\beta}}^*) = \lim_{n\to\infty} E\left(\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})\right) = \int \dots \int \mathbf{x}\,\mathbf{d}G(\mathbf{x}), \tag{3.18}$$

where G is the cumulative distribution function of $\hat{\boldsymbol{\beta}}^*$. In order to simplify the notations to describe ADB and ADR results, let T_1 and T_2 be $\chi^2_{q+2}(\Delta^*)$ and $\chi^2_{q+4}(\Delta^*)$ random variables, respectively. The distribution function of a non-central χ^2 variable with d degrees of freedom and the non-centrality parameter Δ^* is denoted by $H_d(x,\Delta^*) = P(\chi^2_d(\Delta^*) \le x)$. Also, let $\chi^2_{q,\alpha}$ be the α-level critical value of the central χ^2 distribution. Using Lemma 3.1, we present the asymptotic biases of the estimators in the following Theorem.

Theorem 3.2. *Under the sequence of local alternatives in Equation (3.16) and the regularity conditions, the ADBs of the estimators are*

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{UE}) &= \boldsymbol{0},\\
ADB(\hat{\boldsymbol{\beta}}^{RE}) &= -\boldsymbol{\mathcal{J}}\boldsymbol{\xi},\\
ADB(\hat{\boldsymbol{\beta}}^{LS}) &= -\delta\,\boldsymbol{\mathcal{J}}\boldsymbol{\xi},\\
ADB(\hat{\boldsymbol{\beta}}^{PT}) &= -\boldsymbol{\mathcal{J}}\boldsymbol{\xi}\boldsymbol{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*),\\
ADB(\hat{\boldsymbol{\beta}}^{SP}) &= -\delta\,\boldsymbol{\mathcal{J}}\boldsymbol{\xi}\boldsymbol{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*),\\
ADB(\hat{\boldsymbol{\beta}}^{S}) &= -(q-2)\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E\Big(\frac{1}{T_1}\Big),\\
ADB(\hat{\boldsymbol{\beta}}^{S+}) &= ADB(\hat{\boldsymbol{\beta}}^{S}) - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*)\\
&\qquad + (q-2)\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E\left(\frac{I(T_1 < q-2)}{T_1}\right)
\end{aligned}$$

where $\Delta^* = \boldsymbol{\xi}^T(\boldsymbol{R}\,\boldsymbol{B}^{-1}\boldsymbol{R}^T)^{-1}\boldsymbol{\xi}$.

Proof: for detailed proof see Appendix 1.

Remark 3.3. *To compare the ADB of the estimators, let* $\boldsymbol{\psi} = -\boldsymbol{\mathcal{J}}\boldsymbol{\xi}/\sqrt{\Delta^*}$ *where* $\Delta^* = \boldsymbol{\xi}^T(\boldsymbol{R}\,\boldsymbol{B}^{-1}\boldsymbol{R}^T)^{-1}\boldsymbol{\xi}$. *Based on Theorem 3.2, the ADB of all estimators is a scalar function of* Δ^* *with the vector* $\boldsymbol{\psi}$. *The scale factor* $\sqrt{\Delta^*}$ *of the ADB of* $\hat{\boldsymbol{\beta}}^{RE}$ *is an unbounded function of* Δ^*, *but the scale factors in the ADBs of other estimators are bounded in* Δ^*. *Since* $E(\frac{1}{T_1})$ *is a decreasing function of* Δ^*, *the ADB of the shrinkage and positive shrinkage estimators starts from 0 at* $\Delta^* = 0$, *increases to a maximum, and then decreases towards 0, as* Δ^* *increases.*

Now, we define the asymptotic distributional risk of $\hat{\boldsymbol{\beta}}^*$ by

$$ADR(\hat{\boldsymbol{\beta}}^*; \mathbf{Q}) = \int \ldots \int \mathbf{x}^T\mathbf{Q}\,\mathbf{x}\,\mathbf{d}G(\mathbf{x}) = trace(\mathbf{Q\Sigma}), \tag{3.19}$$

where $\boldsymbol{\Sigma}$ is the dispersion matrix for the distribution $G(\mathbf{x})$ and defined as:

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^*) = \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})^T\Big) = \int \ldots \int \mathbf{x}^T\mathbf{x}\,\mathbf{d}G(\mathbf{x}). \tag{3.20}$$

Next, we present the asymptotic distributional risks (ADR) of the estimators in the following Theorem.

Theorem 3.4. *Under the sequence of local alternatives in Equation (3.16) and the usual regularity conditions, the ADRs of the estimators are*

$$\begin{aligned}
ADR(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{Q}) &= trace(\boldsymbol{Q}\mathbf{B}^{-1}),\\
ADR(\hat{\boldsymbol{\beta}}^{RE}; \boldsymbol{Q}) &= ADR(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{Q}) - trace(\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{R}\mathbf{B}^{-1}) + \boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi},\\
ADR(\hat{\boldsymbol{\beta}}^{LS}; \boldsymbol{Q}) &= ADR(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{Q}) - \delta\,(2-\delta)\,trace(\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{R}\mathbf{B}^{-1}) + \delta^2\,\boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi},\\
ADR(\hat{\boldsymbol{\beta}}^{PT}; \boldsymbol{Q}) &= ADR(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{Q}) - trace(\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{R}\mathbf{B}^{-1})\boldsymbol{H}_{q+2}(\chi^2_{q,\alpha}; \Delta^*)\\
&\quad + [2\boldsymbol{H}_{q+2}(\chi^2_{q,\alpha}; \Delta^*) - \boldsymbol{H}_{q+4}(\chi^2_{q,\alpha}; \Delta^*)]\,\boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi},\\
ADR(\hat{\boldsymbol{\beta}}^{SP}; \boldsymbol{Q}) &= ADR(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{Q}) - \delta\,(2-\delta)\,trace(\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{R}\mathbf{B}^{-1})\boldsymbol{H}_{q+2}(\chi^2_{q,\alpha}; \Delta^*)\\
&\quad + [2\,\delta\boldsymbol{H}_{q+2}(\chi^2_{q,\alpha}; \Delta^*) - \delta\,(2-\delta)\boldsymbol{H}_{q+4}(\chi^2_{q,\alpha}; \Delta^*)]\,\boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi},\\
ADR(\hat{\boldsymbol{\beta}}^{S}; \boldsymbol{Q}) &= ADR(\hat{\boldsymbol{\beta}}^{UE}; \boldsymbol{Q}) - (q-2)\,trace(\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{R}\mathbf{B}^{-1})\Big\{2\,E\Big[\frac{1}{T_1}\Big] - (q-2)\,E\Big[\frac{1}{T_1^2}\Big]\Big\}\\
&\quad + (q-2)\Big\{2\,E\Big[\frac{1}{T_1} - \frac{1}{T_2}\Big] + (q-2)\,E\Big[\frac{1}{T_2^2}\Big]\Big\}\,\boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi},\\
ADR(\hat{\boldsymbol{\beta}}^{S+}; \boldsymbol{Q}) &= ADR(\hat{\boldsymbol{\beta}}^{S}; \boldsymbol{Q}) - trace(\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{R}\,\boldsymbol{B}^{-1})\,E\Big((1-\frac{q-2}{T_1})^2\,I(T_1 < q-2)\Big)\\
&\quad - E\Big((1-\frac{q-2}{T_2})^2\,I(T_2 < q-2)\Big)\,\boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\\
&\quad + 2\,E\Big((1-\frac{q-2}{T_1})\,I(T_1 < q-2)\Big)\,\boldsymbol{\xi}^T\,\boldsymbol{\mathcal{J}}^T\,\boldsymbol{Q}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}.
\end{aligned}$$

Proof: for detailed proof see Appendix 2.

Remark 3.5. *To compare the asymptotic distributional risks of the estimators, when* $\boldsymbol{\xi} = \boldsymbol{0}$*, that is* $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$*, the restricted estimator dominates the unrestricted estimator. However, when* $\boldsymbol{\xi}$ *moves away from* $\boldsymbol{0}$ *vector, the ADR of the restricted estimator becomes unbounded. When the linear restriction is correct, the preliminary test estimator has a smaller risk than that of the shrinkage estimator. As* $\boldsymbol{\xi}$ *moves away from* $\boldsymbol{0}$*, the shrinkage preliminary test estimator dominates the preliminary test estimator. By comparing the risks of the shrinkage, positive shrinkage, and unrestricted estimators, we can see* $ADR(\hat{\boldsymbol{\beta}}^{S^+}) \leq ADR(\hat{\boldsymbol{\beta}}^{S}) \leq ADR(\hat{\boldsymbol{\beta}}^{UE})$ *for* $\boldsymbol{\xi} \geq \boldsymbol{0}$.

4. Simulation study

In this section, we present the details of the performance of the proposed estimators with respect to the unrestricted estimator using Monte Carlo simulation with statistical software R. We have used the **glmmTMB** package to obtain an unrestricted estimation of the fixed effect coefficients. In our study, the criterion for comparing the performance of any estimator is the simulated relative efficiency (SRE). We consider the negative binomial mixed model (NBMM) with $n = 60$ subjects and $n_i = 7$ observations for all i. We generate the count responses from NBMM as:

$$\boldsymbol{\mu_{ij}} = exp(\mathbf{x}_{ij}{}^T\boldsymbol{\beta} + \mathbf{u}_i).$$

We consider generating $p = 7, 13, 17, 20$. Each of the p fixed effect covariates $\mathbf{x}_{ij} = (\mathbf{x}_{ij1}, \mathbf{x}_{ij2}, \ldots, \mathbf{x}_{ijp})^T$ are generated from a separate n_i-multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\sigma^2\mathbf{I}$, where $\sigma^2 = 0.02$ and $\mathbf{I}$ is a $n_i \times n_i$ identity matrix. The random effects $\mathbf{u}_i$ are generated from a n_i-multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = diag(0.2,\ 0.5,\ 0.8,\ 0.3,\ 0.1,\ 0.4,\ 0.3)$ is a $n_i \times n_i$ diagonal matrix.
We are interested in testing the hypothesis $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where $\mathbf{R}$ is a $q \times p$ matrix of full row rank, and $\mathbf{r}$ is a $q \times 1$ vector of known constants. We consider a special case of this hypothesis with $\mathbf{R} = [\mathbf{0}_{q\times(p-q)}, \mathbf{I}_q]$ and $\mathbf{r} = \mathbf{0}_{q\times 1}$, where $\mathbf{0}_{a\times b}$ is a $a \times b$ matrix (vector) of zeros. Now, we partition the fixed effect coefficients vector of $\boldsymbol{\beta}$ as $\boldsymbol{\beta} = (\boldsymbol{\beta}^T{}_1, \boldsymbol{\beta}^T{}_2)^T$ where, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are $p_1 \times 1$ and $q \times 1$ vectors contain active and inactive fixed parameters, respectively, such that $p = p_1 + q$. Therefore, p_1 and q are the number of active and inactive fixed parameters, respectively. We set $p_1 = 4$ throughout the study and $q = 3,\ 9,\ 13,\ 16$ so, $p = 7,\ 13,\ 17,\ 20$.

We next define the distance between the simulation model and the restricted model as $\Delta = ||\boldsymbol{\beta} - \boldsymbol{\beta}^{true}|| = \sum_{i=1}^{p}(\beta_i - \beta_i^{true})^2$, where $||.||$ is the Euclidean norm, $\boldsymbol{\beta}$ is the parameter vector in the simulated model, and $\boldsymbol{\beta}^{true}$ is the true parameter. We set $\Delta = \{0.0, 0.5, 1.0, 1.5, 2.0\}$ in our simulations. Samples are generated using the over-dispersion parameter $\nu = 1.25$ and the true parameter $\boldsymbol{\beta}^{true} = (\boldsymbol{\beta}_1^{true}, \boldsymbol{\beta}_2^{true})^T$ as $\boldsymbol{\beta}_1^{true} = (2.1, 1.5, -1.2, -1.3)^T$ and $\boldsymbol{\beta}_2^{true} = \mathbf{0}_q{}^T$. Based on the definition of Δ, when candidate restricted model is correct ($\Delta = 0$), we have $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{true}$ and $\boldsymbol{\beta}_2 = \boldsymbol{\beta}_2^{true}$. When $\Delta > 0$, we have $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_1^{true}$ and $\boldsymbol{\beta}_2 = (a, \underbrace{0, 0, ..., 0}_{q-1})^T$, where a is a scalar, so that $a = \sqrt{\Delta}$. Hence, $a = \{0.0, 0.70, 1.0, 1.22, 1.41\}$. The number of replications is set to $1,000$ for all cases.

We define the simulated mean squared errors (SMSE) of the improved estimators in section 2 as follows:

$$SMSE(\hat{\boldsymbol{\beta}}^*) = \sum_{i=1}^{p}(\beta_i^{true} - \hat{\beta}_i^*)^2,$$

where β_i^{true} is ith element of the true parameter $\boldsymbol{\beta}^{true}$ and $\hat{\boldsymbol{\beta}}^*$ is one of the proposed estimators defined in the previous section. The criterion for comparing the performance of any estimator $\hat{\boldsymbol{\beta}}^*$ in our simulation is the simulated relative efficiency (SRE) and is defined as:

$$SRE(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\beta}}^{UE}) = \frac{SMSE(\hat{\boldsymbol{\beta}}^{UE})}{SMSE(\hat{\boldsymbol{\beta}}^*)}.$$

A value of SRE greater than one indicates that $\hat{\boldsymbol{\beta}}^*$ performs better than $\hat{\boldsymbol{\beta}}^{UE}$.

4.1 The results of simulation

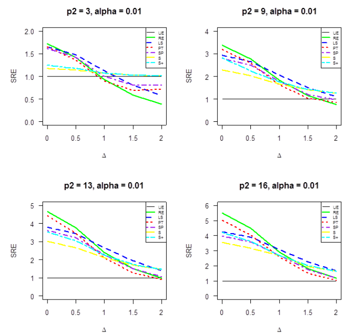
The simulated relative efficiencies (SREs) for all the proposed estimators are reported in Tables 1, 2 and Figures 1a, 1b, and 1c for the shrinkage intensity parameter $\delta = 0.50, 0.75$ and the level critical value of the test statistic $\alpha = 0.01, 0.05, 0.10$. When $\Delta = 0$, the SRE for all the proposed estimators are greater than one and increase as the number of inactive predictors q is increased. The restricted estimator has maximum SRE in all configurations in this case. The linear shrinkage estimator depends on the choice of δ. For the higher value of δ, its SRE approaches to the SRE of the restricted estimator. The performance of the preliminary test estimator is better than the shrinkage preliminary test estimator at $\Delta = 0$. The simulation results reveal that as $\Delta > 0$, the SREs of the restricted estimator and the other estimators decline sharply. When Δ moves away from zero, the SRE of the SP is greater than that of the PT. At $\Delta \geq 0$, the performance of the positive shrinkage estimator is much better than the shrinkage estimator.

Table 1: The *SREs* of the estimators with respect to the unrestricted estimator for $\delta = 0.50$ and $0 \leq \Delta \leq 2$

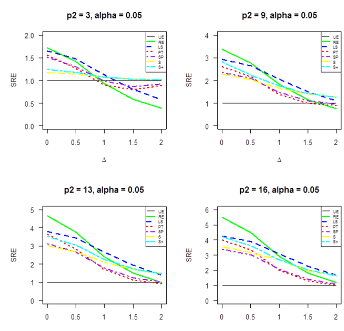
q	Δ	*RE*	*LS*	*PT*			*SP*			*S*	S^+
				α			α				
				0.01	0.05	0.10	0.01	0.05	0.10		
3	0.0	1.724	1.460	1.684	1.564	1.432	1.438	1.370	1.292	1.178	1.246
	0.5	1.419	1.392	1.364	1.269	1.195	1.343	1.257	1.191	1.137	1.173
	1.0	0.927	1.228	0.891	0.893	0.906	1.110	1.039	1.020	1.066	1.073
	1.5	0.588	1.028	0.682	0.788	0.853	0.919	0.936	0.952	1.023	1.028
	2.0	0.388	0.838	0.705	0.880	0.926	0.886	0.954	0.971	1.009	1.010
9	0.0	3.368	2.108	3.184	2.602	2.344	2.052	1.852	1.750	2.282	2.518
	0.5	2.769	2.038	2.581	2.153	1.855	1.955	1.746	1.582	2.033	2.222
	1.0	1.808	1.855	1.605	1.379	1.259	1.650	1.416	1.290	1.656	1.740
	1.5	1.146	1.615	1.021	0.991	0.985	1.268	1.128	1.076	1.397	1.420
	2.0	0.757	1.367	0.832	0.892	0.923	1.031	1.002	0.996	1.252	1.254
13	0.0	4.672	2.433	4.455	3.674	3.098	2.388	2.200	2.030	3.009	3.359
	0.5	3.790	2.370	3.500	2.838	2.338	2.278	2.035	1.816	2.685	3.045
	1.0	2.240	2.184	2.122	1.717	1.515	1.964	1.639	1.469	2.135	2.266
	1.5	1.510	1.928	1.279	1.145	1.091	1.507	1.277	1.181	1.731	1.761
	2.0	0.989	1.654	0.927	0.944	0.962	1.162	1.060	1..029	1.485	1.490
16	0.0	5.488	2.573	5.021	4.023	3.382	2.492	2.282	2.112	3.573	4.201
	0.5	4.486	2.513	4.058	3.347	2.668	2.404	2.190	1.941	3.171	3.625
	1.0	2.893	2.338	2.567	1.997	1.722	2.150	1.782	1.585	2.481	2.675
	1.5	1.817	2.092	1.499	1.284	1.193	1.661	1.369	1.247	1.963	2.020
	2.0	1.194	1.822	1.040	1.004	0.994	1.254	1.108	1.061	1.642	1.656

5. A real data

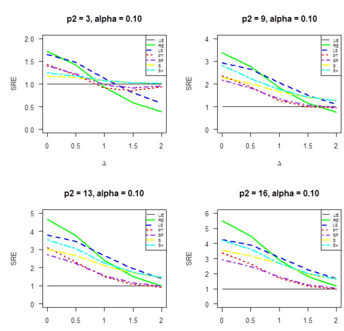
In this section, we have considered the **salamanders**' dataset, which is available in the package of glmmTMB in R software. This data was acquired from Dryad (Price et al. (2016)). The data contains $N = 644$ observations with $n = 161$ subjects, and each of them was sampled $n_i = 4$ times. The dependent variable and covariates are described in Table 3. The mean and variance of the data are 1.33 and 6.95, respectively. Since the variance is more than the mean, the data are over-dispersed. So, a NBMM is an appropriate model for describing this data. We have considered seven covariates **mined** (X_1), **cover** (X_2), **sample** (X_3),



(a) The SREs of the estimators with respect to the unrestricted estimator for $q = 3, 9, 13, 16$, $\alpha = 0.01$ and $\delta = 0.75$



(b) The SREs of the estimators with respect to the unrestricted estimator for $q = 3, 9, 13, 16$, $\alpha = 0.05$ and $\delta = 0.75$



(c) The SREs of the estimators with respect to the unrestricted estimator for $q = 3, 9, 13, 16$, $\alpha = 0.10$ and $\delta = 0.75$

Table 2: The $SREs$ of the estimators with respect to the unrestricted estimator for $\delta = 0.75$ and $0 \leq \Delta \leq 2$

q	Δ	RE	LS	PT			SP			S	S^{+}
				α			α				
				0.01	0.05	0.10	0.01	0.05	0.10		
3	0.0	1.724	1.649	1.684	1.564	1.432	1.615	1.510	1.394	1.178	1.246
	0.5	1.419	1.474	1.364	1.269	1.195	1.412	1.304	1.221	1.137	1.173
	1.0	0.927	1.123	0.891	0.893	0.906	1.001	0.984	0.977	1.066	1.073
	1.5	0.588	0.805	0.682	0.788	0.853	0.805	0.867	0.907	1.023	1.028
	2.0	0.388	0.576	0.705	0.880	0.926	0.798	0.919	0.951	1.009	1.010
9	0.0	3.368	2.922	3.184	2.602	2.344	2.791	2.359	2.157	2.282	2.815
	0.5	2.769	2.642	2.581	2.153	1.855	2.476	2.089	1.814	2.033	2.222
	1.0	1.808	2.055	1.605	1.379	1.259	1.776	1.478	1.326	1.656	1.740
	1.5	1.146	1.500	1.021	0.991	0.985	1.200	1.089	1.050	1.397	1.420
	2.0	0.757	1.088	0.832	0.892	0.923	0.949	0.958	0.967	1.252	1.254
13	0.0	4.672	3.795	4.455	3.674	3.098	3.639	3.146	2.736	3.009	3.539
	0.5	3.790	3.443	3.500	2.838	2.338	3.212	2.662	2.230	2.685	3.045
	1.0	2.420	2.678	2.122	1.717	1.515	2.304	1.815	1.580	2.135	2.266
	1.5	1.510	1.951	1.279	1.145	1.091	1.502	1.266	1.172	1.731	1.761
	2.0	0.989	1.414	0.927	0.944	0.962	1.081	1.021	1.006	1.485	1.490
16	0.0	5.488	4.259	5.021	4.023	3.382	3.990	3.368	2.933	3.573	4.201
	0.5	4.486	3.891	4.058	3.347	2.668	3.580	3.036	2.487	3.171	3.625
	1.0	2.893	3.073	2.567	1.997	1.722	2.703	2.068	1.767	2.675	2.675
	1.5	1.817	2.273	1.499	1.284	1.193	1.734	1.399	1.264	1.963	2.020
	2.0	1.194	1.665	1.040	1.004	0.994	1.199	1.080	1.042	1.642	1.656

DOP (X_4), **Wtemp** (X_5), **DOY** (X_6), and **spp**(X_7) as the fixed covariates and **site** as the random covariate. To determine the active and inactive covariates, we have used the variable selection method based on the Akaike information criterion (AIC). This criterion shows that the coefficients of mined, cover, DOY, and spp are the significant fixed parameters, and the coefficient of the sample, DOP, and Wtemp are inactive fixed parameters. Hence, the restricted subspace is $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7)^T$, and $\mathbf{R}$ and $\mathbf{r}$ is given by

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The above restriction can be written as $(\beta_3, \beta_4,\ \beta_5)^T = (0, 0, 0)^T$. So, $p = 7$, $p_1 = 4$, and $q = 3$. We have assumed $\delta = 0.75$ and $\alpha = 0.01$.
In order to examine the performance of the various estimation strategies, we have chosen $m = 300$ observations with replacement $M = 1,000$ times from the original

dataset using bootstrap sampling,. The point estimates, standard errors, and the relative efficiencies (RE) of the significant fixed parameters are reported in Table 4. The results completely agree with the theory established in Section 3 and our numerical results in Section 4.

6. Conclusion

The main aim of this article was the estimation of the fixed parameters in the negative binomial mixed model based on shrinkage estimators and comparing their performance with the unrestricted estimator when certain prior subspace information is available. We computed the properties of the suggested estimators based on simulated relative efficiency using Monte Carlo experiments in R statistical software. The simulation study results revealed that when the subspace information was correct, the SREs of all estimators were greater than one and increased as the number of inactive fixed parameters (q) was increased and the performance of the restricted estimator was better than the other estimators. The SRE of the positive shrinkage estimator was better than the shrinkage estimator at $\Delta \geq 0$.

Acknowledgments

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

Table 3: The list of variables for the salamanders' dataset

Variable	Description
Response variable	
count	number of salamanders observed
covariate	
mined	factor indicating whether the site was affected by mountain top removal coal mining
cover	amount of cover objects in the stream (scaled)
sample	repeated sample
DOP	Days since precipitation (scaled)
Wtemp	water temperature (scaled)
DOY	day of year (scaled)
spp	abbreviated species name, possibly also life stage
site	name of a location where repeated samples were taken

References

Arashi, M. and Roozbeh, M. (2019), Some improved estimation strategies in high-dimensional semiparametric regression models with application to the Riboflavin production data, *Statistical Papers*, **60(3)**, 317-336.

Ahmed, S. E. (1992), Shrinkage preliminary test estimation in multivariate normal distributions, *Journal of statistical computation and simulation*, **43(3-4)**, 177-195.

Asl, M. N., Bevrani, H., Belaghi, R. A. and Mansson, K. (2021). Ridge-type shrinkage estimators in generalized linear models with an application to prostate cancer data. *Statistical Papers*, **62**, 1043-1085.

Hossain, S., Ahmed, S. E. and Doksum, K. A. (2015) Shrinkage, pretest, and penalty estimators in generalized linear models, *Statistical Methodology*, **24**, 52-68.

Hossain, S. and Howlader, H.A. (2015), Estimation techniques for regression model with zero-inflated poisson data,*International Journal of Statistics and Probability*, **4(4)**, 64-76.

Hossain, S., Thomson, T. and Ahmed, S. E. (2018), Shrinkage estimation in linear mixed models for longitudinal data, *Metrika*, **81(5)**, 569-586.

Judge, G. G. and Bock, M. E. and Bock, M. E. (1978), *The Statistical implication of pre-test and Stein-rule estimators in econometrics*, **25**, North-Holland.

McCulloch, C. E., Searle, S. R. and Neuhaus, J. R. (2008), *Generalized, linear, and mixed models*, 2nd ed. Wiley, Hoboken, New Jersey.

Noori Asl, M., Bevrani, H. and Arabi Belaghi, R. (2020), Penalized and ridge-type shrinkage estimators in Poisson regression model, *Communications in Statistics-Simulation and Computation*, online.

Price, S. J., Muncy, B. L., Bonner, S. J., Drayer, A. N. and Barton, C. D. (2016), Effects of mountaintop removal mining and valley filling on the occupancy and abundance of stream salamanders,*Journal of Applied Ecology*, **53(2)**, 459-468.

Roozbeh, M., Arashi, M. and Hamzah, N. A. (2020), Generalized cross validation for simultaneous optimization of tuning parameters in ridge regression, *Iranian Journal of Science and Technology, Transactions A: Science*, **44(2)**, 473-485.

Table 4: Estimates, standard errors (in parentheses), and SREs of the coefficients for the active fixed parameters with respect to the unrestricted estimator for the salamanders' dataset

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_6$	$\hat{\beta}_7$	Relative Efficiency
UE	1.152 (0.425)	0.176 (0.194)	1.418 (0.318)	0.133 (0.053)	1.000
RE	−0.111 (0.172)	0.454 (0.223)	0.273 (0.130)	0.138 (0.058)	3.283
LS	0.280 (0.207)	0.385 (0.202)	0.559 (0.147)	0.137 (0.054)	2.530
PT	−0.107 (0.192)	0.454 (0.223)	0.275 (0.147)	0.138 (0.058)	3.261
SP	0.283 (0.218)	0.385 (0.202)	0.561 (0.158)	0.137 (0.054)	2.518
S	0.623 (0.630)	0.329 (0.232)	0.797 (0.520)	0.137 (0.053)	1.721
S^+	0.672 (0.500)	0.320 (0.215)	0.840 (0.381)	0.136 (0.053)	1.746

Saleh, A. M. E. (2006), *Theory of preliminary test and Stein-type estimation with applications*, John Wiley, New York.

Saleh, A. M. E., Arashi, M., and Kibria, B. G. (2019), *Theory of ridge regression estimation with applications*, John Wiley and Sons.

Thomson, T. and Hossain, S. (2018), Efficient shrinkage for generalized linear mixed models under linear restrictions, The Indian Journal of Statistics, *Sankhya A*, **80(2)**, 385-410.

Yuzbasi, B., Arashi, M. and Ahmed, S. E. (2020), Shrinkage estimation strategies in generalized ridge regression models: low/high-dimension regime, *International Statistical Review*, **88(1)**, 229-251.

Zandi, Z., Bevrani, H. and Arabi Belaghi, R. (2021), Using shrinkage strategies to estimate fixed effects in zero-inflated negative binomial mixed model, *Communications in Statistics-Simulation and Computation*, online.

Appendix 1. Proof of Theorem 3.2

The following Lemma is needed for the derivation of the bias and risk functions.

Lemma 6.1. *Let $\boldsymbol{y}$ be a q-dimensional random vector distributed as $\mathcal{N}_q(\boldsymbol{\mu_y}, \boldsymbol{\Sigma_y})$. Then, for any measurable function φ, we have*

$$E[\boldsymbol{y}\,\varphi(\boldsymbol{y}^T\boldsymbol{y})] = \boldsymbol{\mu_y}\,E[\varphi(\chi^2_{q+2}(\Delta^*))], \tag{6.21}$$

$$E[\boldsymbol{y}^T\boldsymbol{y}\,\varphi(\boldsymbol{y}^T\boldsymbol{y})] = \boldsymbol{\Sigma_y}\,E[\varphi(\chi^2_{q+2}(\Delta^*))] + \boldsymbol{\mu_y}^T\,\boldsymbol{\mu_y}E[\varphi(\chi^2_{q+4}(\Delta^*))], \tag{6.22}$$

where Δ^* is the non-centrality parameter.

Proof : See Judge et al. (1978).

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

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{UE}) &= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta})] = E(\mathbf{Z}) = \mathbf{0} \\
ADB(\hat{\boldsymbol{\beta}}^{RE}) &= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta})] = E(\mathbf{V}) = -\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi} \\
ADB(\hat{\boldsymbol{\beta}}^{LS}) &= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{LS} - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} E[\sqrt{n}\,(\delta\,\hat{\boldsymbol{\beta}}^{RE} + (1-\delta)\,\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}) - \delta\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})] \\
&= E(\mathbf{Z}) - \delta\,E(\mathbf{W}) \\
&= -\delta\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi} \\
ADB(\hat{\boldsymbol{\beta}}^{SP}) &= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{SP} - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \delta\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha}) - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} E[\sqrt{n}(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta}) - \delta\,\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})] \\
&= E(\mathbf{Z}) - \delta\,E(\mathbf{W}\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})) \\
&= -\delta\,E(\mathbf{W}\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})),
\end{aligned}$$

based on Equation (6.21), we can write

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{SP}) &= -\delta\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E[I(\chi^2_{q+2}(\Delta^*) \le \chi^2_{q,\alpha})] \\
&= -\delta\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,P(\chi^2_{q+2}(\Delta^*) \le \chi^2_{q,\alpha}) \\
&= -\delta\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*).
\end{aligned}$$

If $\delta = 1$,

$$ADB(\hat{\boldsymbol{\beta}}^{PT}) = -\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*).$$

In a similar way, we can obtain

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{S}) &= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{S} - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{RE} + (1-(q-2)\,\mathbf{D}_n^{-1})\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} \Big[E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{RE} - \boldsymbol{\beta})] + E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})] \\
&\qquad - (q-2)\,E[\mathbf{D}_n^{-1}\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})]\Big] \\
&= \lim_{n\to\infty} \Big[E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \boldsymbol{\beta})] - (q-2)\,E[\mathbf{D}_n^{-1}\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})]\Big] \\
&= E(\mathbf{Z}) - (q-2)\,E(\mathbf{D}_n^{-1}\mathbf{W}) \\
&= -(q-2)\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E\Big[\frac{1}{T_1}\Big].
\end{aligned}$$

$$\begin{aligned}
ADB(\hat{\boldsymbol{\beta}}^{S^+}) &= \lim_{n\to\infty} E[\sqrt{n}\,(\hat{\boldsymbol{\beta}}^{S^+} - \boldsymbol{\beta})] \\
&= \lim_{n\to\infty} E[\sqrt{n}\,[\hat{\boldsymbol{\beta}}^{S} - \boldsymbol{\beta} - (1-(q-2)\,\mathbf{D}_n^{-1})\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\,I(\mathbf{D}_n < q-2)] \\
&= ADB(\hat{\boldsymbol{\beta}}^{S}) - \lim_{n\to\infty} E\sqrt{n}\,[(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE})\,(1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n < q-2)] \\
&= ADB(\hat{\boldsymbol{\beta}}^{S}) - E[\mathbf{W}\,(1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n < q-2)] \\
&= ADB(\hat{\boldsymbol{\beta}}^{S}) - E[\mathbf{W}\,I(\mathbf{D}_n < q-2)] + (q-2)\,E[\mathbf{W}\,\mathbf{D}_n^{-1}\,I(\mathbf{D}_n < q-2)] \\
&= ADB(\hat{\boldsymbol{\beta}}^{S}) - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) + (q-2)\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E\bigg(\frac{I(T_1 < q-2)}{T_1}\bigg).
\end{aligned}$$

Appendix 2. Proof of Theorem 3.4

We first derive the asymptotic covariance of the estimators as defined in Equation (3.20)

$$
\begin{aligned}
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{UE}) &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^{UE}-\boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{UE}-\boldsymbol{\beta})^T\Big)\\
&= \lim_{n\to\infty} E(\mathbf{Z}_n\,\mathbf{Z}_n{}^T)\\
&= E(\mathbf{Z}\,\mathbf{Z}^T)\\
&= Var(\mathbf{Z}) + E(\mathbf{Z})\,E(\mathbf{Z}^T)\\
&= \mathbf{B}^{-1}\\
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{RE}) &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^{RE}-\boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{RE}-\boldsymbol{\beta})^T\Big)\\
&= \lim_{n\to\infty} E(\mathbf{V}_n\,\mathbf{V}_n{}^T)\\
&= E(\mathbf{V}\,\mathbf{V}^T)\\
&= Var(\mathbf{V}) + E(\mathbf{V})\,E(\mathbf{V}^T)\\
&= \mathbf{B}^{-1} - \mathcal{J}\,\mathbf{R}\,\mathbf{B}^{-1} + (\mathcal{J}\,\boldsymbol{\xi})\,(\mathcal{J}\,\boldsymbol{\xi})^T
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{LS}) &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^{LS}-\boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{LS}-\boldsymbol{\beta})^T\Big)\\
&= \lim_{n\to\infty} E[(\mathbf{Z}_n-\delta\,\mathbf{W}_n)\,(\mathbf{Z}_n-\delta\,\mathbf{W}_n)^T]\\
&= E[(\mathbf{Z}-\lambda\,\mathbf{W})\,(\mathbf{Z}-\lambda\,\mathbf{W})^T]\\
&= E[\mathbf{Z}\,\mathbf{Z}^T] - 2\delta\,E[\mathbf{Z}\,\mathbf{W}^T] + \delta^2\,E[\mathbf{W}\,\mathbf{W}^T]\\
&= \mathbf{B}^{-1} - 2\delta\,\underbrace{E[\mathbf{Z}\,\mathbf{W}^T]}_{e_1} + \delta^2\,[\mathcal{J}\,\mathbf{R}\,\mathbf{B}^{-1} + (\mathcal{J}\,\boldsymbol{\xi})\,(\mathcal{J}\,\boldsymbol{\xi})^T]
\end{aligned}
$$

Using the conditional expectation, e_1 becomes

$$
\begin{aligned}
e_1 &= E[\mathbf{Z}\,\mathbf{W}^T]\\
&= E\big(E[\mathbf{Z}\,\mathbf{W}^T\,|\mathbf{W}]\big)\\
&= E\big(\,\mathbf{W}^T\,E[\mathbf{Z}\,|\,\mathbf{W}]\big)
\end{aligned}
$$

Based on Lemma 3.1, we have

$$
\begin{pmatrix}\mathbf{Z}_n\\ \mathbf{W}_n\end{pmatrix} \xrightarrow{D} \begin{pmatrix}\mathbf{Z}\\ \mathbf{W}\end{pmatrix} \sim \mathcal{N}_{2p}\left[\begin{pmatrix}\mathbf{0}\\ \mathcal{J}\boldsymbol{\xi}\end{pmatrix}, \begin{pmatrix}\mathbf{B}^{-1} & \mathcal{J}\mathbf{R}\mathbf{B}^{-1}\\ \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1} & \mathcal{J}\mathbf{R}\,\mathbf{B}^{-1}\end{pmatrix}\right]
$$

Using the conditional expectation of a multivariate normal distribution, we can write

$$E[\mathbf{Z}\,|\,\mathbf{W}] = E\Big[\mathbf{Z} + \Big(\frac{cov(\mathbf{Z},\mathbf{W})}{\sigma(\mathbf{Z})\,\sigma(\mathbf{W})}\Big)\Big(\frac{\sigma(\mathbf{Z})}{\sigma(\mathbf{W})}\Big)(\mathbf{W} - E(\mathbf{W}))\Big],$$

based on Lemma 3.1,

$$E[\mathbf{Z}] = \mathbf{0} \quad , \quad E[\mathbf{W}] = \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi} \quad , \quad \sigma^2[\mathbf{Z}] = \mathbf{B}^{-1},$$

$$\sigma^2[\mathbf{W}] = \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1} \quad , \quad cov[\mathbf{Z},\mathbf{W}] = \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}$$

So,

$$\begin{aligned} E[\mathbf{Z}\,|\,\mathbf{W}] &= E\Big[\mathbf{0} + \frac{\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}}{\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}}(\mathbf{W} - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\Big] \\ &= E[\mathbf{W} - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}] \end{aligned}$$

Hence, we have

$$\begin{aligned} e_1 &= E\big[\,(\mathbf{W} - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,\mathbf{W}^T\big] \\ &= E[\mathbf{W}\,\mathbf{W}^T] - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E[\mathbf{W}] \\ &= \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1} \end{aligned}$$

Therefore

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{LS}) = \mathbf{B}^{-1} - \delta\,(2-\delta)\,\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1} + \delta^2\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T.$$

Next we obtain $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{SP})$ as follows

$$\begin{aligned} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{SP}) &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^{SP} - \boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{SP} - \boldsymbol{\beta})^T\Big) \\ &= \lim_{n\to\infty} E[\,\{\mathbf{Z}_n - \delta\,\mathbf{W}_n\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})\}\,\{\mathbf{Z}_n - \delta\,\mathbf{W}_n\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})\}^T\,] \\ &= \lim_{n\to\infty} E(\mathbf{Z}_n\,\mathbf{Z}_n^T) - 2\delta \lim_{n\to\infty} E(\mathbf{Z}_n\,\mathbf{W}_n^T\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})) \\ &\quad + \delta^2 \lim_{n\to\infty} E(\mathbf{W}_n\,\mathbf{W}_n^T\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})) \\ &= E(\mathbf{Z}\,\mathbf{Z}^T) - 2\delta\,E(\mathbf{Z}\,\mathbf{W}^T\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})) + \delta^2\,E(\mathbf{W}\,\mathbf{W}^T\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha})) \\ &= \mathbf{B}^{-1} - 2\delta\,\underbrace{E(\mathbf{Z}\,\mathbf{W}^T\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha}))}_{e_2} + \delta^2\,\underbrace{E(\mathbf{W}\,\mathbf{W}^T\,I(\mathbf{D}_n) \le \chi^2_{q,\alpha}))}_{e_3}, \end{aligned}$$

using Equation (6.22), we have

$$e_3 = \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) + (\boldsymbol{\mathcal{J}}\,\boldsymbol{\delta})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\delta})^T\,\mathbf{H}_{q+4}(\chi^2_{q,\alpha};\Delta^*),$$

and by using conditional expectation, e_2 becomes

$$\begin{aligned}
e_2 &= E[\mathbf{Z}\,\mathbf{W}^T\, I(\mathbf{D}_n) \le \chi^2_{q,\alpha})] \\
&= E\big[E(\mathbf{Z}\,\mathbf{W}^T\, I(\mathbf{D}_n) \le \chi^2_{q,\alpha}) \,|\, \mathbf{W})\big] \\
&= E\big[E(\mathbf{Z} \,|\, \mathbf{W})\,\mathbf{W}^T\, I(\mathbf{D}_n) \le \chi^2_{q,\alpha})\big] \\
&= E\big[\,\{E(\mathbf{Z}) + (\mathbf{W} - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\}\mathbf{W}^T\, I(\mathbf{D}_n) \le \chi^2_{q,\alpha})\big] \\
&= \underbrace{E[\mathbf{W}\,\mathbf{B}^T\, I(\mathbf{D}_n) \le \chi^2_{q,\alpha})]}_{e_3} - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\, E[\mathbf{W}^T\, I(\mathbf{D}_n) \le \chi^2_{q,\alpha})] \\
&= \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) - (\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,[\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) - \mathbf{H}_{q+4}(\chi^2_{q,\alpha};\Delta^*)]
\end{aligned}$$

Therefore,

$$\begin{aligned}
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{SP}) = \mathbf{B}^{-1} &- \delta\,(2-\delta)\,\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) \\
&+ (\boldsymbol{\mathcal{J}}\,\boldsymbol{\beta})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\beta})^T\,[2\delta\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) - \delta\,(2-\delta)\,\mathbf{H}_{q+4}(\chi^2_{q,\alpha};\Delta^*)].
\end{aligned}$$

For $\delta = 1$, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{PT})$ reduces to

$$\begin{aligned}
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{PT}) = \mathbf{B}^{-1} &- \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) \\
&+ (\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,[2\,\mathbf{H}_{q+2}(\chi^2_{q,\alpha};\Delta^*) - \mathbf{H}_{q+4}(\chi^2_{q,\alpha};\Delta^*)].
\end{aligned}$$

Now we obtain $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{S})$ as follows:

$$\begin{aligned}
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{S}) &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^{S} - \boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{S} - \boldsymbol{\beta})^T\Big) \\
&= \lim_{n\to\infty} E\Big[\sqrt{n}\Big(\hat{\boldsymbol{\beta}}^{RE} + (1-(q-2)\mathbf{D}_n^{-1})\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}\Big) \\
&\quad \times \sqrt{n}\Big(\hat{\boldsymbol{\beta}}^{RE} + (1-(q-2)\mathbf{D}_n^{-1})\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}\Big)^T\Big] \\
&= \lim_{n\to\infty} E[(\mathbf{Z}_n - (q-2)\,\mathbf{D}_n^{-1}\,\mathbf{W}_n)\,(\mathbf{Z}_n - (q-2)\,\mathbf{D}_n^{-1}\,\mathbf{W}_n)^T] \\
&= E[(\mathbf{Z} - (q-2)\,\mathbf{D}_n^{-1}\,\mathbf{W})\,(\mathbf{Z} - (q-2)\,\mathbf{D}_n^{-1}\,\mathbf{W})^T] \\
&= E[(\mathbf{Z}\,\mathbf{Z}^T] - 2(q-2)\,\underbrace{E[\mathbf{W}\,\mathbf{Z}^T\,\mathbf{D}_n^{-1}]}_{e_4} + (q-2)^2\,\underbrace{E[\mathbf{W}\,\mathbf{W}^T\,\mathbf{D}_n^{-2}]}_{e_5},
\end{aligned}$$

similar to e_1, we can write e_4 as follows:

$$\begin{aligned}
e_4 &= E[\mathbf{W}\,\mathbf{Z}^T\,\mathbf{D}_n^{-1}] \\
&= E[\mathbf{W}\,\mathbf{W}^T\,\mathbf{D}_n^{-1}] - \boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,E[\mathbf{W}\,\mathbf{D}_n^{-1}] \\
&= \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,E\Big[\frac{1}{T_1}\Big] - (\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,\Big(E\Big[\frac{1}{T_1}\Big] - E\Big[\frac{1}{T_2}\Big]\Big),
\end{aligned}$$

and by using Equation (6.22), e_5 becomes

$$\begin{aligned} e_5 &= E[\mathbf{W}\,\mathbf{W}^T\,\mathbf{D}_n^{-2}] \\ &= \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,E\Big[\frac{1}{T_1^2}\Big] + (\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,E\Big[\frac{1}{T_2^2}\Big] \end{aligned}$$

Therefore,

$$\begin{aligned} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^S) &= \mathbf{B}^{-1} - 2(q-2)\,\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,E\Big[\frac{1}{T_1}\Big] \\ &+ 2(q-2)\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T \left\{E\Big[\frac{1}{T_1}\Big] - E\Big[\frac{1}{T_2}\Big]\right\} \\ &+ (q-2)^2\,\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,E\Big[\frac{1}{T_1^2}\Big] + (\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,E\Big[\frac{1}{T_2^2}\Big] \\ &= \mathbf{B}^{-1} + (q-2)\,\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\left\{(q-2)\,E\Big[\frac{1}{T_1^2}\Big] - 2E\Big[\frac{1}{T_1}\Big]\right\} \\ &+ (q-2)\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\Big\{-2E\Big[\frac{1}{T_2}\Big] + 2E\Big[\frac{1}{T_1}\Big] + (q-2)\,E\Big[\frac{1}{T_2^2}\Big]\Big\} \end{aligned}$$

Finally, we can write $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{S^+})$ as follows:

$$\begin{aligned} \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{S^+}) &= \lim_{n\to\infty} E\Big(\sqrt{n}(\hat{\boldsymbol{\beta}}^{S^+} - \boldsymbol{\beta})\,\sqrt{n}(\hat{\boldsymbol{\beta}}^{S^+} - \boldsymbol{\beta})^T\Big) \\ &= \lim_{n\to\infty} E\Big[\sqrt{n}\,\Big(\hat{\boldsymbol{\beta}}^S - (1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n < q-2)\,(\hat{\boldsymbol{\beta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}\Big) \\ &\times \sqrt{n}\,\Big(\hat{\boldsymbol{\beta}}^S - (1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n < q-2)\,(\hat{\boldsymbol{\delta}}^{UE} - \hat{\boldsymbol{\beta}}^{RE}) - \boldsymbol{\beta}\Big)^T\Big] \\ &= \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^S) - 2E[\mathbf{W}\,\mathbf{V}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n < q-2)] \\ &- 2E[\mathbf{W}\,\mathbf{W}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})^2\,I(\mathbf{D}_n < q-2)] \\ &+ E[\mathbf{W}\,\mathbf{W}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})^2\,I(\mathbf{D}_n < q-2)] \\ &= \boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^S) - 2\,\underbrace{E[\mathbf{W}\,\mathbf{V}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n < q-2)]}_{e_6} \\ &- \underbrace{E[\mathbf{W}\,\mathbf{W}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})^2\,I(\mathbf{D}_n < q-2)]}_{e_7}, \end{aligned}$$

now we obtain e_6

$$
\begin{aligned}
e_6 &= E[\mathbf{W}\,\mathbf{V}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})\,I(\mathbf{D}_n<q-2)]\\
&= E[\mathbf{W}\,E\{\mathbf{V}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})\;I(\mathbf{D}_n<q-2)\,|\,\mathbf{W}\}]\\
&= E[\mathbf{W}\,E\{-\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}+\mathbf{0}\;\times(\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1})^{-1}\,(\mathbf{W}-\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\}^T\\
&\times(1-(q-2)\,\mathbf{D}_n^{-1})\;I(\mathbf{D}_n<q-2)]\\
&= -E[\mathbf{W}\,\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi}\,(1-(q-2)\,\mathbf{D}_n^{-1})\;I(\mathbf{D}_n<q-2)]\\
&= -(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,E\Big[\Big(1-\frac{q-2}{T_1}\Big)\,I(T_1<q-2)\Big],
\end{aligned}
$$

and based on Equation (6.22), e_7 becomes

$$
\begin{aligned}
e_7 &= E[\mathbf{W}\,\mathbf{W}^T\,(1-(q-2)\,\mathbf{D}_n^{-1})^2\,I(\mathbf{D}_n<q-2)]\\
&= \boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{I}^{-1}\,E\Big[\Big(1-\frac{q-2}{T_1}\Big)^2\,I(T_1<q-2)\Big]\\
&+(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,E\Big[\Big(1-\frac{q-2}{T_2}\Big)^2\,I(T_2<q-2)\Big].
\end{aligned}
$$

Therefore, $\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{S+})$ becomes

$$
\begin{aligned}
\boldsymbol{\Sigma}(\hat{\boldsymbol{\beta}}^{S+}) &= \mathbf{V}(\hat{\boldsymbol{\beta}}^{S})+2(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,E\Big[\Big(1-\frac{q-2}{T_1}\Big)\,I(T_1<q-2)\Big]\\
&-(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})\,(\boldsymbol{\mathcal{J}}\,\boldsymbol{\xi})^T\,E\Big[\Big(1-\frac{q-2}{T_1}\Big)^2\,I(T_2<q-2)\Big]\\
&-\boldsymbol{\mathcal{J}}\,\mathbf{R}\,\mathbf{B}^{-1}\,E\Big[\Big(1-\frac{q-2}{T_1}\Big)^2\,I(T_1<q-2)\Big].
\end{aligned}
$$

Now, the proof of Theorem 3.4 can be derived using the above results by following the definition of ADR.