HIGH-DIMENSIONAL BIAS-CORRECTED INFERENCE, WITH APPLICATIONS TO FMRI STUDIES

by

Xiaonan Zhu

B.S. in Statistics, Shandong University, China, 2012

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Xiaonan Zhu

It was defended on

August 2nd, 2019

and approved by

Satish Iyengar, PhD, Department of Statistics

Zhao Ren, PhD, Department of Statistics

Yu Cheng, PhD, Department of Statistics

Milos Hauskrecht, PhD, Department of Computer Science

Michele Bertocci, PhD, School of Medicine

Dissertation Advisors: Satish Iyengar, PhD, Department of Statistics,

Zhao Ren, PhD, Department of Statistics

HIGH-DIMENSIONAL BIAS-CORRECTED INFERENCE, WITH APPLICATIONS TO FMRI STUDIES

Xiaonan Zhu, PhD

University of Pittsburgh, 2019

In neuroimaging studies, measures of neural structure and function are used to try to predict clinical outcomes of patients. Identifying biomarkers that reflect underlying neuropathological processes can provide promising neural targets for future therapeutic interventions. This identification is typically done using linear or generalized linear models (GLM) with many covariates and relatively few subjects. Thus, regularization is used to select the salient covariates in the model. In this thesis, we compare the performance of the least absolute shrinkage and selection operator (LASSO) regression, adaptive LASSO regression, debiased LASSO regression, and regularized zero-inflated Poisson (ZIP) regression model in two simulation settings. The performance of LASSO regression with Poisson and Gaussian models are similar but for all these approaches the zero-inflated model outperforms the rest. We apply these approaches to the data from the Longitudinal Assessment of Manic Symptoms (LAMS) study. We then study the bias correction of GLM and the application on ZIP data. We apply a decorrelated score approach to address Poisson distributed data and introduce Cornish-Fisher correction to the decorrelated score test. In high-dimension settings, the Cornish-Fisher correction can improve the performance of

decorrelated score test for ZIP data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 BACKGROUND ON STATISTICAL INFERENCE

Statistical inference can be thought of as the process of drawing conclusions about a population from a sample. It is the combination of model selection, estimation and formulation of a hypothesis test. One of most familiar examples of statistical inference is variable selection in regression. We consider the high-dimensional linear regression model

$$Y = X\beta + \epsilon, \tag{1.1}$$

where Y is an n-dimensional response vector and X is an $n \times p$ fixed or random design matrix. In high dimensional settings $p \gg n$. β is an unknown p dimension parameter and ϵ is an n-dimensional error vector with independent and identically distributed (i.i.d.) entries ϵ_i for $i = 1, 2, ..., n$. We assume $\mathbb{E}\epsilon_i = 0$ and $Var\epsilon_i = \sigma^2$. The error vector ϵ and design matrix X are independent for a random design. The goal is to identify the non-zero entries of β. Least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 1996) is one of the first coefficient methods to solve such problems. Various methods have been proposed to improve LASSO for model selection with weaker assumptions on signals, such as adaptive LASSO (Zou, 2006) and some of other non-convex penalized methods such as smoothly clipped absolute

deviations (SCAD) penalty (Fan and Li, 2001) and minimax concave penalty (MCP) (Zhang, 2010). More recently, statistical inference such as p-value and confidence interval for each coefficient was made possible using bias-corrected LASSO (Zhang and Zhang (2014); Bühlmann (2013); Mitra and Zhang (2016); Taylor and Tibshirani (2016)). The idea of statistical inference and variable selection for linear regression model has also been extended to the generalized linear models (GLM) (Van de Geer et al., 2014) and latent variable models (Wang et al., 2014a). A recent overview can be found in (Dezeure et al., 2015). Motivated by our data, zero-inflated models might be more appropriate. Here we mainly review one method, namely, the zero-inflated Poisson (ZIP) model in high dimension.

1.2 MOTIVATION AND STRUCTURE OF THIS THESIS

In neuroimaging studies, measures of neural structure and function are used to try to predict clinical outcomes of patients. Identifying biomarkers that reflect underlying neuropathological processes can provide promising neural targets for future therapeutic interventions. In Bertocci et al. (2016), neural activity measured by functional magnetic resonance imaging (fMRI) using 80 youth from 3 clinical sites are analyzed and LASSO regression in the Poisson model is used to assess its ability to predict future levels of behavioral and emotional dysregulation in psychiatrically unwell youth.

In this thesis, we compare the performance of five statistical inference approaches in Chapter 2 and apply these approaches to the data from Bertocci et al. (2016) in Chapter 3. In Chapter 4, we study the bias correction of a GLM and the application on ZIP data. We apply a decorrelated score approach to address Poisson distributed

data and then introduce Cornish-Fisher correction to the decorrelated score test. We also discuss extensions of current approaches in Chapter 5.

2.0 ANALYSIS OF ZERO-INFLATED POISSON DATA

This Chapter consists of four parts. In Section 2.1, we state the statistical framework for the analysis of zero-inflated Poisson data. In Section 2.2, we review five models in high-dimensional estimation, including two LASSO regression models (2.2.1.1 and 2.2.1.2), adaptive LASSO model (2.2.3), debiased LASSO model (2.2.4) and regularized zero-inflated Poisson (RZIP) model (2.2.5). We also compare the results from different models given non-inflated data (2.2.6). Next, in Section 2.3, we compare the performance of the five approaches through two simulation studies.

2.1 STATISTICAL FRAMEWORK

Let Y_i be the outcome observed for the ith subject, $i = 1, 2, ..., n$. As for the covariates, let X_k be the kth predictor and β_k be the corresponding regression coefficient, for $k = 1, 2, ..., p$. In addition, let β_0 be the intercept term in the regression model. Our goal is to obtain a subset of predictors and the estimate of their coefficients β_k that are significant in the model. We call the subset of predictors with non-zero coefficients the support of regression model. We write $\|\cdot\|_1$ for vector l_1 norm and $\|\cdot\|_2$ for vector l_2 norm. We use $\mathbf{1}_n$ to denote the n-dimensional vector

$(1, 1, ...1)^T$ and $\mathbf{I}_n$ to denote $n \times n$ identity matrix. Throughout the thesis we assume $p > n$.

2.2 REVIEW OF EXISTING METHODS

2.2.1 LASSO regression

LASSO regression is an analysis method introduced in Tibshirani (1996). The key idea of LASSO regression is penalizing the l_1 norm of the regression coefficients in order to select a subset of covariates instead of all of them in the regression model. Although the original LASSO is formulated for linear regression, it has been extended to a large variety of statistical models. In this thesis, we apply LASSO regression in the Poisson and Gaussian models.

2.2.1.1 Gaussian model In the high-dimensional linear regression model (1.1), we assume the outcome Y follows a Gaussian distribution with mean μ and unknown variance σ^2. For each observation Y_i $(i = 1, 2, ..., n)$, let $\mathbb{E}Y_i = \mu_i$, $\mu_i = \beta_0 + \sum_j X_{ij}\beta_j$ and $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$; then the Gaussian LASSO estimators are defined as

$$\begin{aligned} \{\hat{\beta}_0, \hat{\beta}\} &= \arg\min_{\beta_0,\beta} \frac{1}{n}\|Y - \beta_0\mathbf{1}_n - X\beta\|_2^2 + \lambda\|\beta\|_1 \\ &= \arg\min_{\beta_0,\beta} \frac{1}{n}\sum_i |Y_i - \mu_i|^2 + \lambda\|\beta\|_1. \end{aligned} \tag{2.1}$$

2.2.1.2 Poisson model For the high-dimensional Poisson regression model, suppose that the outcome Y follows a Poisson distribution with mean μ. For each

observation Y_i $(i = 1, 2, ..., n)$, let $\mathbb{E}Y_i = \mu_i$, $\log(\mu_i) = \beta_0 + \sum_j X_{ij}\beta_j$, and $\beta = (\beta_1, \beta_2, ...\beta_p)^T$. Then the Poisson LASSO estimators are defined as

$$\{\hat{\beta}_0, \hat{\beta}\} = \arg\min_{\beta_0,\beta} -\frac{1}{n}\sum_i (Y_i \log(\mu_i) - \mu_i) + \lambda\|\beta\|_1. \tag{2.2}$$

2.2.1.3 Anscombe transform and Gaussian model In the previous section, the LASSO extension to Poisson model works well for certain applications. Another way of modeling the effect of covariates on the count type response is through the Anscombe transform (Anscombe, 1948), which is a variance-stabilizing transformation that transforms a Poisson distributed random variable into one with an approximately Gaussian distribution. Let Y_i be the ith outcome and $\mathbb{E}Y_i = \mu_i$. Then the Anscombe transformed ith outcome Y_i^* is defined as

$$Y_i^* = 2\sqrt{Y_i + \frac{3}{8}}. \tag{2.3}$$

We apply the Anscombe transform to the outcome Y assuming that the Anscombe transformed outcome Y^* is distributed as a Gaussian. Keeping the notation of Section 2.2.1.2, the Gaussian LASSO estimators are defined as

$$\{\hat{\beta}_0, \hat{\beta}\} = \arg\min_{\beta_0,\beta} \frac{1}{n}\|Y^* - \beta_0 - X\beta\|_2^2 + \lambda\|\beta\|_1 \tag{2.4}$$

2.2.2 Relationship between Poisson and Gaussian models

For the Anscombe transformed outcome Y_i^* in equation (2.3), if the original outcome Y_i follows a Poisson distribution with mean μ_i, then

$$\mathbb{E}Y_i^* = 2e^{-\mu_i} \sum_{k=0}^{\infty} \frac{\mu_i^k}{k!} \sqrt{k + \frac{3}{8}} \tag{2.5}$$

Suppose there exists an invertible transform $\mathcal{T}$ such that

$$\mathcal{T}\left(2e^{-\mu_i} \sum_{k=0}^{\infty} \frac{\mu_i^k}{k!} \sqrt{k + \frac{3}{8}} \right) = \mu_i; \tag{2.6}$$

then we can analyze the Anscombe transformed data with a Gaussian model and compare the corresponding Poisson log-likelihood with a Poisson model. We take the derivative of $\mathbb{E}Y_i^*$

$$\begin{aligned}
(\mathbb{E}Y_i^*)' &= \frac{d}{d\mu_i} \left(2e^{-\mu_i} \sum_{k=0}^{\infty} \frac{\mu_i^k}{k!} \sqrt{k + \frac{3}{8}} \right) \\
&= -2e^{-\mu_i} \sum_{k=0}^{\infty} \frac{\mu_i^k}{k!} \sqrt{k + \frac{3}{8}} + 2e^{-\mu_i} \sum_{k=1}^{\infty} \frac{k\mu_i^{k-1}}{k!} \sqrt{k - 1 + \frac{11}{8}} \\
&= 2e^{-\mu_i} \sum_{k=0}^{\infty} \frac{\mu_i^k}{k!} \left(\sqrt{k + \frac{11}{8}} - \sqrt{k + \frac{3}{8}} \right) > 0
\end{aligned}$$

The positive derivative indicates that the transform $\mathcal{T}$ is unique. However, unfortunately there is no closed form for $\mathcal{T}$, so we use a grid search to find the corresponding Poisson mean μ_i when the Gaussian mean is $\mathbb{E}Y_i^*$.

2.2.3 Adaptive LASSO regression for Gaussian model

In this section, we apply the adaptive LASSO estimation approach proposed by Zou (2006). By the definition of l_1 norm $\|\cdot\|_1$, the penalty term $\lambda\|\beta\|_1$ in LASSO regression model (2.1) can be written as $\lambda\sum_j 1\times|\beta_j|$, in which all coefficients β_j are penalized equally with weight factors equal to 1. In this section, we assign different weights to different coefficients and consider the weighted LASSO. Let β_j be the jth regression coefficient and w_j be the corresponding unknown weight factor, $\beta=(\beta_1,\beta_2,...,\beta_p)^T$, then the weighted LASSO likelihood function will be

$$\frac{1}{n}\|Y^*-\beta_0\mathbf{1}_n-X\beta\|_2^2+\lambda\sum_j w_j|\beta_j|. \tag{2.7}$$

We consider the adaptive LASSO regression model with the weighted factor $w_j=1/|\beta_j^{init}|^\gamma$:

$$\{\hat{\beta}_0,\hat{\beta}\}=\arg\min_{\beta_0,\beta}\frac{1}{n}\|Y^*-\beta_0\mathbf{1}_n-X\beta\|_2^2+\lambda\sum_j\frac{|\beta_j|}{|\widehat{\beta}_j^{init}|^\gamma}, \tag{2.8}$$

where $\widehat{\beta}_j^{init}$ is calculated from a $\sqrt{n}$ consistent estimator, for example, the ordinary least squares estimator $\widehat{\beta}_j^{OLS}$. γ is a pre-specified known constant (e.g. $\gamma=1$). To adjust the cases in which $\widehat{\beta}_j^{OLS}=0$, we use the initial estimator thus $\widehat{\beta}_j^{init}=|\widehat{\beta}_j^{OLS}|+1/\sqrt{n}$.

2.2.4 Debiased LASSO regression for Gaussian model

The debiased LASSO estimator was introduced in Zhang and Zhang (2014). In the LASSO regression model in Section 2.2.1.1, an l_1 penalty term added to the classical least squares likelihood function controls the number of non-zero coefficients. Meanwhile, this LASSO penalty term also introduces bias to the least squares estimate of β. For any predictor X_j, the corresponding univariate linear regression estimator can be written as

$$\widehat{\beta}_j^{linear} = \frac{z_j^T y}{z_j^T x_j} = \beta_j + \sum_{k \neq j} \frac{z_j^T x_k \beta_k}{z_j^T x_j} + \frac{z_j^T \epsilon}{z_j^T x_j}, \tag{2.9}$$

where $\|z_j\|_2 = 1$, $z_j^T x_j \neq 0$. This representation of linear estimator suggests a bias correction with a nonlinear initializer $\widehat{\beta}^{init}$:

$$\widehat{\beta}_j = \widehat{\beta}_j^{init} + \frac{z_j^T \{y - X\widehat{\beta}^{init}\}}{z_j^T x_j}, \tag{2.10}$$

where z_j is a relaxed orthogonalization of x_j against other predictor vectors X_{-j}. If $\widehat{\gamma}_j$ is the vector of coefficients from the lasso regression of x_j on X_j, i.e.

$$\widehat{\gamma}_j = \arg\min_b \frac{1}{2n} \|x_j - X_{-j} b\|_2^2 + \lambda_j \|b\|_1, \tag{2.11}$$

the lasso-generated score is $z_j = x_j - X_{-j}\widehat{\gamma}_j$ (Zhang and Zhang, 2014).

2.2.5 Regularized zero-inflated Poisson (RZIP) regression model

In this section, we introduce the regularized zero-inflated Poisson regression model. In practice, when we deal with morbidity data or clinical visit data, the outcomes are counts in which zero or positive integer values are observed. Often the number of zeros in the outcome cannot be accommodated by the Poisson model. In RZIP model, instead of assuming the outcome Y as a Poisson distributed variable, we assume that Y_i is generated from either the Poisson state or the zero state via a latent variable. Let Z_i be the zero state indicator for the ith observation Y_i. In other words, $Z_i = 1$ if Y_i is from the zero state and $Z_i = 0$ if Y_i is from the Poisson state with mean μ_i. Let $P(Z_i = 1) = \pi_i$ and $P(Z_i = 0) = 1 - \pi_i$, then the marginal probability of Y is:

$$P(Y_i = 0) = \pi_i + (1 - \pi_i)\exp(-\mu_i), \tag{2.12}$$

$$P(Y_i = y_i) = (1 - \pi_i)\mu_i^{y_i}\exp(-\mu_i)/y_i!,\ y_i > 0. \tag{2.13}$$

The log-likelihood function l_{RZIP} of $(Y_1, ...Y_n)$ is then given by

$$\begin{aligned} l_{RZIP} &= \sum_{Y_i=0} \log(\pi_i + (1 - \pi_i)\exp(-\mu_i)) + \sum_{Y_i\neq 0} \log((1 - \pi_i)\mu_i^{y_i}\exp(-\mu_i)/y_i!) \\ &= \sum_{Y_i=0} \log(\pi_i + (1 - \pi_i)\exp(-\mu_i)) + \sum_{Y_i\neq 0} \log(1 - \pi_i) \\ &\quad + \sum_{Y_i\neq 0} y_i \log \mu_i - \sum_{Y_i\neq 0} \mu_i - \sum_{Y_i\neq 0} \log(y_i!). \end{aligned} \tag{2.14}$$

The EM algorithm provided in Wang et al. (2014b) can be applied to the selected significant predictors and to estimate their coefficients. In this section, based on the structure of zero state we apply two different models: a uniform zero state probability model and varying zero state probability model. In the first model, we assume that for each Y_i, $P(Z_i = 1) = \pi$ for all $i = 1, 2, ..., n$. In the second model, we assume

that $P(Z_i = 1) = \pi_i$ are potentially different and fit a logistic regression model for the zero state probability. We can obtain the penalized log-likelihood function with some penalty function p_{RZIP} for the RZIP model,

$$\{\hat{\beta}, \hat{\pi}\} = \arg\max_{\beta,\pi} l_{RZIP} - np_{RZIP} \tag{2.15}$$

where $\beta = (\beta_1, \beta_2, ..., \beta_p)^T$.

2.2.5.1 Uniform zero state probability In the first case, we assume that the probability that an outcome is from the zero state is a fixed value π for all observations. If $\log \mu_i = \beta_0 + \sum_k X_{ij}\beta_k$ and $X_{i0} = 1$, then the log-likelihood function l_{RZIP} is defined as:

$$\begin{aligned} l_{RZIP} &= \sum_{Y_i=0} \log\left[\pi + (1-\pi)\exp(-\exp(\sum_j X_{ij}\beta_j))\right] + \sum_{Y_i\neq 0} \log(1-\pi) \\ &+ \sum_{Y_i\neq 0}\left[y_i \sum_j X_{ij}\beta_j - \exp(\sum_j X_{ij}\beta_j)\right] - \sum_{Y_i\neq 0} \log(y_i!). \end{aligned} \tag{2.16}$$

The LASSO type penalty function is:

$$p_{RZIP,U} = \lambda_1 \|\beta\|_1. \tag{2.17}$$

Then (2.15) can be written as,

$$\{\hat{\beta}, \hat{\pi}\} = \arg\max_{\beta,\pi} l_{RZIP} - np_{RZIP,U}. \tag{2.18}$$

2.2.5.2 Varying zero state probability If the probability that an outcome is from the zero state varies for observations, we fit a logistic regression model to estimate the zero state probability. Let $\log \mu_i = \beta_0 + \sum_j X_{ij}\beta_j$, $\log[\pi_i/1-\pi_i)] = \gamma_0 + \sum_k X_{ik}\gamma_k$ and $X_{i0} = 1$, then the log-likelihood function l_{RZIP} for is:

$$\begin{aligned} l_{RZIP} &= \sum_{Y_i=0} \log(\exp(\sum_k X_{ik}\gamma_k) + \exp(-\exp(\sum_j X_{ij}\beta_j))) \\ &+ \sum_{Y_i\neq 0}(y_i \sum_j X_{ij}\beta_j - \exp(\sum_j X_{ij}\beta_j)) \\ &- \sum_{Y_i\neq 0} \log(1+\exp(\sum_k X_{ik}\gamma_k)) - \sum_{Y_i\neq 0} \log(y_i!) \end{aligned} \tag{2.19}$$

Let $\beta = (\beta_1, \beta_2, ...\beta_p)^T$ and $\gamma = (\gamma_1, \gamma_2, ..., \gamma_p)^T$, then the LASSO type penalty function is:

$$p_{RZIP,D} = \lambda_1\|\beta\|_1 + \lambda_2\|\gamma\|_1. \tag{2.20}$$

Thus (2.15) can be written as

$$\{\hat{\beta}, \hat{\gamma}\} = \arg\max_{\beta,\gamma} l_{RZIP} - np_{RZIP,D}. \tag{2.21}$$

2.2.6 LASSO regression without zero-inflated observations

In Section 2.2.5, we can detect the zero state observations by estimating the zero state indicator Z_i. In this section, assuming we have access to the latent indicator Z_i, we are able to drop the zero state observations and then all the remaining observations are the Poisson state observations. We analyze the Poisson state data with the models in Section 2.2.1.1, 2.2.1.2, 2.2.3 and 2.2.4.

2.3 SIMULATION STUDIES

We mimic the data from motivating study in Bertocci et al. (2016) and simulate ZIP data from two settings. We call them transformed RZIP data and simulated RZIP data. For both simulated data sets, we apply nine different approaches from five models and compare the support and Poisson log-likelihood. To evaluate the variable selection performance of these models, we compare the ROC curve of LASSO regression for Poisson, LASSO regression for Gaussian, LASSO Bias correction and adaptive LASSO model.

In both settings, $n = 80$ observations are generated from a zero inflated Poisson model and $p = 107$ covariates are specified. The zero state probability $\pi = 0.2$ for all observations. X_{ij} is generated from the standard normal distribution. We use non-zero intercept in the Poisson regression model. A brief discussion about the intercept term in the Poisson model is in Section 4.3.

2.3.1 Transformed RZIP data

For transformed RZIP data, the outcomes are generated from a zero-inflated Poisson model. The zero state probability is set as 0.2 for all observations. The Poisson state mean μ_i for each observation Y_i is transformed from the corresponding Gaussian mean η_i by transformation $\mathcal{T}$ such that $\mathcal{T}(\eta_i) = \mu_i$.

We set 107 predictors and specify the Gaussian mean η_i for each observation as:

$$\eta_i = \beta_0 + \sum_j X_{ij}\beta_j, \qquad j = 1, 2, ...107. \tag{2.22}$$

We fix $\beta_0 = 4.5$, $\beta_1 = \beta_2 = ... = \beta_{16} = 0.4$, $\beta_{17} = \beta_{18} = ... = \beta_{107} = 0$, and generate X_{ij} independently from the standard normal distribution.

2.3.1.1 Support and Poisson log-likelihood

We applied seven different approaches to analyze the transformed RZIP data. For those approaches, the number of predictors selected were quite different. In the following table (Table 1), we summarize the support and Poisson log-likelihood for the different methods.

Table 1: Summary of support and Poisson log-likelihood

Method	**Support**	**Poisson log-likelihood**
Truth	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	-3.2025
LASSO regression for Poisson	0, 4, 6, 8, 13, 14, 16, 17, 18, 27, 38, 50, 51, 68, 70, 72, 87, 100, 102, 107	-2.5452
LASSO regression for Gaussian	0, 14, 38, 51, 70, 102	-2.8644
Debiased LASSO	1, 14, 38, 51, 70, 102	-2.8525
Adaptive LASSO	0, 1, 7, 14, 17, 51, 87	-3.0286
Non-inflated Truth	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	-2.0844
Zero-inflated Poisson	4, 6, 8, 10, 13, 14, 18, 27, 38, 66, 68, 70, 72, 77, 94, 98	
Non-inflated LASSO (Poisson)	0, 1, 4, 5, 6, 7, 8, 10, 12, 13, 14, 16, 17, 18, 27, 34, 35, 36, 38, 39, 42, 43, 46, 47, 50, 54, 58, 59, 61, 65, 66, 68, 69, 70, 71, 72, 76, 77, 79, 80, 81, 83, 89, 90, 91, 92, 95, 98, 99, 100, 103, 104, 105, 107	-1.7008

2.3.1.2 ROC curve In Section 2.2.1.2, Section 2.2.1.1 and Section 2.2.6, with different values of tuning parameter λ, the supports we got were different. In Section 2.2.4 and Section 2.2.6, with different values of Type-I error α, the debiased LASSO supports were also different. In Section 2.2.3 and Section 2.2.6, with different γ, the adaptive LASSO support we got were different.

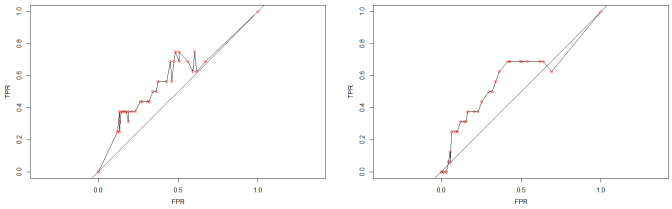
In the analysis above, we use 10-fold cross-validation method to set up the values of these tuning parameters. The object function for tuning is the "deviance", which uses squared-error for Gaussian and deviance for Poisson regression model. We choose the largest value of lambda such that error is within one standard error of the minimum cross-validated error.

To evaluate the variable selection performance of these models, we compared the ROC curve for LASSO regression for Poisson, LASSO regression for Gaussian, debiased LASSO and adaptive LASSO. In LASSO regression for Poisson model and LASSO regression for Gaussian model, for the different values of tuning parameter λ, we compared the True Positive Rate (TPR, sensitivity) and False Positive Rate (FPR, 1-specificity), where

$$TPR = \frac{\#Correctly\ selected\ predictors}{\#True\ predictors} \tag{2.23}$$

$$FPR = \frac{\#Incorrectly\ selected\ predictors}{\#False\ predictors} \tag{2.24}$$

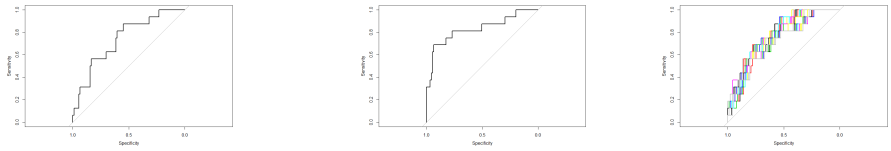
For each λ value, we took 20 replications and use the mean of 20 (FPR, TPR) to generate ROC curve (Figure 1).



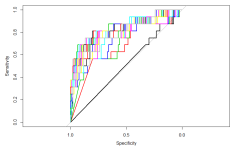
(a) Poisson, Whole data (b) Gaussian, Whole data

Figure 1: ROC Curve for LASSO Regression

In debiased LASSO, for the different values of Type-I error α level, we took 20 replications and use the mean p-values for each predictor, and then calculated (FPR, TPR) to generate the ROC curve (Figure 2).



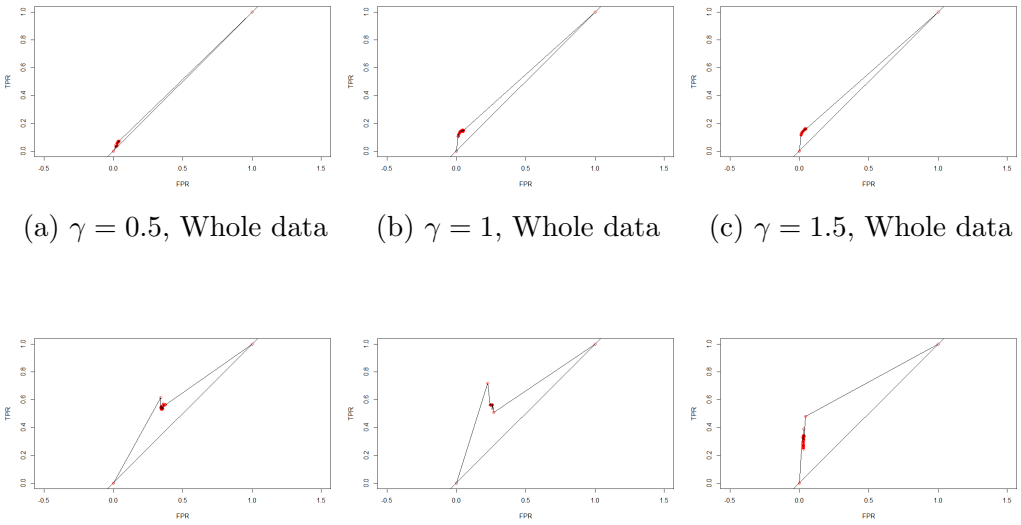
(a) Whole Data (b) Non-inflated Data (c) Whole Data



(d) Non-inflated Data

Figure 2: ROC Curve for Bias Correction

In debiased LASSO, for the different initial LASSO tuning parameters, the debiased LASSO support may be different. We use all the observations in the whole data and use only the estimated Poisson state observations in the non-inflated data. We generated different ROC curve for 40 different initial LASSO tuning parameters and selected the ROC curve with the largest area under curve for whole data and non-inflated data respectively. In adaptive LASSO, for the different γ, we took 20 replications and use the mean of 20 (FPR, TPR) to generate ROC curve (Figure 3).



(a) $\gamma = 0.5$, Whole data (b) $\gamma = 1$, Whole data (c) $\gamma = 1.5$, Whole data

(d) $\gamma = 0.5$, Non-inflated (e) $\gamma = 1$, Non-inflated (f) $\gamma = 1.5$, Non-inflated

Figure 3: ROC Curve for Adaptive LASSO Regression

Remark 2.3.1. *For both whole data and non-inflated data, the performance of LASSO regression with Poisson and Gaussian are similar. For all four methods, the non-inflated data model outperform the whole data model.*

2.3.2 Simulated RZIP data

For the simulated RZIP data, the outcomes are generated from the zero-inflated Poisson model. The zero state probability is set as 0.2 for all observations. We set 107 predictors and specify the Poisson mean μ_i for each observation as:

$$\log \mu_i = \beta_0 + \sum_j X_{ij}\beta_j, \qquad j = 1, 2, ...107. \tag{2.25}$$

We fix $\beta_0 = 1.7$, $\beta_1 = \beta_2 = ... = \beta_{16} = 0.2$, $\beta_{17} = \beta_{18} = ... = \beta_{107} = 0$, X_{ij} is generated from the standard normal distribution.

2.3.2.1 Support and Poisson log-likelihood We applied seven different approaches to analyze the simulated RZIP data. For those approaches, the number of predictors selected were quite different. In Table 2, we summarize the support and Poisson log-likelihood for the different approaches.

Table 2: Summary of support and Poisson log-likelihood

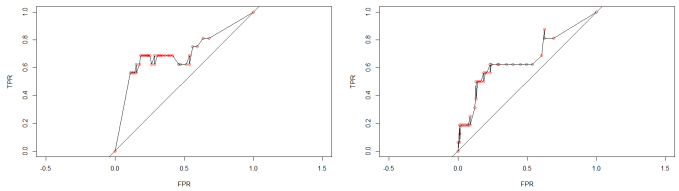
Method	**Support**	**Poisson log-likelihood**
Truth	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	-3.1668

Continued on Next Page...

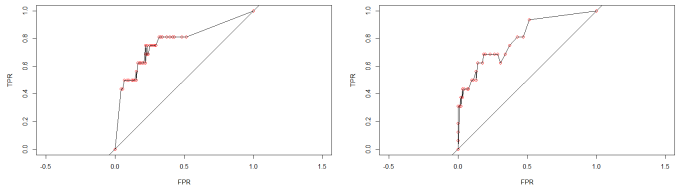
Table 2 – Continued

Method	Support	Poisson log-likelihood
LASSO regression for Poisson	0, 2, 4, 5, 6, 7, 8, 10, 13, 15, 16, 19, 20, 30, 34, 36, 37, 45, 48, 51, 58, 65, 66, 67, 72, 77, 79, 91, 96, 97, 99, 102, 106	-2.2665
LASSO regression for Gaussian	0, 2, 5, 6, 7, 8, 13, 15, 16, 19, 20, 23, 34, 36, 45, 48, 58, 65, 66, 72, 77, 79, 91, 97	-2.3379
Debiased LASSO	1, 6, 7, 16, 17, 20, 37, 67	-4.1391
Non-inflated true model	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16	-2.3739
Zero-inflated Poisson	0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 15, 16, 19, 20, 22, 34, 36, 41, 42, 43, 48, 49, 66, 68, 79, 91, 96, 97, 99, 100, 102, 106	
Non-inflated LASSO for Poisson	0, 1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 15, 16, 19, 20, 22, 25, 27, 33, 34, 36, 41, 42, 43, 48, 49, 63, 66, 75, 80, 88, 91, 94, 96, 97, 99, 100, 102, 106	-1.8987

2.3.2.2 ROC curve Keeping the same notation in Section 2.3.1, for each λ value, we take 20 replications and use the mean of 20 (FPR, TPR) to generate ROC curve (Figure 4).



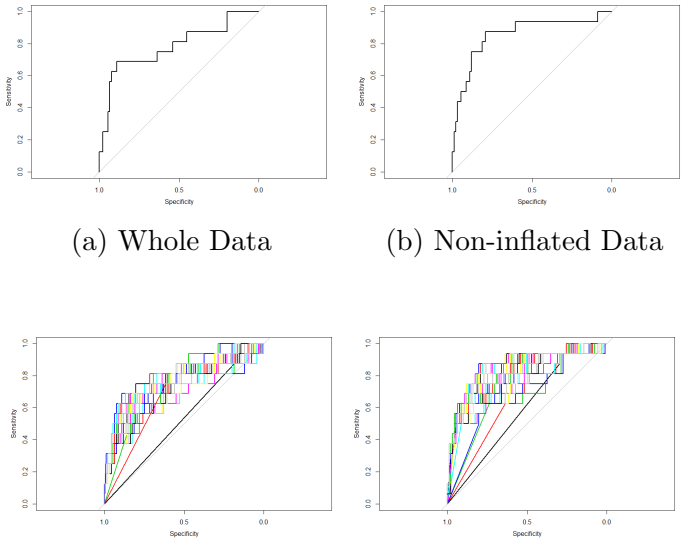
(a) Poisson, Whole data (b) Gaussian, Whole data



(c) Poisson, Non-inflated (d) Gaussian, Non-inflated

Figure 4: ROC Curve for LASSO Regression

In debiased LASSO, for the different values of Type-I error α level, we took 20 replications and used the mean p-values for each predictor, and then calculated (FPR, TPR) to generate ROC curve (Figure 5). In debiased LASSO, for the different initial LASSO tuning parameters, the debiased supports may be different. We use all the observations in the whole data and use only the estimated Poisson state observations in the non-inflated data. We generated different ROC curve for 40 different initial LASSO tuning parameters and selected the ROC curve with the largest area under curve for whole data and non-inflated data respectively.



(a) Whole Data (b) Non-inflated Data

(c) Whole Data (d) Non-inflated Data

Figure 5: ROC Curve for Debiased LASSO

Remark 2.3.2. *For both whole data and non-inflated data, the performance of LASSO regression with Poisson and Gaussian are similar. For all three methods, the non-inflated data model outperform the whole data model.*

3.0 ANALYSIS OF DATA FROM MOTIVATING PROBLEM

3.1 REVIEW OF DATA IN MOTIVATING PROBLEM

In Bertocci et al. (2016), the measures of neural structure and function along with clinical, demographic, genetic and environmental factors of 80 youth were collected in the Longitudinal Assessment of Manic Symptoms (LAMS) study. The aim of the LAMS study was to identify measures of neural function and structure predicting future behavioral and emotional dysregulation in a large group of youth. The severity of future behavioral and emotional dysregulation was measured by the Parent General Behavior Inventory-10 Item Mania Scale (PGBI-10M)). In the study, PGBI-10M scores were obtained on or near the day of scan (TIME1) and at follow-up interviews after neuroimaging scans (TIME2). Linear regression model using the LASSO method for variable selection was used in the data analysis. In the regression model, TIME2:PGBI-10 is the outcome variable, TIME1:PGBI-10 and other TIME1 clinical and demographic variables serve as predictor variables. There are 107 predictors, including TIME1:PGBI-10, in the regression model. TIME1 measures included the blood-oxygen-level-dependent, functional connectivity and diffusion imaging (DI) neuroimaging measures, Mania Rating Scale (KMRS), Depression Rating Scale (KDRS) and diagnoses (attention deficit hyperactivity disorder

(ADHD), bipolar spectrum disorder, major depressive disorder, disruptive behavior disorder, anxiety disorder), age, IQ, sex, medication status (taking versus not taking each psychotropic medication class: stimulant, non-stimulant ADHD, mood stabilizer, antipsychotic and antidepressant psychotropic medications), scan site and days between TIME1:PGBI-10M and TIME2:PGBI-10M. The predictors are standardized before the analysis.

The outcome variable, TIME2:PGBI-10, is of count type. There are 80 subjects in the study and 21 out of 80 outcomes are zero. We use a Poisson distribution to build the regression model for variable selection. For convenience, the following histogram (Figure 6) shows the marginal distribution of the outcome variable.

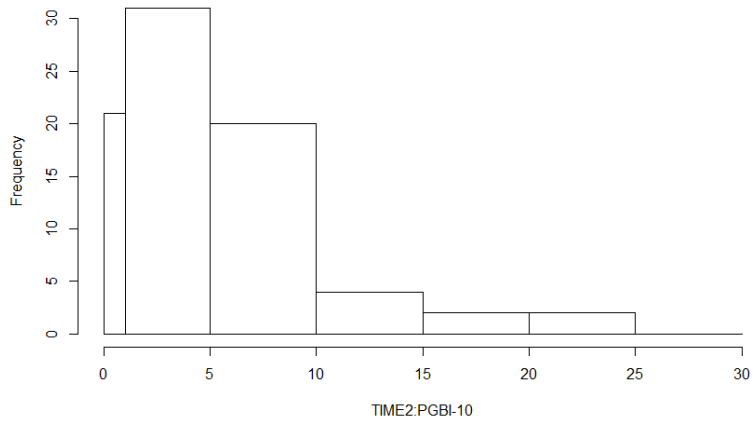


Figure 6: Histogram of TIME2:PGBI-10

3.2 LASSO REGRESSION FOR POISSON MODEL

Because there are more predictors than observations, in the analysis we use a LASSO penalty term to control the number of predictors.

We replicated the results in Bertocci et al. (2016), in the LASSO regression for Poisson model with the following covariate indexes selected: 0, 1, 5, 6, 9, 62, 63, 92. Here index 0 means the intercept term.

Unlike the Gaussian model, the Poisson model uses a non-linear log link function, hence the intercept term cannot be eliminated simply by standardising the response variable Y. The intercept term can be considered a measure of background contamination (see Hunt et al. (2019) for details). A brief discussion of the intercept term in the Poisson regression model is in Section 4.3. To study the influence of the intercept term in the Poisson model, we also analyze the LASSO regression model without an intercept term β_0:

$$\log(\mu_i) = \sum_j X_{ij}\beta_j \tag{3.1}$$

$$\hat{\beta} = \arg\min_{\beta} -\frac{1}{n}\sum_i (Y_i \log(\mu_i) - \mu_i) + \lambda\|\beta\|_1. \tag{3.2}$$

Without the intercept term β_0, only predictor 1 is selected.

3.3 LASSO REGRESSION FOR GAUSSIAN MODEL

3.3.1 Gaussian model

With the Anscombe transformed response variable Y^*, in the LASSO regression for Gaussian model, ten predictors with the following indexes are selected: 0, 1,

5, 6, 9, 29, 44, 62, 71, 92. Figure 7 shows the residual Q-Q plot: Compared to the

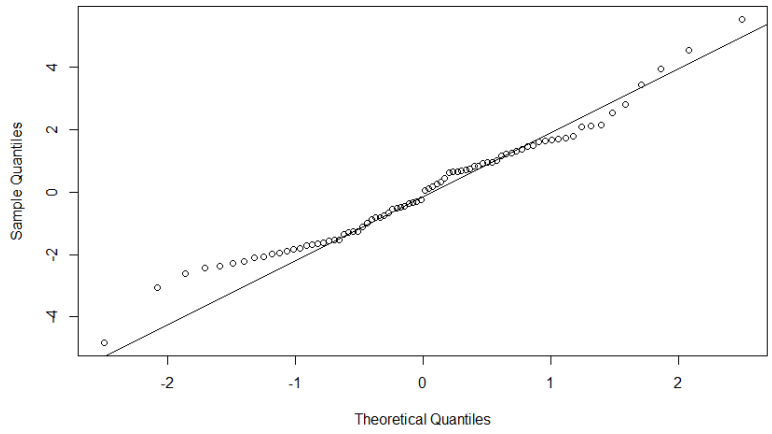


Figure 7: Transformed LASSO Q-Q Plot

Poisson model, this Gaussian model has the similar number of support variables, and both models select variables 0, 1, 5, 6, 9, 62, 92, which shows that the Anscombe transform is reasonable in dealing with this data set. However, based on the Q-Q plot, the distribution of residuals indicate that further adjustments are required for this approach, especially in the tails.

Remark 3.3.1 (without intercept term)**.** *We also analyze the LASSO regression model without intercept term* β_0*, i.e.* $\hat{\beta} = \arg\min_\beta \frac{1}{n}\|Y^* - X\beta\|_2^2 + \lambda\|\beta\|_1$ *However, without the intercept term* β_0*, no predictors were selected.*

3.3.2 Bias correction of LASSO regression for Gaussian model

With the LASSO initializer, seven predictors with the following indexes are selected: 1, 2, 42, 51, 71, 76, 92. Table 3 shows the predictors with the smallest 20

p-values.

Table 3: Bias correction of LASSO supports

LASSO	**Coef**	**Upper**	**Lower**	**Sig Ind**	***p*-value**	**Index**
Zclosest_ESM	4.1144	4.5170	3.7117	1	<0.0001	1
Zfolldays	0.7184	1.1147	0.3220	1	0.0004	2
ZKMRS_score	0.3248	0.7664	-0.1168	0	0.1494	5
ZKDRS_score	0.368	0.7603	-0.0238	0	0.0656	6
Anx	-0.3616	0.1246	-0.8477	0	0.1449	12
Zrcst_RD_adj	-0.3951	0.0586	-0.8487	0	0.0879	29
Zlcab_ICV	-0.6989	-0.0584	-1.3393	1	0.0325	42
Zlcst_ICV	0.4605	0.9831	-0.0621	0	0.0841	46
Zlilf_ICV	0.5641	1.2903	-0.1621	0	0.1279	48
Zrslfp_ICV	-0.5360	-0.0182	-1.0538	1	0.0425	51
Zrslft_ICV	0.5836	1.2120	-0.0449	0	0.0688	53
Zrccg_length	-0.3912	0.1080	-0.8904	0	0.1246	63
Zlslft_length	0.4725	1.0511	-0.1061	0	0.1095	70
Zrslft_length	0.4602	0.9003	0.0201	1	0.0404	71
ZFmin_L1_adj	-0.5529	0.1469	-1.2527	0	0.1215	75
Zlatr_L1_adj	0.7010	1.3413	0.0606	1	0.0319	76
Zlslfp_L1_adj	-0.3756	0.1472	-0.8985	0	0.1591	86
Zvs484652	1.0338	2.0283	0.0393	1	0.0416	92
ZBA40rtPar	-0.4156	0.1105	-0.9417	0	0.1215	96
ZBA45lftIFG	-0.4148	0.1331	-0.9628	0	0.1379	103

Compared to the support in Section 2.2.1.2, the variables selected in this section are quite different. However, if we consider variables with the smallest 20 p-values, most of the variables from Section 2.2.1.2 are included. Also, the p-values for the overlapping variables in these 20 variables are all smaller than to 0.15. Table 4 shows the missed variables from Section 2.2.1.2. The heat map in Figure 8 shows the covariance of the 20 variables and the 2 missed variables. The brighter color on the heat map indicates the stronger correlation between two variables. From the heat map, variable 42, 46, 48, 51, and 53, variable 75, 76 and 86 are correlated, which can also explain why only one of these variables in each group is significant.

Table 4: Missed variables in bias correction for LASSO model

LASSO	**Coefficient**	**Upper**	**Lower**	**Sig Ind**	p**-value**	**Index**
sex	0.3304	0.9500	-0.2893	0	0.2961	9
Zlccg_length	-0.2765	0.2701	-0.8230	0	0.3215	62

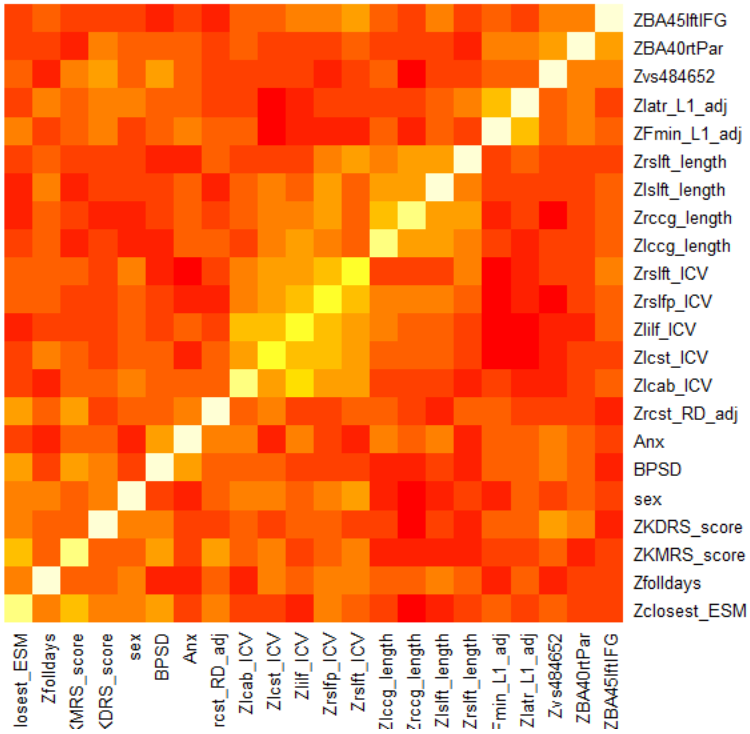


Figure 8: De-biased LASSO Heat Map

Remark 3.3.2. *In Section 2.2.1.1 and Section 2.2.4, the value of the tuning parameter λ in equation (2.1) is selected via 10-fold cross-validation. In Section 2.2.4, Type-I error α is set as 0.05. Table 5 shows the Poisson log-likelihood for the different tuning parameters and different values of the Type-I error. The bold numbers are the maximum for each column.*

Table 5: Poisson log-likelihood with different λ and α

		Bias correction					
λ	**Initial**	α**=0.05**	α**=0.06**	α**=0.07**	α**=0.08**	α**=0.09**	α**=0.10**
0.3000	-2.8350	-3.0547	-3.0156	-2.8022	-2.8358	-2.8442	-2.7513
0.3205	**-2.8307**	-3.0547	-3.0156	-2.8022	-2.9215	-2.9005	-2.7513
0.3410	**-2.8307**	-3.0547	-3.0156	-2.9190	-2.9215	-2.9005	-2.7513
0.3615	-2.8548	-3.0547	-3.0156	-2.7842	-2.9215	-3.0560	-2.7513
0.3821	-2.8548	-3.1845	-3.0156	**-2.6816**	-2.9215	-3.0560	-2.7256
0.4026	-2.9433	**-2.9645**	**-2.8916**	**-2.6816**	-2.9215	-3.0560	-2.7256
0.4231	-2.9433	-2.9749	**-2.8916**	-2.8358	-2.7322	-2.7981	**-2.5851**
0.4436	-2.9967	-3.0570	**-2.8916**	-2.8877	-2.7322	-2.6751	**-2.5851**
0.4641	-2.9967	-3.0570	**-2.8916**	-2.7732	**-2.6046**	-2.6751	**-2.5851**
0.4846	-2.9967	-3.0570	-3.0308	-2.7732	**-2.6046**	-2.6751	**-2.5851**
0.5051	-3.0978	-3.0570	-2.9272	-2.7732	**-2.6046**	-2.6751	**-2.5851**
0.5256	-3.0978	-3.0527	-2.9272	-2.7475	-2.6959	-2.6751	**-2.5851**
0.5462	-3.0952	-3.0681	-2.9272	-2.7475	-2.6959	-2.6751	-2.7481
0.5667	-3.0952	-3.0681	-2.9272	-2.7475	-2.6897	-2.6751	-2.7481
0.5872	-3.1414	-3.0681	-2.9272	-2.7475	-2.6897	-2.7766	-2.7495
0.6077	-3.4471	-3.0681	-2.9272	-2.7475	-2.6897	-2.7766	-2.7495

3.3.3 Regularized zero-inflated Poisson regression model

3.3.3.1 Uniform zero state probability Let $\log \mu_i = \beta_0 + \sum_j X_{ij}\beta_j$, then the log-likelihood function l_{ZIP} is

$$\begin{aligned} l_{RZIP} &= \sum_{Y_i=0} \log \left[\pi + (1-\pi) \exp(-\exp(\sum_j X_{ij}\beta_j)) \right] + \sum_{Y_i \neq 0} \log 1 - \pi \\ &+ \sum_{Y_i \neq 0} \left[y_i \sum_j X_{ij}\beta_j - \exp(\sum_j X_{ij}\beta_j) \right] - \sum_i \log(y_i!) \end{aligned}$$

Analysis Results We apply the EM algorithm proposed in Wang et al. (2014b) and select 20 variables for the Poisson state (Table 6).

Table 6: RZIP with uniform π model supports

RZIP Poisson	Coef	Index	RZIP Poisson	Coef	Index
(Intercept)	1.6156	0	X1Zlccg_length	-0.2194	62
X1Zclosest_ESM	0.2671	1	X1Zrccg_length	-0.0816	63
X1ZAge_At_Scan	0.0246	3	X1Zlslfp_length	-0.0666	68
X1ZBase_IQ	-0.0738	4	X1ZFmin_L1_adj	-0.0143	75
X1ZKMRS_score	0.0310	5	X1Zlslfp_L1_adj	-0.0640	86
X1Anx	-0.0463	12	X1Zrslfp_L1_adj	-0.0060	87
X1antidepressant	-0.0363	15	X1Zrunc_L1_adj	-0.0419	91
X1Zlcab_ICV	-0.1767	42	X1Zvs32658	0.1157	95
X1Zrslft_ICV	0.0755	53	X1ZBA40lftPar	0.0916	99
X1Zratr_length	-0.1153	59	X1ZCorpCal	-0.0356	101
π	**0.2381**				

For each observation, Z_i is the zero state indicator. There are 21 zero observations in the outcome variable. Based on estimated Z_i (Table 7), 20 of those observations are from the zero state. The bold observation is the one that is **not** from the zero state.

Table 7: Zero state vs. Poisson state for zero realization

Obs	Z_i	π_i	μ_i	$(1-\pi_i)e^{-\mu_i}$	**Obs**	Z_i	π_i	μ_i	$(1-\pi_i)e^{-\mu_i}$
2	1	0.2381	15.5077	0.0000	41	1	0.2381	5.1663	0.0043
3	1	0.2381	5.6993	0.0026	53	1	0.2381	2.8517	0.0440
7	1	0.2381	1.7669	0.1302	59	1	0.2381	6.6310	0.0010
8	1	0.2381	2.4949	0.0629	61	1	0.2381	8.0007	0.0003
9	1	0.2381	2.8625	0.0435	63	1	0.2381	4.3810	0.0095
16	1	0.2381	3.2294	0.0302	**64**	**0**	**0.2381**	**1.1380**	**0.2442**
18	1	0.2381	6.4322	0.0012	69	1	0.2381	6.7047	0.0009
29	1	0.2381	6.3400	0.0013	73	1	0.2381	2.7682	0.0478
34	1	0.2381	5.3038	0.0038	75	1	0.2381	6.7352	0.0009
37	1	0.2381	2.9922	0.0382	79	1	0.2381	4.2438	0.0109
39	1	0.2381	5.2263	0.0041					

3.3.3.2 Varying zero state probability

Let $\log \mu_i = \beta_0 + \sum_k X_{ij}\beta_j$, $\log(\pi_i/1-\pi_i) = \gamma_0 + \sum_k X_{ik}\gamma_k$ and $X_{i0} = 1$, then the log-likelihood function l_{RZIP} is:

$$\begin{aligned} l_{RZIP} &= \sum_{Y_i=0} \log \left[\exp(\sum_k X_{ik}\gamma_k) + \exp(-\exp(\sum_j X_{ij}\beta_j)) \right] \\ &\quad + \sum_{Y_i \neq 0} \left[y_i \sum_j X_{ij}\beta_j - \exp(\sum_j X_{ij}\beta_j) \right] \\ &\quad - \sum_{Y_i \neq 0} \log \left[1 + \exp(\sum_k X_{ik}\gamma_k) \right] - \sum_{Y_i \neq 0} \log(y_i!) \end{aligned} \tag{3.3}$$

We apply the EM algorithm proposed in Wang et al. (2014b). 20 variables for the Poisson state and 4 variables for the zero state are selected (Table 8).

Table 8: RZIP model support

RZIP Poisson	Coef	Index	RZIP Poisson	Coef	Index
Intercept	1.6220	0	**X1Zlccg_length**	-0.2241	62
X1Zclosest_ESM	0.2609	1	**X1Zrccg_length**	-0.0840	63
X1ZAge_At_Scan	0.0279	3	**X1Zlslfp_length**	-0.0652	68
X1ZBase_IQ	-0.0706	4	**X1ZFmin_L1_adj**	-0.0047	75
X1ZKMRS_score	0.0288	5	**X1Zlslfp_L1_adj**	-0.0675	86
X1Anx	-0.0492	12	**X1Zrslfp_L1_adj**	-0.0075	87
X1antidepressant	-0.0338	15	**X1Zrunc_L1_adj**	-0.0405	91
X1Zlcab_ICV	-0.1731	42	**X1Zrunc_L1_adj**	-0.0405	91
X1Zrslft_ICV	0.0737	53	**X1ZBA40lftPar**	0.0962	99
X1Zratr_length	-0.1106	59	**X1ZCorpCal**	-0.0396	101
RZIP Zero	**Coef**	**Index**	**RZIP Zero**	**Coef**	**Index**
Intercept	-1.1675	0	**X1BPSD**	-0.0827	11
X1Zclosest_ESM	-0.0535	1	**X1Zrslft_length**	-0.3110	71

For each observation, Z_i is the zero state indicator. There are 21 zero observations in the outcome variable. Based on estimated Z_i (Table 9), 20 of those observations are from the zero state. The bold observation is the one that is **not** from the zero state. The average value of π_i for the 21 observations is 0.2844.

Table 9: Zero state vs. Poisson state for zero realization

Obs	Z_i	π_i	μ_i	$(1-\pi_i)e^{-\mu_i}$	**Obs**	Z_i	π_i	μ_i	$(1-\pi_i)e^{-\mu_i}$
2	1	0.2087	14.9644	0.0000	41	1	0.1950	5.2536	0.0042
3	1	0.2431	5.8048	0.0023	53	1	0.2648	2.9345	0.0391
7	1	0.2070	1.7320	0.1403	59	1	0.3194	6.6985	0.0008
8	1	0.3937	2.6470	0.0430	61	1	0.2611	8.1030	0.0002
9	1	0.2798	2.8780	0.0405	63	1	0.2812	4.5047	0.0079
16	1	0.2456	3.2786	0.0284	**64**	**0**	**0.2463**	**1.1019**	**0.2504**
18	1	0.2681	6.4601	0.0011	69	1	0.2104	6.8943	0.0008
29	1	0.3368	6.3961	0.0011	73	1	0.3816	2.8357	0.0363
34	1	0.4500	5.3251	0.0027	75	1	0.3419	6.7849	0.0007
37	1	0.2770	3.0494	0.0343	79	1	0.3009	4.3391	0.0091
39	1	0.2601	5.4070	0.0033					

Compared to the support from Section 2.2.1.2, the Poisson state support included more predictors. However, variable 6, 9, and 92 were still not included. For both uniform π model and different π_i model, only observation No. 64 is from the Poisson state, and all the remaining zero observations are from the zero state. The heat map (Figure 9) of these 20 variables and Section 2.2.1.2 is as below. From the heat map, variable 6, 92 and 95, variable 6 and 42 are correlated, which can also explain why

variable 6, 9, and 92 were not included.

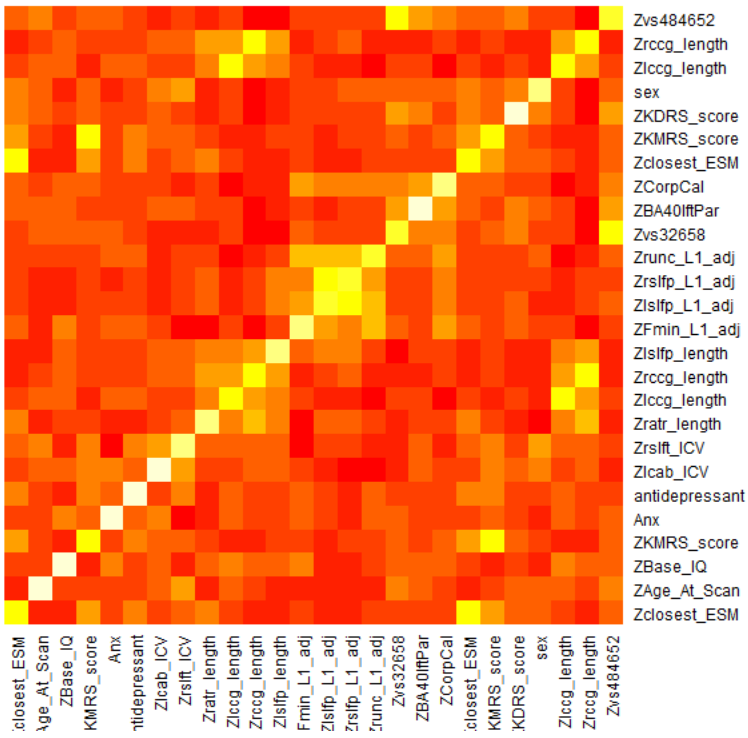


Figure 9: RZIP Support Heat Map

3.3.4 LASSO regression without zero-inflated observations

In Section 2.2.5, we detect the zero state observations by estimating Z_i. In this section, we drop the zero state observations and only use observations from the Poisson state. In the new data set, there are 60 observations and 107 predictors. The predictors are standardized before analysis. Figure 10 shows the histogram of the Poisson state outcome. We then analyze the new data with the methods in Section 2.2.1.1 and Section 2.2.1.2.

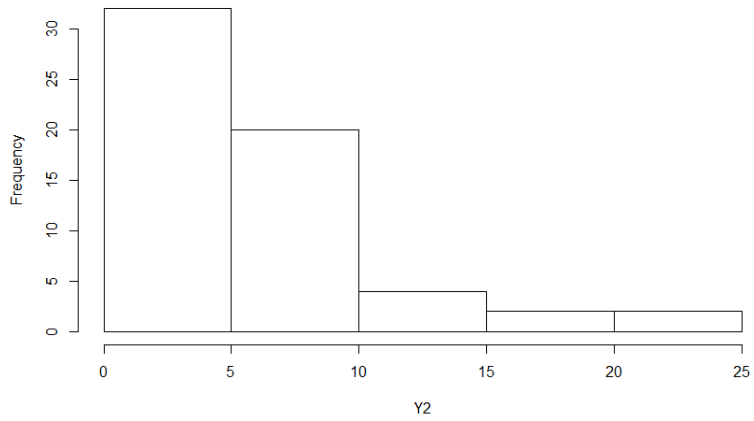


Figure 10: Histogram of Poisson State Outcome

3.3.4.1 Poisson model We repeat the analysis in Section 2.2.1.2 and 16 predictors with the following indexes selected: 0, 1, 3, 4, 5, 12, 42, 53, 59, 62, 63, 68, 86, 91, 95, 99.

3.3.4.2 Gaussian model We repeat the analysis in Section 2.2.1.1 and 14 predictors with the following indexes selected: 0, 1, 3, 4, 9, 12, 42, 59, 62, 63, 68, 86, 95, 99.

3.3.4.3 Bias correction of Gaussian model We repeat the analysis in Section 2.2.4 and 8 predictors with the following indexes are selected: 1, 2, 42, 53, 63, 75, 97, 100. Table 10 shows the predictors with the smallest 20 p-values.

Remark 3.3.3. *In Section 2.2.4, we provide the Poisson log-likelihood for the differ-*

ent tuning parameters and different values of Type-I error. Table 11 shows the Poisson log-likelihood for the different tuning parameters and different values of Type-I error for the non-inflated case. The bold numbers are the maximum in each column.

Table 10: Non-inflated bias correction of LASSO supports

LASSO	**Coef**	**Upper**	**Lower**	**Sig Ind**	***p*-value**	**Index**
Zclosest_ESM	4.9170	5.2301	4.6040	1	0.0000	1
Zfolldays	0.7796	1.1058	0.4534	1	0.0000	2
ZAge_At_Scan	0.2262	0.5487	-0.0962	0	0.1690	3
sex	0.3527	0.8062	-0.1008	0	0.1275	9
Anx	-0.3738	0.1456	-0.8932	0	0.1584	12
Zrcst_RD_adj	-0.2618	0.1166	-0.6402	0	0.1751	29
Zlilf_RD_adj	0.4167	0.9822	-0.1488	0	0.1487	30
ZFM_ICV	-0.2925	0.1667	-0.7517	0	0.2118	38
Zlcab_ICV	-0.4422	-0.0105	-0.8738	1	0.0447	42
Zrslft_ICV	0.4678	0.8722	0.0635	1	0.0234	53
Zrccg_length	-0.6053	-0.2048	-1.0059	1	0.0031	63
ZFmin_L1_adj	-0.4071	-0.0141	-0.8001	1	0.0423	75
Zlslfp_L1_adj	-0.2595	0.1368	-0.6558	0	0.1994	86
Zrslfp_L1_adj	-0.4597	0.0285	-0.9479	0	0.0649	87
Zrunc_L1_adj	-0.3732	0.1390	-0.8855	0	0.1533	91
Zvs32658	0.7025	1.4921	-0.0871	0	0.0812	95
ZBA8rtPFC	-0.4387	-0.0153	-0.8620	1	0.0423	97

Continued on Next Page...

Table 10 – Continued

LASSO	**Coef**	**Upper**	**Lower**	**Sig Ind**	p**-value**	**Index**
ZBA40lftPar	0.2338	0.5937	-0.1261	0	0.2030	99
ZBA6lftmotot	0.4330	0.8243	0.0418	1	0.0301	100
ZCorpCal	-0.3211	0.0181	-0.6603	0	0.0636	101

Table 11: Poisson log-likelihood with different λ and α

		Bias correction					
λ	**Initial**	α**=0.05**	α**=0.06**	α**=0.07**	α**=0.08**	α**=0.09**	α**=0.10**
0.1000	**-1.9222**	-2.7818	-2.7302	-2.6908	-2.6629	-2.6629	-2.5959
0.1256	-2.0013	-2.8589	-2.8589	-2.7470	-2.6930	-2.6930	-2.6930
0.1513	-2.1135	-2.8589	-2.8060	-2.7469	-2.6930	-2.6819	-2.6930
0.1769	-2.1843	-2.8589	-2.8060	-2.7469	-2.5867	-2.6819	-2.5810
0.2026	-2.2609	-2.6531	-2.6792	-2.6514	-2.5867	-2.5810	-2.5867
0.2538	-2.2667	-2.6566	-2.7423	-2.6566	-2.6531	-2.5539	-2.5867
0.2795	-2.2667	**-2.6706**	-2.7423	-2.6566	-2.6566	-2.6073	-2.5922
0.3051	-2.3438	**-2.6706**	-2.7568	-2.6566	-2.6325	-2.5862	-2.6437
0.3308	-2.4184	-2.7568	-2.7568	-2.6437	-2.6437	-2.5909	-2.6437
0.3564	-2.4184	-2.7609	**-2.6780**	-2.6437	-2.6437	-2.5909	-2.5059
0.3821	-2.5236	-2.7609	**-2.6780**	-2.6437	-2.6141	-2.4737	-2.5059

Continued on Next Page...

Table 11 – Continued

λ	**Initial**	**Bias correction**					
		α=0.05	**α=0.06**	**α=0.07**	**α=0.08**	**α=0.09**	**α=0.10**
0.4077	-2.6230	-2.7609	**-2.6780**	-2.6437	-2.5920	-2.4737	-2.4561
0.4333	-2.6230	-2.7609	-2.7609	-2.6437	-2.5920	-2.3787	**-2.3726**
0.4590	-2.6520	-2.7609	-2.7609	-2.6437	**-2.5426**	**-2.3372**	**-2.3726**
0.4846	-2.7286	-2.7609	-2.7609	**-2.6423**	**-2.5426**	**-2.3372**	**-2.3726**
0.5103	-2.7590	-2.7609	-2.7609	**-2.6423**	-2.5512	**-2.3372**	**-2.3726**
0.5359	-2.7590	-2.7609	-2.7609	**-2.6423**	-2.5512	-2.3440	-2.3735
0.5615	-2.7590	-2.7225	-2.7609	**-2.6423**	-2.5954	-2.3440	-2.3778
0.5872	-2.7590	-2.7225	-2.6987	-2.6652	-2.5478	-2.4274	-2.3778
0.6000	-2.7590	-2.7225	-2.6987	-2.6294	-2.5684	-2.4274	-2.3811

3.4 SUMMARY

We apply the approaches that are described in the review and summarize the supports and Poisson log-likelihood for different approaches in Table 12.

We replicate the results in Bertocci et al. (2016), in the LASSO regression for the Poisson model to get the same covariate indexes: 0, 1, 5, 6, 9, 62, 63, 92. Here index 0 means the intercept term. For other approaches, the number of predictors selected are quite different. However, among the results, variables 1, 5, 62, 63 and 92 are selected by most approaches. In Section 2.2.1.1, Section 2.2.4 and Section 2.2.6,

we apply the Anscombe transform to the outcome variable. For the results of those three sections, we apply the transformation (2.6) and calculate the corresponding Poisson means. From the table, the non-inflated LASSO regression for Poisson model (different zero state probability) provides the largest Poisson log-likelihood.

Table 12: Summary of support and Poisson log-likelihood

Method	**Support**	**Poisson log-likelihood**
LASSO for Poisson (without intercept)	0 , 1, 5, 6, 9, 62, 63, 92 (1)	-3.1722 (-6.1754)
LASSO for Gaussian (without intercept)	0, 1, 5, 6, 9, 29, 44, 62, 71, 92 (NA)	-2.8780 (NA)
Debiased LASSO for Gaussian	1, 2, 42, 51, 71, 76, 92	-3.1217
Regularized zero-inflated Poisson regression uniform zero state	0, 1, 3, 4, 5, 12, 15, 42, 53, 59, 62, 63, 68, 75, 86, 87, 91, 95, 99, 101	NA
Regularized zero-inflated Poisson regression Poisson (zero)	0, 1, 3, 4, 5, 12, 15, 42, 53, 59, 62, 63, 68, 75, 86, 87, 91, 95, 99, 101 (0, 1, 11, 71)	NA
Non-inflated LASSO for Poisson	0, 1, 3, 4, 5, 12, 42, 53, 59 62, 63, 68, 86, 91, 95, 99	-2.4224

Continued on Next Page...

Table 12 – Continued

Method	**Support**	**Poisson log-likelihood**
(same zero state probability)		
Non-inflated LASSO for Poisson (different zero state probability)	0, 1, 3, 4, 5, 6, 7, 11, 15, 20, 25, 30, 33, 40, 42, 44, 47, 49, 56, 59, 63, 67, 68, 70, 71, 75, 84, 89, 95, 101, 104	-1.9240

4.0 BIAS CORRECTION OF GLM

In linear regression model $Y = X\beta + \epsilon$, $\beta = (\beta_1, \beta_2, ..., \beta_p)$, we have the loss function

$$l(X, Y) = \frac{1}{n}\|Y - X\beta\|_2^2 \tag{4.1}$$

which is a convex function in β with second order partial derivative existing. We define $\widehat{\Sigma} = \frac{\partial}{\partial\beta\partial\beta^T} l(X, Y)/n$. Given the regularized LASSO estimator

$$\widehat{\beta}^{init} = \arg\min_{\beta}(l(X, Y) + \lambda\|\beta\|_1), \tag{4.2}$$

the debiased LASSO estimator defined in Section 2.2.4 takes the form of

$$\widehat{\beta} = \widehat{\beta}^{init} - \widehat{\Theta}\frac{\partial}{\partial\beta} l(X, Y)|_{\widehat{\beta}^{init}} \tag{4.3}$$

where $\widehat{\Theta}$ is the approximate inverse of Σ (see details in Van de Geer et al. (2014)).

To estimate $\widehat{\Theta}$, we consider the LASSO type optimization:

$$\widehat{\gamma}_i = \arg\min_{\gamma}(\widehat{\Sigma}_{i,i} - 2\widehat{\Sigma}_{i,-i}\gamma + \gamma^T\widehat{\Sigma}_{-i,-i}\gamma + 2\lambda_i\|\gamma\|_1), \tag{4.4}$$

where $\widehat{\Sigma}_{i,-i}$ is the ith row of $\widehat{\Sigma}$ without the ith element, and $\widehat{\Sigma}_{-i,-i}$ is the sub-matrix of $\widehat{\Sigma}$ without the ith row and ith column for $i = 1, 2, ..., p$. We denote $\widehat{\tau}_i^2 = \widehat{\Sigma}_{i,i} - \widehat{\Sigma}_{i,-i}\widehat{\gamma}_i$, and $\widehat{T} = diag(\widehat{\tau}_1, \widehat{\tau}_2, ..., \widehat{\tau}_p)$,

$$\widehat{C} = \begin{pmatrix} 1 & -\widehat{\gamma}_{1,2} & \cdots & -\widehat{\gamma}_{1,p} \\ -\widehat{\gamma}_{2,1} & 1 & \cdots & -\widehat{\gamma}_{2,p} \\ \vdots & \ddots & \vdots & \vdots \\ -\widehat{\gamma}_{p,1} & -\widehat{\gamma}_{p,21} & \cdots & -\widehat{\gamma}_{p,p} \end{pmatrix}, \tag{4.5}$$

Then we can define $\widehat{\Theta}$ as

$$\widehat{\Theta} = \widehat{T}^{-1}\widehat{C}. \tag{4.6}$$

In Van de Geer et al. (2014), the debiased LASSO estimator proposed in (4.3) is extended to GLM with convex loss functions. For example, in the logistic regression model, Y follows a $Bernoulli(\pi)$ distribution with $logit(\pi) = X\beta$, $\beta = (\beta_1, \beta_2, ..., \beta_p)$. If X_i is the ith row of design matrix X, then the loss function is:

$$\begin{aligned} l(X, Y) &= \ln \prod_i \{expit(X_i\beta)^{Y_i}(1 - expit(X_i\beta)^{(1-Y_i)})\} && (4.7) \\ &= \sum_i \{Y_i \ln(X_i\beta) - \ln(1 + e^{X_i\beta})\} && (4.8) \end{aligned}$$

We can define the debiased LASSO estimator for logistic regression model by updating the function in (4.3) with the new loss function and corresponding Σ.

In the Poisson regression model, suppose Y follows a Poisson distribution with mean μ. Let $\ln(\mu) = X\beta$, $\beta = (\beta_1, \beta_2, ..., \beta_p)$, X_i be the ith row of design matrix X, then the loss function is:

$$\begin{aligned} l(X, Y) &= \sum_i Y_i \ln \mu_i - \sum_i \mu_i && (4.9) \\ &= \sum_i \{Y_i X_i\beta - e^{X_i\beta}\} && (4.10) \end{aligned}$$

The corresponding $\widehat{\Sigma}$ is:

$$\widehat{\Sigma} = \frac{\partial}{\partial\beta\partial\beta^T} l(X, Y)/n. \tag{4.11}$$

4.1 DECORRELATED SCORE METHOD

In Ning et al. (2017), a decorrelated score method is proposed for statistical inference of GLM, which extended Rao's score test to high dimensional settings. To test the hypothesis $H_0 : \beta_1 = 1$ against $H_a : \beta_1 \neq 0$, the decorrelated score function is:

$$S(\beta_1, \beta_{-1}) = \nabla_{\beta_1} l(\beta_1, \beta_{-1}) - w^T \nabla_{\beta_{-1}} l(\beta_1, \beta_{-1}) \tag{4.12}$$

with $w = \mathbf{I}^{-\mathbf{1}}_{\beta_{-1},\beta_{-1}} \mathbf{I}_{\beta_{-1},\beta_1}$ and $\beta_{-1} = (\beta_2, \beta_3, ..., \beta_p)^T$. $\mathbf{I}_{\beta_{-1},\beta_1}$, $\mathbf{I}_{\beta_{-1},\beta_{-1}}$ and $\mathbf{I}_{\beta_1,\beta_1}$ are the corresponding partitions of the Fisher Information matrix $\mathbf{I} = -\mathbb{E}_\beta(\nabla^2 l(\beta))$, i.e,

$$\mathbf{I} = \begin{pmatrix} \mathbf{I}_{\beta_1,\beta_1} & \mathbf{I}_{\beta_1,\beta_{-1}} \\ \mathbf{I}_{\beta_{-1},\beta_1} & \mathbf{I}_{\beta_{-1},\beta_{-1}} \end{pmatrix}, \tag{4.13}$$

The estimated decorrelated score function $\widehat{S}(0, \tilde{\beta}_{-1})$ is

$$\widehat{S}(0, \tilde{\beta}_{-1}) = \frac{1}{n} \sum_{i=1}^{n} \{\nabla_{\beta_1} l_i(0, \tilde{\beta}_{-1}) - \bar{w}^T \nabla_{\beta_{-1}} l_i(0, \tilde{\beta}_{-1})\}, \tag{4.14}$$

with the estimated weight $\bar{w}$ for decorrelated score function

$$\bar{w} = \arg\min_w \frac{1}{2n} \sum_{i=1}^{n} \{\nabla_{\beta_1} l_i(\tilde{\beta}) - w^T \nabla_{\beta_{-1}} l_i(\tilde{\beta})\}^2 + \lambda' \|w\|_1 \tag{4.15}$$

In the Poisson regression model, with covariate $X = (X_1, X_{-1})$ and coefficient $\beta = (\beta_1, \beta_{-1})^T$, the decorrelated score function for testing $H_0 : \beta_1 = 0$ is:

$$\widehat{S}(0, \tilde{\beta}_{-1}) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - exp(\tilde{\beta}_{-1}^T X_{-1,i}))(X_{1,i} - \bar{w}^T X_{-1,i}) \tag{4.16}$$

The test statistic U_n is

$$\widehat{U}_n \quad = \quad n^{1/2}\widehat{S}(0,\tilde{\beta}_{-1})\widehat{I}^{-1/2}_{\beta_1|\beta_{-1}} \tag{4.17}$$

where $\widehat{I}_{\beta_1|\beta_{-1}} = \frac{1}{n}\sum_{i=1}^{n}\nabla^2_{\beta_1,\beta_1}l_i(\tilde{\beta}) - \bar{w}^T\nabla^2_{\beta_{-1},\beta_1}l_i(\tilde{\beta})$. We reject H_0 if U_n is larger than the critical value z_α.

4.2 CORNISH-FISHER ADJUSTED TEST

The Cornish-Fisher expansion, first proposed in Cornish and Fisher (1938), is an asymptotic expansion used to approximate the quantiles of a probability distribution based on its first few cumulants to try to improve the Gaussian approximation in the central limit theorem. One application of the Cornish-Fisher expansion is to estimate Value at Risk (VaR). When the return of a portfolio is close to Gaussian distribution, Cornish-Fisher expansion will provide an accurate estimation of the q_{th} quantile.

Under the assumption that log-return of the portfolio X is Gaussian distributed with mean $\mu(X)$ and variance $Var(X)$, the VaR is

$$VaR = \mu(X) + \sqrt{Var(X)}z_\alpha \tag{4.18}$$

where z_α is the VaR critical value for the confidence level α. The Cornish-Fisher expansion takes the higher moments of X into account and modified the critical value q_α as

$$q_\alpha = z_\alpha + \frac{(z_\alpha^2-1)S(X)}{6} + \frac{(z_\alpha^3-3z_\alpha)K(X)}{24} - \frac{(2z_\alpha^3-5z_\alpha)S^3(X)}{36} \tag{4.19}$$

where $S(X)$ is the skewness, $K(X)$ is kurtosis of X.

In the decorrelated score test, the test statistic $\widehat{U}_n = n^{1/2}\widehat{S}(0,\tilde{\beta}_{-1})\widehat{I}^{-1/2}_{\beta_1|\beta_{-1}}$ is asymptotically Gaussian. We calculate the sample skewness $S(U)$, sample kurtosis $K(U)$:

$$S(U) = \frac{\sum_{i=1}^{n}(U_i-\bar{U})^3/n}{s^3(U)} \tag{4.20}$$

$$K(U) = \frac{\sum_{i=1}^{n}(U_i-\bar{U})^4/n}{s^4(U)} \tag{4.21}$$

where $U_i = \{\nabla_{\beta_1} l_i(0,\tilde{\beta}_{-1}) - \bar{w}^T\nabla_{\beta_{-1}} l_i(0,\tilde{\beta}_{-1})\}\widehat{I}^{-1/2}_{\beta_1|\beta_{-1}}$, $\bar{U}$ is the sample mean, $s(U)$ is the sample standard deviation. The Cornish -Fisher critical value for decorrelated score test is

$$q_\alpha = z_\alpha + \frac{(z_\alpha^2-1)S(U)}{6} + \frac{(z_\alpha^3-3z_\alpha)K(U)}{24} - \frac{(2z_\alpha^3-5z_\alpha)S^3(U)}{36} \tag{4.22}$$

In the simulation study, we also consider the Cornish-Fisher expansion with first-order skewness only, and Cornish-Fisher expansion with first-order skewness and first-order kurtosis.

4.3 INFLUENCE OF INTERCEPT TERM

For the Gaussian regression model, $\mathbb{E}Y_i = \alpha + X_i\beta$ for $i = 1, 2, ..., n$. We can remove the intercept term intercept term α by standardizing the predictors. However, For the Poisson regression model, $\ln \mathbb{E}Y_i = \alpha + X_i\beta$ for $i = 1, 2, ..., n$. The typical standardizing method cannot deal with the intercept term α well since it has influence on both the mean and variance of the observation. In Hunt et al. (2019), the intercept term was referred as background contamination and a mapping of both Y and X are required for further consideration. In the simulation study, we let $\alpha = 0$ to simplify the model.

4.4 SIMULATION STUDY

In the simulation study, we generated new data sets to investigate the performance of different models. In particular, the response variable Y is generated from a zero-inflated Poisson distribution. The sample size is $n = 200$. The covariate X is generated from a d-dimensional multivariate normal distribution $N(0, \Sigma)$, where $d = 100, 200, 500$, and Σ is a Toeplitz matrix with $\Sigma_{jk} = \rho^{|j-k|}$. The correlation parameter $\rho = 0, 0.25, 0.4, 0.6, 0.75$. To perform the hypothesis of $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$, we specified the regression coefficient $\beta = (0, \beta_s, \mathbf{0})$ with $\beta_s = (1, 1, 1)$. The intercept term is α. The probability that Y_i is from the zero state is fixed at $\pi = 0$ and 0.2. For observations from the Poisson state, we generated Y_i with

$$\ln \mathbb{E}Y_i = X_i\beta \tag{4.23}$$

To begin with, we set $\pi = 0$ so that all the observations were from the Poisson state. We then set $\pi = 0.2$, in which about 20% of all observations were from the zero state and the rest were from the Poisson state. We applied the EM algorithm in Section 2.2.5, LASSO regression for Poisson model in Section 2.2.1.2 and decorrelated score method in Section 4.1.

We consider three Cornish-Fisher adjusted score tests. In the first test, we only include the first-order skewness in the model. Keeping the notation in Section 4.2, the corresponding Cornish-Fisher critical value for decorrelated test is:

$$q_\alpha = z_\alpha + \frac{(z_\alpha^2 - 1)S(U)}{6}. \tag{4.24}$$

In the second test, we include both the first-order skewness and first-order kurtosis in the expansion. The Cornish-Fisher critical value is:

$$q_\alpha = z_\alpha + \frac{(z_\alpha^2 - 1)S(U)}{6} + \frac{(z_\alpha^3 - 3z_\alpha)K(U)}{24}. \tag{4.25}$$

In the third test, we use the Cornish-Fisher critical value defined in equation (4.22):

$$q_\alpha = z_\alpha + \frac{(z_\alpha^2 - 1)S(U)}{6} + \frac{(z_\alpha^3 - 3z_\alpha)K(U)}{24} - \frac{(2z_\alpha^3 - 5z_\alpha)S^3(U)}{36}.$$

For the non-inflated data, $\pi = 0$, all the outcomes are from Poisson state. The type I error at 5% significance level table for 500 replications is as follows:

Table 13: Averaged type I error when $\pi = 0$

Methods	d	$\rho = 0$	$\rho = 0.25$	$\rho = 0.4$	$\rho = 0.6$	$\rho = 0.75$
CF I	100	0.1122	0.0865	0.0669	0.1393	0.3101
CF II	100	0.2164	0.1791	0.1866	0.2308	0.4747
CF III	100	0.1042	0.0765	0.0909	0.1639	0.3291
CF I	200	0.0980	0.1340	0.1280	0.1960	0.3988
CF II	200	0.2160	0.2460	0.2180	0.3060	0.5371
CF III	200	0.0980	0.1240	0.0960	0.1620	0.3888
CF I	500	0.0840	0.1167	0.0964	0.1487	0.3333
CF II	500	0.1700	0.2354	0.1968	0.2546	0.5135
CF III	500	0.0780	0.1167	0.0965	0.1767	0.3789

As the correlation parameter ρ increase from 0 to 0.75, the type I error increased for all of these methods. Among the three Cornish-Fisher expansions, the expansion with all three terms performs the best, hence we use this Cornish-Fisher critical value for the decorrelated score test.

5.0 DISCUSSION AND FUTURE WORK

5.1 MEASURING THE SIGNAL-TO-NOISE RATIO IN GLM

5.1.1 Signal-to-noise ratio in linear model

The signal-to-noise ratio (SNR) is widely used in science and engineering that measures the strength of a signal relative to the background noise. A standard definition of the SNR is:

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}, \tag{5.1}$$

where σ^2_{signal} is the variability introduced by the signal and σ^2_{noise} is the variability due to noise. An alternative definition of SNR in regression problems is as the ratio of regression coefficient to standard deviation of the noise:

$$SNR = \frac{|\beta|}{\sigma} \tag{5.2}$$

SNR is commonly expressed in decibels as $10 \log_{10}(SNR)$, the higher the higher SNR, the stronger the signal or information in the signal relative to the background noise. A definition of the SNR for GLM is proposed in Czanner et al. (2015).

In linear regression model (1.1), the covariate structure can be represented as $X\beta = X_1\tilde{\beta}_1 + X_2\tilde{\beta}_2$, where $X_1\tilde{\beta}_1$ is the covariate related to the signal, $X_2\tilde{\beta}_2$ is the covariate not related to the signal, i.e. $\tilde{\beta}_2 = 0$. The SNR is defined as

$$SNR_{X_1} = \frac{SSR(X_2) - SSR(X)}{SSR(X)} \tag{5.3}$$

where $SSR(X_2)$ is the regression sum of squares for X_2 and $SSR(X)$ is the regression sum of squares for X.

5.1.2 Extension of SNR to RZIP model

In Czanner et al. (2015), SNR is proposed for point process GLM which replaced the regression sum of square SSR with the deviance of regression model, Dev. The SNR for GLM is

$$SNR_{X_1} = \frac{Dev(X_2) - Dev(X)}{Dev(X)}, \tag{5.4}$$

and Dev is

$$Dev(X) = -2\log\frac{L(y, X\widehat{\beta})}{L(y, y))}, \tag{5.5}$$

where $L(y, X\widehat{\beta})$ is the likelihood evaluated at the MLE $\widehat{\beta}$ and $L(y, y)$ is the saturated likelihood. The paper used a Volterra series expansion of the conditional intensity function of a spiking neuron. Their supporting document defined SNR for the regression model and compared SNR with R^2, F-test and LR test. SNR defined by deviance in GLM is related to K-L divergence. It also discussed the idea of bias correction for SNR. Czanner et al. (2015) also suggested an approximate bias-corrected SNR estimate:

$$SNR_{X_1} = \frac{Dev(X_2) - Dev(X) + \dim(\beta_2) - \dim(\beta)}{Dev(X) + \dim(\beta)}. \tag{5.6}$$

We will adjust SNR of GLM base on the zero-inflated model and compare SNR of Poisson model and Anscombe transformed Gaussian model in the future.

5.2 HIGH-DIMENSIONAL EM ALGORITHM FOR RZIP MODEL

Balakrishnan et al. (2017) showed the application of EM algorithm to latent variable models and Wang et al. (2014a) extended the results to high-dimensional cases. The ZIP model (especially the uniform zero-state probability model) can be treated as a latent variable model hence we can adapt the high-dimensional EM algorithm to ZIP.

Suppose that the outcome Y and the zero state indicator Z have a joint density function f_{θ^*}. Let Ω be the parameter space. For each $\theta \in \Omega$, we let $k_\theta(z|y)$ denote the conditional density of z given y. The finite-sample Q function is defined as:

$$Q_n(\theta|\theta') = \frac{1}{n}\sum_{i=1}^{n}\int_{\mathcal{Z}} k_{\theta'}(z|y_i)\log f_{\theta^*}(y_i, z)dz \tag{5.7}$$

Wang et al. (2014a) assumed that Q_n is differentiable in its first argument and showed the the high-dimensional EM algorithm with a truncation step. For ZIP model, Q_n can be found from the function l_{RZIP}

$$l_{ZIP} = \sum_{Y_i=0}\log(\pi_i + (1-\pi_i)\exp(-\mu_i)) + \sum_{Y_i\neq 0}\log((1-\pi_i)\mu_i^{y_i}\exp(-\mu_i)/y_i!) \tag{5.8}$$

It would be of interest to apply the high-dimensional EM algorithm to the ZIP model and compare the results with the EM algorithm in Wang et al. (2014b).

6.0 BIBLIOGRAPHY

Anscombe, F. J. (1948). The transformation of poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254.

Balakrishnan, S., Wainwright, M. J., and Yu, B. (2017). Statistical guarantees for the em algorithm: From population to sample-based analysis. *The Annals of Statistics*, 45(1):77–120.

Bertocci, M. A., Bebko, G., Versace, A., Fournier, J. C., Iyengar, S., Olino, T., Bonar, L., Almeida, J. R., Perlman, S. B., Schirda, C., and Travis, M. (2016). Predicting clinical outcome from reward circuitry function and white matter structure in behaviorally and emotionally dysregulated youth. *Molecular Psychiatry*, 21(9):1194.

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli*, 19(4):1212–1242.

Cornish, E. A. and Fisher, R. A. (1938). Moments and cumulants in the specification of distributions. *Revue de l'Institut international de Statistique*, pages 307–320.

Czanner, G., Sarma, S. V., Ba, D., Eden, U. T., Wu, W., Eskandar, E., Lim, H. H., Temereanca, S., Suzuki, W. A., and Brown, E. N. (2015). Measuring the signal-

to-noise ratio of a neuron. *Proceedings of the National Academy of Sciences*, page 201505545.

Dezeure, R., Bühlmann, P., Meier, L., and Meinshausen, N. (2015). High-dimensional inference: Confidence intervals, p-values and r-software hdi. *Statistical Science*, 30(4):533–558.

Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360.

Hunt, X. J., Reynaud-Bouret, P., Rivoirard, V., Sansonnet, L., and Willett, R. (2019). A data-dependent weighted lasso under poisson noise. *IEEE Transactions on Information Theory*, 65(3):1589–1613.

Mitra, R. and Zhang, C.-H. (2016). The benefit of group sparsity in group inference with de-biased scaled group lasso. *Electronic Journal of Statistics*, 10(2):1829–1873.

Ning, Y., Liu, H., et al. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics*, 45(1):158–195.

Taylor, J. and Tibshirani, R. (2016). Post-selection inference for l1-penalized likelihood models. *preprint arXiv:1602.07358*.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Van de Geer, S., Bühlmann, P., Ritov, Y., and Dezeure, R. (2014). On asymptotically

optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202.

Wang, Z., Gu, Q., Ning, Y., and Liu, H. (2014a). High dimensional expectation-maximization algorithm: Statistical optimization and asymptotic normality. *preprint arXiv:1412.8729*.

Wang, Z., Ma, S., Wang, C.-Y., Zappitelli, M., Devarajan, P., and Parikh, C. (2014b). Em for regularized zero-inflated regression models with applications to postoperative morbidity after cardiac surgery in children. *Statistics in Medicine*, 33(29):5192–5208.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942.

Zhang, C.-H. and Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101(476):1418–1429.