Sparse estimation in high-dimensional zero-inflated Poisson regression model

Mengmeng Jiang$^{1,3}$ and Hang Zhang$^3$

$^1$School of mathematics and statistics, Xidian University, Xi’an, Shaanxi 710126, China
$^2$College of mechanical and Electrical Engineering, Henan Agricultural University, Zhengzhou, Henan 450002, China
$^3$Email: jangmeng72773@163.com

Abstract. In the high-dimensional zero-inflated Poisson (ZIP) regression model, the traditional Gauss-Newton iteration method and Expectation Maximization (EM) algorithm cannot do the variable selection. We propose an Expectation-Maximization-Regularization (EMR) algorithm. In the E step, we calculate the expected value of the implicit variable under the existing estimated value. In the MR step, the "regularization + likelihood function" method is used. We introduce the elastic net penalty and use the coordinate descent method to obtain the sparse estimate. The simulation results show that the EMR method makes the parameter estimation and variable selection simultaneously, and the estimation accuracy is high. Finally, applying the EMR method to forecast the number of municipal engineering Public-Private-Partnership (PPP) projects in Hebei Province, this model has a strong explanation.

1. Introduction

Counting data exists widely in many fields such as finance, insurance, epidemics, etc. For discrete counting data, Poisson distribution or negative binomial distribution can be generally used for processing. However, when there are too many zeros in the data, there will be zero inflation. Mullahy proposed the Hurdle model to separate zero and non-zero data in the data. It was first determined whether zero event occurred or not, and then the number of the event was determined when the event occurred at least once. The non-zero count came from a zero truncated Poisson distribution or a zero truncated negative binomial distribution [1, 2]. Lambert proposed a ZIP model [3]. The difference between the ZIP model and the Hurdle model was zero in the data. Zeros in the ZIP might come from zero count process or the Poisson count process. Fahrmeir and Echavarria studied zero-inflated additive model [4]. Ghosh et al. studied the problem of zero-inflated Bayes model [5]. Xie et al. studied the generalized Poisson mixed regression model [6].

In the study of the ZIP model, Lambert used the Gauss-Newton method and EM algorithm to estimate parameters. Wang et al. analyzed a bivariate ZIP regression model [7]. Long et al. applied the random effect ZIP model to a motivational interviewing-based safer sex intervention trial [8]. When the variable dimension is high, the above studies could not do the variable selection. If all the variables are taken into account, the cost will be large, and the accuracy of the prediction may not be good. Therefore, the variable selection in the high-dimensional ZIP regression model parameter estimation is useful.
In this paper, the regularization term is added in the M step of EM algorithm, and the EMR algorithm makes the parameter estimation and the variable selection simultaneously. The model is interpretive and has good prediction effect.

2. ZIP regression model

When observations are counting data, Poisson distribution processing is generally used, and the Poisson distribution expectation equals variance. However, we sometimes found that there has a large number of zeros in the data and the variance is greater than expectation, reflecting the phenomenon of zero inflation. This kind of data can be regarded as a combination of zero count data and Poisson distribution data, or a combination of zero count data and negative binomial distribution data. This paper only considers the combination of zero count data and Poisson distributed data. We have response variables \( Y \), the ZIP model has the form

\[
P(Y_i = y_i) = \begin{cases} 
\phi_i + (1 - \phi_i) \exp(-\lambda_i), & y_i = 0, \\
\left(1 - \phi_i\right) \frac{\lambda_i^y \exp(-\lambda_i)}{y_i!}, & y_i > 0.
\end{cases}
\]

(1)

In equation (1), \( \lambda_i > 0 \), \( i = 1, 2, L, n \). \( \lambda = (\lambda_1, \lambda_2, L, \lambda_n) \) are Poisson parameter, \( \phi = (\phi_1, \phi_2, L, \phi_n) \) are zero inflation parameter, indicating probabilities that observed values in the zero counting process are zero. In the analysis of zero inflation data explanatory variables and response variables, covariate variables are introduced [9]. We can apply linkage functions to the zero inflation parameters \( \phi \) and the Poisson parameters \( \lambda \), respectively

\[
\begin{align*}
\text{logit}(\phi) &= \ln\left(\frac{\phi}{1-\phi}\right) = Zb, \\
\ln(\lambda) &= X\beta,
\end{align*}
\]

(2)

where \( X = (x_1, x_2, L, x_n)^T \), \( x_i \in R^p \) and \( Z = (z_1, z_2, L, z_n)^T \), \( z_i \in R^r \) are covariate variables, \( b = (b_1, b_2, L, b_p)^T \) and \( \beta = (\beta_1, \beta_2, L, \beta_r)^T \) are regression coefficients. In this paper, we use \( \ln(\lambda) = \log_e (\lambda) \). The log-likelihood function has the form

\[
l_{ZIP}(b, \beta; Y) = \sum_{i=1}^{n} I_{(y_i=0)} \ln(\phi_i + (1-\phi_i) e^{-\lambda_i}) + \sum_{i=1}^{n} I_{(y_i>0)} [\ln(1-\phi_i) - \lambda_i + y_i \ln(\lambda_i) - \ln(y_i!)].
\]

(3)

where \( I \) is the indicator function. The score function of \( \theta = (b^T, \beta^T)^T \) is \( U(\theta) = \frac{\partial l_{ZIP}}{\partial \theta} \). Since there is no explicit solution when maximizing the equation (3), it can be solved by Gauss-Newton iteration method or EM algorithm.

3. EMR algorithm

3.1. Algorithm Introduction

We introduce a binary indicator \( u_i \), \( u_i = 1 \) indicating that the observation \( y_i \) comes from the degenerate distribution; \( u_i = 0 \) indicating that the observation \( y_i \) comes from non-degenerate distributions. We can regarded \( u = (u_1, u_2, L, u_n) \) as missing data, and \( Y = (x_i, z_i, y_i), i = 1, 2, L, n \) as observed data, so complete-data could be recorded as \( Y_n = (u, Y_c) \). The log-likelihood function of the complete-data is

\[
l(\theta|Y_n) = \sum_{i=1}^{n} u_i \ln\left(\frac{\phi_i}{1-\phi_i}\right) + \ln(1-\phi_i) + \sum_{i=1}^{n} (1-u_i)[y_i \ln(\lambda_i) - \ln(y_i!)].
\]

(4)

The procedure is outlined as followings:
E Step: Expectation

\[
Q(\theta | \theta^{(t)}) = E \left\{ \ln \left( \frac{\prod_{i=1}^{n} \left( y_i^{\theta} \right)^{y_i} \left( 1 - y_i^{\theta} \right)^{1 - y_i}}{1 - \prod_{i=1}^{n} \left( 1 + \exp(-\lambda \phi_i) \right)^{-1}} \right) \right\} = L_1(b) + L_2(\theta). \tag{5}
\]

where the \( \theta^{(t)} \) is the parameter estimate at step \( t \), and

\[
L_1(b) = \sum_{i=1}^{n} E \left( u_i \left| r_i \right. \right) \ln \left( \frac{\prod_{i=1}^{n} \left( y_i^{\theta} \right)^{y_i} \left( 1 - y_i^{\theta} \right)^{1 - y_i}}{1 - \prod_{i=1}^{n} \left( 1 + \exp(-\lambda \phi_i) \right)^{-1}} \right) \ln \left( \frac{\prod_{i=1}^{n} \left( y_i^{\theta} \right)^{y_i} \left( 1 - y_i^{\theta} \right)^{1 - y_i}}{1 - \prod_{i=1}^{n} \left( 1 + \exp(-\lambda \phi_i) \right)^{-1}} \right),
\]

\[
L_2(\theta) = \sum_{i=1}^{n} \left( 1 - E \left( u_i \left| r_i \right. \right) \right) \ln \left( \frac{\prod_{i=1}^{n} \left( y_i^{\theta} \right)^{y_i} \left( 1 - y_i^{\theta} \right)^{1 - y_i}}{1 - \prod_{i=1}^{n} \left( 1 + \exp(-\lambda \phi_i) \right)^{-1}} \right),
\]

\[
E \left( u_i \left| r_i \right. \right) = \left( 1 + \exp \left( -\lambda \phi_i \right) \right)^{-1}. \tag{8}
\]

MR Step: The maximum likelihood estimation of model parameters are obtained by maximizing the penalty likelihood function

\[
\max_{b \in \mathbb{R}^m} \left[ L_1(b) - \lambda P_u(b) \right]. \tag{9}
\]

\[
\max_{\theta \in \mathbb{R}^p} \left[ L_2(\theta) - \lambda P_v(\theta) \right]. \tag{10}
\]

The next section describes the detailed estimation of parameters in the MR step.

3.2. Sparse estimation of the model
The following has a detailed description of how to use \( L_1(b) \) and \( L_2(\theta) \) to get the sparse estimates in MR steps. Ridge regression cannot produce a parsimonious model [10]. In the p>n case, the lasso has the limitations of the most choice of n variables [11]. In this paper, we use the elastic net penalty, the elastic net penalty is a combination between the ridge regression penalty and the lasso penalty. This penalty is useful where there are many correlated predictor variables [12]. This paper only introduces the sparse estimation of \( b \), and the same method can be used to obtain the estimate of \( \beta \). We have a predictor vector \( Z \), and assume that \( z_{ij} \) are standardized: \( \sum_{i=1}^{n} z_{ij} = 0 \), \( \sum_{j=1}^{p} z_{ij} = 0 \), for \( j = 1 \sim p \). We maximize the penalty likelihood function defined by

\[
\max_{b \in \mathbb{R}^m} \left[ L_1(b) - \lambda P_u(b) \right] = \min_{b \in \mathbb{R}^m} \left[ -L_1(b) + \lambda P_u(b) \right] \tag{11}
\]

\[
= \min_{b \in \mathbb{R}^m} \sum_{i=1}^{n} \left[ E \left( u_i \left| r_i \right. \right) \ln \left( \frac{\prod_{i=1}^{n} \left( y_i^{\theta} \right)^{y_i} \left( 1 - y_i^{\theta} \right)^{1 - y_i}}{1 - \prod_{i=1}^{n} \left( 1 + \exp(-\lambda \phi_i) \right)^{-1}} \right) + \lambda \sum_{i=1}^{p} \left( \frac{1}{2} (1 - \alpha) \beta_i^2 + \alpha \beta_i^2 \right) \right],
\]

where \( E \left( u_i \left| r_i \right. \right) \) is valuated using parameters at step \( t \). Substituting \( \phi = \frac{\exp(Zb)}{1 + \exp(Zb)} \) to \( L_1(b) \), the log-likelihood function can be written in a more explicit form

\[
L_1(b) = \sum_{i=1}^{n} E \left( u_i \left| r_i \right. \right) z_i b - \ln(1 + \exp(z_i b)), \tag{12}
\]

a concave function of the parameter. If the current estimates are \( b^{(t)} \), we form a Taylor expansion, which is

\[
L_1(b) = -\frac{1}{2} \sum_{i=1}^{n} w_i (h_i - z_i b)^2 + c \left( b^{(t)} \right)^2, \tag{13}
\]
\[ h_i = z_i^T b^{(t)} + \frac{y_i - \varphi^{(t)}(z_i)}{\varphi^{(t)}(z_i)(1 - \varphi^{(t)}(z_i))}, \]

(14)

\[ w_i = \varphi^{(t)}(z_i)(1 - \varphi^{(t)}(z_i)). \]

(15)

where \( \varphi^{(t)}(z_i) \) is valued using parameters at step \( t \), and the last term is constant. We use Bayesian Information Criterion (BIC) to select the tuning parameter. Parameter estimation steps are as follows:

STEP1: using the current estimate to update the quadratic approximation.

STEP2: using coordinate descent algorithm [13] to implement \( \min_{b \in \mathbb{R}} [-L_0(b) + \lambda P_d(b)] \).

4. Simulation studies

4.1. Parameter estimation

In the following example, we compare the proposed EMR algorithm with the Gauss-Newton iterative and EM algorithm. The simulations are conducted using R code. First we generate data that obeys the ZIP model. We assume that the design matrix \( \mathbf{Z} = \mathbf{X} \), and they are multivariate normal distribution with the mean of 0 and the correlation between \( x \) and \( x \) is \( \rho = \rho^4 \) is \( \rho = 0.1 \). \( \alpha = (2,0,1,0,0,2,1,0,0,1,0,0,0)^T \). \( \beta = (0.1,0.2,0.0,0.0,0.2,0.0,0.0,0.0,0.2)^T \). The sample size is 200, Gauss-Newton iteration method, EM algorithm and EMR algorithm results are shown in table 1. From table 1, we know that the parameter estimate between the Gauss-Newton iterative method and the EM algorithm are close, and none of the EMR algorithm can reduce the smaller coefficient to zero, making the parameter estimation and variable selection simultaneous.

| Table 1. Estimated value of test parameters. |
|----------------------------------------------|
| \( \text{Real value} \) | \( \text{Logistic} \) | \( \text{Poisson} \) | \( \text{Gauss-Newton} \) | \( \text{Logistic} \) | \( \text{Poisson} \) | \( \text{EM} \) | \( \text{Logistic} \) | \( \text{Poisson} \) | \( \text{EMR} \) | \( \text{Logistic} \) | \( \text{Poisson} \) |
|----------|----------|----------|----------------|----------|----------|----------------|----------|----------|----------------|----------|----------|----------------|----------|
| Intercept | 0.0      | 0.0      | 1.197          | 1.137    | 1.197    | 1.137          | 0.000    | -0.226    |                |          |          |                |          |
| \( x_1 \) | 2.0      | 0.1      | -0.714         | 0.672    | -0.714   | 0.672          | 4.287    | 0.318     |                |          |          |                |          |
| \( x_2 \) | 0.0      | 0.0      | -0.318         | 0.218    | -0.318   | 0.218          | 0.000    | -0.465    |                |          |          |                |          |
| \( x_3 \) | 1.0      | 0.2      | -1.187         | -0.512   | -1.187   | -0.512         | 0.600    | 1.200     |                |          |          |                |          |
| \( x_4 \) | 0.0      | 0.0      | -0.891         | -0.153   | -0.891   | -0.153         | 0.000    | 0.000     |                |          |          |                |          |
| \( x_5 \) | 0.0      | 0.0      | 0.423          | -0.067   | 0.423    | -0.067         | 0.000    | 0.000     |                |          |          |                |          |
| \( x_6 \) | 2.0      | 0.2      | -0.345         | 0.228    | -0.345   | 0.228          | 1.300    | -0.494    |                |          |          |                |          |
| \( x_7 \) | 1.0      | 0.0      | -0.253         | -0.129   | -0.253   | -0.130         | 0.000    | 0.000     |                |          |          |                |          |
| \( x_8 \) | 0.0      | 0.0      | 1.115          | 0.630    | 1.115    | 0.630          | 0.000    | 0.000     |                |          |          |                |          |
| \( x_9 \) | 0.0      | 0.2      | -1.039         | 0.224    | -1.039   | 0.224          | 0.000    | 0.274     |                |          |          |                |          |
| \( x_{10} \) | 1.0     | 0.0      | 0.102          | -0.065   | 0.102    | -0.065         | 3.354    | 0.000     |                |          |          |                |          |
| \( x_{11} \) | 0.0     | 0.0      | -0.455         | 0.124    | -0.455   | 0.124          | 0.000    | 0.000     |                |          |          |                |          |
| \( x_{12} \) | 0.0     | 0.2      | 0.054          | 0.234    | 0.054    | 0.234          | 0.000    | 0.870     |                |          |          |                |          |

4.2. Model evaluation

In this section, we take \( p=12 \) and \( p=40 \), where the non-zero terms in the zero inflation parameters and Poisson parameters are \( 5 \) and \( 10 \). In addition, the sample size are \( n=100 \) and \( n=200 \), and experiment is repeated 200 times. Results are reported in table 2, in which the column NC presents the average number of nonzero coefficients correctly estimated to be nonzero, and the column NIC depicts the
average number of zero coefficients incorrectly estimated to be nonzero. The MAE depicts mean absolute error. From the numerical simulation results, it can be seen that:

(1) From the model prediction accuracy point of view, the MAE values of the Gauss-Newton iteration method and the EM algorithm are almost equal. With the increase of the sample, the MAE value gradually decreases, and the MAE value of the EMR algorithm is slightly lower than Gauss-Newton and EM algorithm values;

(2) From the aspect of model complexity, the Gauss-Newton iteration method and the EM algorithm estimate all the coefficients as non-zero. While using the EMR algorithm, the value of NC is close to the nonzero number of coefficients of 5 and 10, the value of NIC is close to 0, and the NIC value of logistic process is closer to 0 than Poisson process.

### Table 2. Simulation results.

| Method | MAE Logistic | NIC | MAE Possion | NIC | MAE Logistic | NIC | MAE Possion | NIC |
|--------|--------------|-----|-------------|-----|--------------|-----|-------------|-----|
| P=12 n=100 | 0.705 5 7 5 7 0.496 5 7 5 7 | | | | | | | |
| Newton | | | | | | | | |
| EM | 0.706 5 7 5 7 0.496 5 7 5 7 | | | | | | | |
| EMR | 0.659 4.090 0.640 5.680 0.695 0.480 4.265 0.815 5.470 1.080 | | | | | | | |
| P=40 n=200 | | | | | | | | |
| Newton | 0.418 10 30 10 30 0.246 10 30 10 30 | | | | | | | |
| EM | 0.418 10 30 10 30 0.246 10 30 10 30 | | | | | | | |
| EMR | 0.364 9.410 0.535 10.350 0.620 0.210 9.650 0.750 9.530 0.865 | | | | | | | |

5. Applications to PPP data

5.1. Data introduction

This section examines the relationship between economic indicators and the number of PPP projects in Hebei Province. The PPP project information comes from the data released by the Ministry of Finance project database as of September 30, 2017, including PPP projects in the identification phase, preparation phase, procurement phase, and implementation phase. The data do not include provincial and municipal levels projects in Hebei Province. Nearly 92% of projects are initiated after 2015. Taking into account the lag in the relative economic data of the development plan, the truncated data of 18 indicators in 132 counties and cities in Hebei Province are used for analysis. The county and city indicators are derived from the Statistical Yearbook of China's Counties. See table 3 for specific indicators. When dealing with explanatory variables, we centralize and standardize the treatment to eliminate the influence of different dimensions.

Analyzing the number of PPP projects, we find that the percentage of zero is 47%, the mean was 1.23, and the variance was 2.45, and the variance was greater than the mean value. Our initial judgment data has zero inflation. We use the Vuong test [14] to select the ZIP model and Poisson model. Define

\[ m_i = \ln \left( \frac{P_{zip}(y_i|x_i)}{P_{poisson}(y_i|x_i)} \right), \]  \( (16) \)

where \( P(y_i|x_i) \) represents the predicted probability of the corresponding model in the random variable \( y = y_i \), then the Vuong test statistic of the ZIP model relative to the Poisson regression model.
\[ V = n^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} m_i \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (m_i - \bar{m})^2 \right)^{1/2} / S_n \]  

(17)

where \( \bar{m}, S_n \) are the mean and standard deviation of \( m_i \), respectively, and \( n \) is the sample size. After calculation, the \( V \) value is 2.34, \(|V|>1.96\). The ZIP model is more suitable for fitting the data than the Poisson model.

**Table 3.** Indicators and parameter estimation.

| Variable | Indicator                                           | Logistic(EMR) | Poisson(EMR) |
|----------|-----------------------------------------------------|---------------|--------------|
| Intercept|                                                     | 0.000         | 0.000        |
|          | Administrative area                                 | -2.068        | 0.000        |
|          | Primary industrial product                          | 0.000         | 0.000        |
|          | Secondary industry product                          | 0.000         | 0.491        |
|          | Tertiary industry product                           | 0.000         | 0.000        |
|          | Gross industrial production                         | 1.523         | 0.000        |
|          | Total Population                                    | 0.000         | 0.000        |
|          | The average wage of workers                          | 0.000         | 0.000        |
|          | Total investment in fixed assets                     | 0.000         | -2.657       |
|          | Energy loss per unit of GDP                          | 0.000         | 3.115        |
|          | Local public budget revenue                          | -0.467        | 0.000        |
|          | The local public finance budget expenditure          | 0.169         | 0.000        |
|          | Per capita disposable income of urban residents      | 0.000         | -0.658       |
|          | Per capita net income of rural residents             | 0.000         | 0.012        |
|          | Industrial output above designated size             | 3.394         | 1.250        |
|          | The total retail sales of social consumer goods       | 0.000         | 1.339        |
|          | Total exports                                        | -1.072        | -0.115       |
|          | Actual use of foreign capital                        | -0.048        | -0.881       |
|          | Year-end balance of savings deposits of urban and    | 0.000         | 0.000        |

5.2. Parameter estimation and model evaluation

We take 80% of the sample for training and 20% for testing. The MAE for Gauss-Newton, EM and EMR are 0.530, 0.531, 0.463. The results show that the MAE value obtained by EMR algorithm is less than the other two methods. Because Gauss-Newton iterative method and EM algorithm cannot do the variable selection, the model explanatory is poor, and there is a large MAE. Therefore, this paper analyzes the estimation results of the EMR algorithm, the results are shown in table 3.

First, we analyze whether there are factors affecting the municipal engineering PPP project:

1. The coefficient of gross industrial production is positive. The increase in its value leads to an increase in the probability of a zero dependent variable, that is, the greater the gross industrial production in the region, the more impossible it is to implement a PPP project;

2. The smaller the local public finance budget revenue and the greater the expenditure, the more impossible it is to implement PPP projects. This shows that the PPP project is less likely to be implemented when its own financial resources are insufficient;

3. The larger the actual amount of foreign capital used in the year and the greater the export value, the more likely it is to implement PPP projects. It shows that the more prosperous the domestic and foreign trade in this area, the greater the possibility of implementing PPP projects.
Under the premise that there is a municipal engineering PPP project, we analyze the influencing factors of the number of projects:

1. The greater the total production values of the secondary industry, the greater the number of projects;
2. The energy loss per unit of GDP is large, indicating that the economic development of the area is highly dependent on energy, and may require industrial transformation and increase the number of projects;
3. The more per capita net income of rural residents, the total industrial output value above designated size, and the total retail sales of social consumer goods, the more the number of projects in the region;
4. The larger the actual amount of foreign capital used, the larger the export amount, and the smaller the number of projects, indicating that the potential of the import and export market is greater, the construction of the municipal engineering project may have been relatively complete, and the number of projects needed may be reduced.

6. Conclusions

In the parameter estimation of high dimensional ZIP regression model, the EM algorithm is modified and the EMR algorithm is proposed. This method makes parameter estimation and variable selection simultaneously, and the model is easily explained. The simulation analysis shows that when there are many explanatory variables, the EMR algorithm with elastic net penalty can not only do the variable selection, but also have smaller mean absolute error. Finally, we apply EMR algorithm to analyze influencing factors of the number of county-level municipal engineering PPP projects in Hebei Province. We find that factors such as finance, industry, import and export have a certain relationship with the number of regional PPP projects.

References
[1] Mullahy J 1986 Specification and testing of some modified count data models Journal of Econometrics 33(3) 341-365
[2] Dalrymple M L, Hudson I L and Ford R P K 2003 Finite mixture, zero-inflation Poisson and Hurdle models with application to SIDS Computational Statistics & Data Analysis 41(3-4) 491-504 More references
[3] Lambert D 1992 Zero-inflated Poisson regression, with an application to defects in manufacturing Technometrics 34(1) 1-14
[4] Fahrmeir L and Echavarría L O 2006 Structured additive regression for overdispersed and zero-inflation count data Applied Stochastic Models in Business & Industry 22(4) 351-369
[5] Ghosh S K , Mukhopadhyay P and Lu J C 2006 Bayesian analysis of zero-inflation regression models Journal of Statistical Planning & Inference 136(4) 1360-1375
[6] Xie F C ,Wei B C and Lin J G 2009 Score tests for zero-inflation generalized Poisson mixed regression models Computational Statistics & Data Analysis 53(9) 3478-3489
[7] Wang K, Lee A H, Yau K K and Carrivick P J 2003 A bivariate zero-inflated Poisson regression model to analyze occupational injuries Accident; analysis and prevention 35(4) 625-629
[8] Long D L, Preisser J S, Herring A H and Golin C E 2015 A Marginalized zero-inflated Poisson regression model with random effects Journal of the Royal Statistical Society 64(5) 815-830
[9] Xu X, Yuan W and Meng S 2009 Zero-inflation generalized Poisson regression model and its application to insurance rate making Mathematics in Practice & Theory 24(24) 99-107
[10] Mcdonald G C 2010 Ridge regression Wiley Interdisciplinary Reviews Computational Statistics 1(1) 93-100
[11] Tibshirani R 2011 Journal of the Royal Statistical Society 73(3) 273-282
[12] Zou H and Hastie T 2005 Journal of the Royal Statistical Society 67(2) 301-320
[13] Friedman J, Hastie T and Tibshirani R 2010 Journal of Statistical Software 33(1) 1-22
[14] Vuong Q H 1989 Econometrica 57(2) 307-333