Centre of Excellence
for Science and Innovation Studies

CESIS Electronic Working Paper Series

Paper No. 257

A Ridge Regression estimator for the zero-inflated Poisson model

B. M. Golam Kibria
Kristofer Månsson
Ghazi Shukur

October 2011

The Royal Institute of technology
Centre of Excellence for Science and Innovation Studies (CESIS)
http://www.cesis.se

A Ridge Regression estimator for the zero-inflated Poisson model

by

B. M. Golam Kibria[1] Kristofer Månsson[2] and Ghazi Shukur[2,3]

[1] Department of Mathematics and Statistics, Florida International University, Modesto A. Maidique Campus, Miami, Florida, USA

[2] Department of Economics, Finance and Statistics, Jönköping University, Sweden

[3] Department of Economics and Statistics, Linnaeus University, Sweden

Abstract

The zero inflated Poisson regression model is very common when analysing economic data that comes in the form of non-negative integers since it accounts for excess zeros and over-dispersion of the dependent variable. This model may be used in innovation analysis to see for example the impact on different economic factors has on the number of patents of companies, how frequently mortgages or credit-card fail to meet their financial obligations, the amount of take-over bids firms receives and the number of products different firms export, etc. However, a problem often encountered when analyzing economic data that has not been addressed for this model is multicollinearity. This paper proposes ridge regression estimators and some methods of estimating the ridge parameter *k* for the non-negative model. A simulation study has been conducted to compare the performance of the estimators. Both MSE and MAE are considered as performance criterion. The simulation study shows that some estimators are better than the commonly applied maximum likelihood estimator and some other ridge regression estimators. Some useful estimators are recommended for the practitioners.

Key words: Count data; Innovation analysis; Multicollinearity; Zero Inflated Poisson; Ridge Regression.

JEL Classification: C13; C16; C31

1. Introduction

The Zero Inflated Poisson (ZIP) regression model introduced by Lambert (1992) is a popular choice among researchers in applied economics when the dependent variable comes in the form of non-negative integers or counts. This model may be used to see the impact on different economic factors has on the number of patents of companies, how frequently mortgages or credit-card fail to meet their financial obligations, the amount of take-over bids firms receives and the number of products different firms export, to mention some examples. The ZIP model is often appealing firstly since it divides the dependent variable into two subpopulations where the first one takes on the value zero with probability π_i while the second one is Poisson distributed with probability $1-\pi_i$. Due to this setup the ZIP model may be used when the data contains an excess amount of zeros. Secondly, this model is popular, because it accounts for overdispersion (meaning that the variance of the dependent variable exceeds the mean value) which leads to an underestimation of the variance of the estimated coefficients when applying the standard Poisson regression model.

However, a problem when using the ZIP model is the multicollinearity first defined by Frisch (1934) as the situation when the independent variables are highly inter-correlated. This problem can often be found when using count data models in applied economic research. For instance Gråsjö (2005) had this problem since both government and university R&D were used as explanatory variables when the author analyzed the number of patents for Swedish firms. Another example can be found in Greene (1994) where age, job-experience and income were included in the regression model explaining how frequently credit-card holders failed to meet their financial obligations. The problem of multicollinearity is that it leads to a high variance and difficulties in interpreting the estimates of the coefficients. Hence, multicollinearity makes it difficult to make valid statistical inference and to investigate the impact of different economic factors on the dependent variable. As a solution to this problem for linear regression models Hoerl and Kennard (1970 a,b) proposed that one may use ridge regression (RR) which is a biased shrinkage estimator instead of ordinary least squares (OLS). Hoerl and Kennard (1970 a,b) showed that one may reduce the variance of the estimated coefficients substantially by introducing a small amount of bias. This method was then generalized in order to be used for models estimated by maximum likelihood (ML) such as the logit and Poisson models by Schaeffer et al. (1984), Månsson and Shukur (2011a,b), among others.

The purpose of this paper is to propose a RR estimator for the ZIP model. Furthermore, we are going to generalize some methods of estimating the ridge parameter, first proposed for linear regression by Hoerl and Kennard (1979 a,b), Kibria (2003), Muniz and Kibria (2009) and Kibria et al. (2011), so they can be used for ZIP ridge regression models (ZIP RR). In order to judge the performance of the different estimators we calculate the mean squared error (MSE) and the mean absolute error (MAE) using Monte Carlo simulations. In the simulation study we change factors such as the sample size, the degree of the correlation among the regressors and the value of the intercept of the logit model. The results from the simulation study clearly show that the new RR estimator for the ZIP model outperforms the ML.

The paper is organized as follows: in Section 2, we describe the statistical methodology. The design of the experiment and simulation results are discussed in Section 3. Finally, in Section 4 we give a brief summary and conclusions.

2. Statistical methodology

2.1 The ZIP model

The ZIP regression model is a popular choice for applied researchers when the dependent variables comes in the form of non-negative integers or counts that can be divided into two subpopulation using the following equation:

$$y_i = \begin{cases} 0 \text{ with probability } \pi_i \\ \text{Po}\left(\mu_i\right) \text{ with probability } 1\text{-}\pi_i \end{cases}. \tag{2.1}$$

where $\pi_i = \dfrac{\exp\left(x_i\gamma\right)}{1+\exp\left(x_i\gamma\right)}$ and $\mu_i = \exp\left(x_i\beta\right)$ where x_i is the i*th* row of X and β is a $(p+1)\times 1$ vector of coefficients. The most common method of estimating β is to apply the ML method where the following joint log likelihood should be maximized:

$$l = \sum_{i=1}^{N} y_i \log(\pi_i) + \sum_{i=1}^{N} (1-y_i)\log(1-\pi_i). \tag{2.2}$$

This complex likelihood function is estimated by using the simplex method proposed by Nelder and Mead (1965) where the start-up values come from the iterative weighted least squares (IWLS) algorithm of the individual Poisson and logit estimation. As a modification to this method we propose the following estimator which for the Poisson model in Månsson and Shukur (2011b):

$$\hat{\beta}_{RR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\hat{\beta}_{ML}, \qquad (2.3)$$

where $\hat{\beta}_{ML}$ is the estimates obtained using the simplex method The $\hat{\mathrm{W}}$ is a matrix where the non-diagonal elements equals to zero and the i*th* diagonal element equals to $\hat{\mu}_i$. This estimator is a biased shrinkage estimator in the same spirit as the ones proposed by for example Hoerl and Kennard (1970a,b), Schaeffer et al. (1984) and Månsson and Shukur (2010a,b) for the linear, logit and Poisson models, respectively. The shrinkage parameter *k* may take on values between zero and infinity and when *k* equals zero we have $\hat{\beta}_{RR} = \hat{\beta}_{ML}$. When *k* is greater than zero we have $\left\|\hat{\beta}_{RR}\right\| \leq \left\|\hat{\beta}_{ML}\right\|$. Since $\hat{\beta}_{ML}$ is, on average, too long in the presence of multicollinearity, $\hat{\beta}_{RR}$ is expected to perform better than $\hat{\beta}_{ML}$.

2.2 The Proposed Ridge Parameter Estimators

There is no definite rule of how to estimate the ridge parameter. However, several methods have been proposed for the linear RR model and these will be generalized in this paper to be applicable for ZIP RR. The first estimator is the following:

$$K1 = \hat{k}_{HK1} = \frac{\hat{\sigma}^2}{\hat{\alpha}^2_{\max}},$$

proposed by Hoerl and Kennard (1970a,b), where we define $\alpha^2_{\max}$ to be the maximum element of $\gamma\hat{\beta}_{ML}$, where γ is the eigenvector of $X'\hat{W}X$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(y_i - \hat{\mu}_i\right)^2}{n-p-1}$. A second estimator of *k* introduced by Schaeffer et al. (1984) for logistic RR is the following:

$$K2 = \hat{k}_{HKM} = \frac{1}{\hat{\alpha}^2_{\max}}.$$

Then two estimators proposed for linear regression by Kibria (2003) are evaluated, namely:

$$K3 = \hat{k}_{GM} = \frac{\hat{\sigma}^2}{\left(\prod_{i=1}^{p} \hat{\alpha}_i^2\right)^{\frac{1}{p}}} \text{ and } K4 = \hat{k}_{MED} = Median\left\{m_i^2\right\},$$

where $m_i = \sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}$. Finally, the following estimators from Muniz and Kibria (2009) and Kibria et al. (2011) are considered:

$$K5 = \max\left(\frac{1}{m_j}\right), \quad K6 = \max\left(m_j\right), \quad K7 = \prod_{j=1}^{p}\left(\frac{1}{m_j}\right)^{\frac{1}{p}}, \quad K8 = \prod_{j=1}^{p}\left(m_j\right)^{\frac{1}{p}}$$

$$K9 = median\left(\frac{1}{m_j}\right), K10 = median\left(m_j\right), K11 = \max\left(\frac{1}{q_j}\right), \quad K12 = \max\left(q_j\right),$$

$$K13 = \prod_{j=1}^{p}\left(\frac{1}{q_j}\right)^{\frac{1}{p}}, \quad K14 = \prod_{j=1}^{p}\left(q_j\right)^{\frac{1}{p}}, \quad K15 = median\left(\frac{1}{q_j}\right), \quad K16 = median\left(q_j\right)$$

where $q_j = \frac{\lambda_{\max}}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_j^2}$ and $\lambda_{\max}$ is defined as the maximum eigenvalue of $X'\hat{W}X$.

2.3 The measurements of the performance of the estimators

To investigate the performance of the ZIP RR and the ML methods, we calculate the MSE using the following equation:

$$MSE = \frac{\sum_{i=1}^{R}\left(\hat{\beta}-\beta\right)\left(\hat{\beta}-\beta\right)}{R}, \tag{2.4}$$

and the MAE using the following equation:

$$MAE = \frac{\sum_{i=1}^{R}\left|\hat{\beta}-\beta\right|}{R} \tag{2.5}$$

where $\hat{\beta}$ is the estimator of β obtained from ML or ZIP ridge regression estimators. R equals 2000 which corresponds to the number of replicates used in the Monte Carlo simulation.

3. The Monte Carlo simulation

This section describes how we generate the data and which factors that have been varied in the design of the experiment. Then, a discussion of the results obtained from the Monte Carlo simulations is provided.

3.1 The Design of the Experiment

The key factor varied in the design of the experiment is the degree of correlation between the independent variables (ρ^2). To be able to generate data with different degrees of correlation we use the following formula to obtain the regressors:

$$x_{ij} = \left(1-\rho^2\right)^{(1/2)} z_{ij} + \rho z_{ip}, \qquad i = 1,2,\ldots n,\ j = 1,2,\ldots p \qquad (3.1)$$

where z_{ij} are pseudo-random numbers generated using the standard normal distribution. Three different values of ρ^2 corresponding to 0.85, 0.95 and 0.99 are considered in the simulation study. Then, based on the regressors, the dependent variable of the ZIP regression model is generated according to equation (2.1). We first generate a binary variable using pseudo-random numbers from the binomial distribution where $\pi_i = \frac{\exp(\gamma_0)}{1+\exp(\gamma_0)}$ and then the non-zero values of the binary variable are obtained from the Poisson distribution with $\mu_i = \exp\left(\beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots x_p\beta_p\right)$. The number of regressors (p) equals 2 and 4, the slope parameters are chosen so that $\sum_{j=1}^{p}\beta_j^2 = 1$ and the intercept of the Poisson model is always equal to zero. However, the intercept of the logit model (γ_0) is varied since this factor affects the probability of obtaining zeros and ones. When the intercept equals zero then there is an equal average probability of obtaining ones and zeros. When the intercept is positive then there is a greater probability of obtaining zeros than ones. Accordingly, the value of the intercept is set to be equal to 0, 1 and 2 in the simulation study. Finally, we vary the sample size (n) between 50 and 150. However, when π_i is high and the sample size is low we sometimes obtain a dependent variable consisting of only zeros. Therefore, we need, just as in Månsson and Shukur (2011a,b), to increase the sample size with the intercept of the logit model in order to achieve convergence of the simplex algorithm. The different combinations of intercept and sample sizes can be found in Table 1:

Table 1: Combinations of intercept and sample sizes.

		Sample size						
Intercept	50	75	100	150	200	300	400	500
0	*	*	*	*	*			
1			*	*	*	*		
2					*	*	*	*

3.2 Result discussion

The estimated MSE and MAE are presented for p=2 in Tables 2 and 3 respectively and for p=4 in Tables 4 and 5 respectively. The general pattern according to both measures of performance is that, increasing the degree of correlation and the number of explanatory variables have a negative impact, while increasing the number of observations has a positive impact on both MSE and MAE. All of the proposed ridge regression estimators outperform the ML in the sense that they have smaller MSE and MAE. However, the estimators, K11, K13 and K15 are performing better among the other ridge estimators in most situations, in the sense of lower MSE and MAE although their performances are little poor when sample sizes are large or in the cases of weak to moderate correlation among the regressors. These three estimators may be recommended for practitioners. The gain in terms of MSE and MAE is the highest when the sample size is small, the number of zeros is moderate and the degree of correlation is high.

4. Summary and Conclusions

This paper proposes a new biased shrinkage estimator for the ZIP regression model in order to solve the problem of inflated variance of the classical ML method which is commonly applied to estimate the ZIP model. Furthermore, we generalize some methods of estimating the shrinkage parameter k that were developed for the linear regression model by Hoerl and Kennard (1970(a,b), Kibria (2003), Muniz and Kibria. (2009) and Kibria et al. (2011)). A Monte Carlo simulation study is conducted to compare the performance of the estimators using both the MSE and the MAE criterion. From the simulated results we can clearly see that the variance of the classical ML method becomes inflated in the presence of multicollinearity. On the other hand, the new estimation method works especially well when the degree of correlation is high, the sample size is small, the number of explanatory variables is large and when the problem of excess zeros is moderate. The K11, K13 and K15 estimators are performing better than the ML and rest of ridge regression estimators in the sense of smaller MSE and MAE. We are therefore recommending the K11 K13 and K15 for estimating the ridge parameter for ZIP regression model. We strongly believe that the findings of the paper will be useful for the practitioners.

Table 1: Estimated MSEs of the proposed estimators when p=2

Estimated MSE when the intercept equals 0																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
50	0.332	0.269	0.260	0.229	0.256	0.272	0.226	0.286	0.245	0.282	0.239	0.126	0.325	0.130	0.325	0.130	0.325
75	0.163	0.137	0.133	0.121	0.145	0.146	0.125	0.150	0.133	0.149	0.130	0.094	0.160	0.096	0.160	0.096	0.160
100	0.113	0.097	0.094	0.084	0.112	0.108	0.090	0.109	0.095	0.109	0.093	0.083	0.112	0.084	0.112	0.084	0.112
150	0.059	0.053	0.052	0.046	0.068	0.058	0.050	0.058	0.052	0.058	0.051	0.051	0.059	0.051	0.059	0.051	0.059
ρ=0.95																	
50	0.930	0.666	0.630	0.489	0.469	0.536	0.440	0.630	0.519	0.600	0.492	0.128	0.869	0.146	0.876	0.141	0.875
75	0.387	0.273	0.253	0.201	0.214	0.298	0.210	0.319	0.236	0.312	0.227	0.117	0.367	0.121	0.368	0.121	0.368
100	0.280	0.200	0.186	0.153	0.165	0.240	0.169	0.248	0.188	0.246	0.181	0.128	0.268	0.130	0.269	0.130	0.269
150	0.149	0.119	0.113	0.090	0.099	0.137	0.101	0.140	0.113	0.139	0.108	0.099	0.144	0.100	0.145	0.100	0.145
ρ=0.99																	
50	4.283	2.891	2.741	1.933	1.709	0.502	0.995	0.871	1.342	0.694	1.236	0.057	3.289	0.066	3.473	0.065	3.441
75	1.881	1.122	1.029	0.673	0.613	0.506	0.477	0.734	0.622	0.640	0.571	0.070	1.493	0.082	1.537	0.080	1.532
100	1.194	0.708	0.641	0.436	0.403	0.501	0.368	0.634	0.457	0.587	0.426	0.105	0.986	0.116	1.005	0.115	1.003
150	0.780	0.480	0.434	0.314	0.302	0.461	0.306	0.525	0.367	0.506	0.345	0.158	0.668	0.171	0.676	0.170	0.676
Estimated MSE when the intercept equals 1																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
100	0.297	0.261	0.256	0.227	0.235	0.277	0.235	0.283	0.253	0.282	0.246	0.209	0.292	0.214	0.293	0.213	0.293
150	0.147	0.132	0.129	0.115	0.127	0.142	0.124	0.143	0.131	0.143	0.128	0.124	0.145	0.125	0.145	0.125	0.145
200	0.094	0.086	0.084	0.075	0.090	0.092	0.081	0.092	0.086	0.092	0.084	0.084	0.093	0.085	0.093	0.085	0.093
300	0.058	0.054	0.054	0.049	0.062	0.057	0.052	0.057	0.054	0.057	0.053	0.054	0.057	0.055	0.057	0.055	0.057
ρ=0.95																	
100	0.881	0.749	0.731	0.619	0.586	0.625	0.588	0.686	0.667	0.669	0.638	0.311	0.842	0.341	0.849	0.335	0.848
150	0.384	0.311	0.299	0.247	0.244	0.344	0.268	0.354	0.298	0.352	0.287	0.243	0.371	0.250	0.372	0.249	0.372
200	0.267	0.225	0.216	0.180	0.189	0.249	0.200	0.253	0.218	0.252	0.211	0.201	0.260	0.204	0.260	0.204	0.260
300	0.160	0.139	0.135	0.113	0.114	0.154	0.126	0.156	0.138	0.155	0.133	0.137	0.157	0.138	0.157	0.138	0.157
ρ=0.99																	
100	4.322	3.501	3.410	2.831	2.636	0.997	2.151	1.398	2.524	1.256	2.414	0.207	3.728	0.261	3.858	0.249	3.833
150	1.835	1.406	1.338	1.006	0.916	0.863	0.849	1.075	1.030	1.006	0.964	0.276	1.590	0.327	1.630	0.318	1.624
200	1.176	0.847	0.800	0.576	0.521	0.746	0.539	0.854	0.662	0.821	0.613	0.330	1.033	0.368	1.051	0.363	1.048
300	0.649	0.459	0.427	0.308	0.295	0.505	0.332	0.543	0.391	0.530	0.367	0.306	0.585	0.322	0.590	0.320	0.590
Estimated MSE when the intercept equals 2																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
200	0.380	0.359	0.356	0.329	0.326	0.363	0.338	0.369	0.355	0.367	0.347	0.319	0.376	0.331	0.377	0.328	0.377
300	0.182	0.171	0.169	0.153	0.153	0.179	0.163	0.180	0.172	0.180	0.168	0.172	0.181	0.173	0.181	0.173	0.181
400	0.111	0.107	0.106	0.097	0.105	0.110	0.102	0.111	0.106	0.110	0.104	0.107	0.111	0.108	0.111	0.108	0.111
500	0.086	0.081	0.081	0.074	0.082	0.085	0.079	0.085	0.082	0.085	0.081	0.083	0.085	0.084	0.085	0.084	0.085
ρ=0.95																	
200	1.088	0.987	0.971	0.865	0.794	0.909	0.856	0.965	0.940	0.949	0.908	0.640	1.056	0.704	1.063	0.685	1.061
300	0.501	0.451	0.442	0.395	0.378	0.471	0.420	0.478	0.446	0.477	0.437	0.409	0.490	0.419	0.491	0.418	0.491
400	0.353	0.323	0.317	0.287	0.281	0.339	0.305	0.342	0.321	0.342	0.315	0.309	0.347	0.313	0.348	0.313	0.348
500	0.221	0.203	0.200	0.172	0.166	0.215	0.187	0.217	0.201	0.217	0.195	0.203	0.218	0.204	0.218	0.204	0.218
ρ=0.99																	
200	5.392	4.887	4.819	4.291	4.015	2.062	3.674	2.581	4.117	2.441	3.978	0.705	4.941	0.931	5.067	0.870	5.041
300	2.532	2.229	2.189	1.850	1.703	1.599	1.702	1.819	1.931	1.762	1.844	0.804	2.334	0.975	2.379	0.935	2.371
400	1.471	1.253	1.221	0.973	0.876	1.115	0.950	1.212	1.098	1.186	1.037	0.719	1.365	0.797	1.383	0.784	1.380
500	1.128	0.962	0.936	0.756	0.697	0.910	0.769	0.963	0.872	0.949	0.827	0.652	1.058	0.698	1.068	0.691	1.067

Table 2: Estimated MAE of the proposed estimators when p=2

Estimated MSE when the intercept equals 0																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
50	0.595	0.520	0.509	0.478	0.522	0.551	0.484	0.563	0.500	0.560	0.496	0.372	0.589	0.375	0.589	0.375	0.589
75	0.435	0.394	0.387	0.364	0.402	0.418	0.377	0.423	0.389	0.422	0.385	0.338	0.431	0.341	0.432	0.341	0.431
100	0.368	0.340	0.334	0.313	0.358	0.360	0.324	0.362	0.335	0.361	0.331	0.316	0.365	0.317	0.366	0.317	0.366
150	0.267	0.254	0.251	0.236	0.271	0.263	0.245	0.264	0.251	0.264	0.249	0.247	0.265	0.248	0.265	0.248	0.265
ρ=0.95																	
50	1.005	0.780	0.746	0.639	0.646	0.803	0.642	0.854	0.705	0.840	0.683	0.372	0.974	0.383	0.976	0.381	0.976
75	0.685	0.553	0.528	0.456	0.478	0.613	0.486	0.630	0.524	0.626	0.511	0.382	0.668	0.387	0.669	0.387	0.669
100	0.584	0.483	0.463	0.399	0.423	0.543	0.433	0.553	0.469	0.551	0.456	0.395	0.571	0.399	0.572	0.399	0.572
150	0.424	0.378	0.368	0.315	0.328	0.407	0.340	0.411	0.367	0.410	0.356	0.346	0.417	0.349	0.417	0.349	0.417
ρ=0.99																	
50	2.159	1.515	1.439	1.094	1.033	0.832	0.862	1.059	1.028	0.969	0.970	0.250	1.897	0.264	1.930	0.262	1.925
75	1.499	1.007	0.938	0.720	0.699	0.844	0.664	0.978	0.774	0.932	0.733	0.297	1.337	0.315	1.351	0.313	1.349
100	1.210	0.828	0.769	0.608	0.591	0.826	0.608	0.906	0.693	0.882	0.664	0.365	1.101	0.381	1.108	0.379	1.107
150	0.975	0.695	0.647	0.515	0.512	0.775	0.554	0.817	0.626	0.806	0.599	0.452	0.901	0.466	0.905	0.465	0.905
Estimated MSE when the intercept equals 1																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
100	0.564	0.521	0.513	0.475	0.494	0.548	0.492	0.553	0.515	0.552	0.506	0.475	0.560	0.479	0.560	0.478	0.560
150	0.419	0.396	0.391	0.365	0.386	0.413	0.381	0.414	0.395	0.414	0.389	0.387	0.416	0.388	0.417	0.388	0.417
200	0.331	0.317	0.314	0.294	0.316	0.328	0.306	0.329	0.316	0.329	0.312	0.315	0.330	0.316	0.330	0.316	0.330
300	0.264	0.257	0.256	0.241	0.259	0.262	0.249	0.263	0.256	0.263	0.252	0.257	0.263	0.257	0.263	0.257	0.263
ρ=0.95																	
100	0.940	0.823	0.804	0.708	0.699	0.837	0.730	0.864	0.789	0.857	0.766	0.595	0.920	0.613	0.922	0.610	0.922
150	0.669	0.590	0.576	0.501	0.501	0.639	0.540	0.646	0.580	0.645	0.564	0.539	0.658	0.545	0.659	0.545	0.659
200	0.567	0.516	0.505	0.447	0.459	0.549	0.479	0.553	0.508	0.552	0.496	0.495	0.559	0.498	0.559	0.498	0.559
300	0.437	0.407	0.401	0.357	0.353	0.429	0.381	0.431	0.404	0.430	0.395	0.405	0.432	0.407	0.432	0.407	0.432
ρ=0.99																	
100	2.061	1.631	1.578	1.303	1.240	1.163	1.175	1.342	1.336	1.286	1.281	0.507	1.885	0.558	1.910	0.548	1.906
150	1.443	1.156	1.107	0.897	0.848	1.058	0.880	1.151	1.001	1.124	0.954	0.599	1.340	0.639	1.351	0.633	1.350
200	1.196	0.940	0.898	0.722	0.689	0.985	0.749	1.037	0.849	1.023	0.810	0.658	1.120	0.685	1.126	0.682	1.126
300	0.896	0.722	0.689	0.555	0.543	0.802	0.609	0.824	0.677	0.818	0.650	0.628	0.850	0.640	0.853	0.639	0.853
Estimated MSE when the intercept equals 2																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
200	0.631	0.607	0.603	0.572	0.577	0.621	0.589	0.625	0.606	0.624	0.599	0.590	0.628	0.596	0.629	0.595	0.628
300	0.449	0.433	0.430	0.405	0.406	0.446	0.422	0.447	0.435	0.446	0.430	0.437	0.447	0.437	0.447	0.437	0.447
400	0.364	0.356	0.355	0.337	0.348	0.362	0.347	0.363	0.356	0.362	0.352	0.357	0.363	0.358	0.363	0.358	0.363
500	0.321	0.314	0.313	0.297	0.308	0.320	0.308	0.320	0.315	0.320	0.311	0.317	0.320	0.317	0.320	0.317	0.320
ρ=0.95																	
200	1.056	0.974	0.961	0.876	0.840	0.998	0.910	1.015	0.962	1.011	0.942	0.862	1.041	0.886	1.043	0.881	1.043
300	0.753	0.706	0.696	0.638	0.621	0.735	0.674	0.739	0.703	0.738	0.692	0.689	0.745	0.695	0.746	0.694	0.746
400	0.648	0.615	0.608	0.563	0.552	0.637	0.591	0.639	0.613	0.639	0.605	0.610	0.642	0.613	0.642	0.613	0.642
500	0.509	0.487	0.484	0.438	0.429	0.503	0.463	0.505	0.485	0.504	0.475	0.489	0.505	0.491	0.506	0.491	0.506
ρ=0.99																	
200	2.335	2.072	2.039	1.803	1.685	1.633	1.711	1.789	1.882	1.749	1.821	0.970	2.212	1.089	2.236	1.062	2.232
300	1.684	1.499	1.470	1.268	1.185	1.401	1.266	1.475	1.394	1.456	1.342	1.026	1.609	1.102	1.621	1.087	1.619
400	1.318	1.160	1.134	0.965	0.892	1.180	1.001	1.217	1.099	1.208	1.060	0.968	1.266	1.004	1.273	0.999	1.272
500	1.155	1.026	1.003	0.859	0.810	1.065	0.908	1.087	0.989	1.081	0.955	0.919	1.116	0.942	1.120	0.939	1.120

Table 3: Estimated MSE of the proposed estimators when p=4

Estimated MSE when the intercept equals 0																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
50	1.514	1.398	1.334	0.969	0.977	0.962	0.750	1.149	1.070	1.115	1.084	0.519	1.400	0.626	1.436	0.628	1.430
75	0.417	0.380	0.359	0.235	0.242	0.368	0.218	0.394	0.314	0.390	0.319	0.279	0.398	0.295	0.401	0.297	0.400
100	0.254	0.237	0.225	0.158	0.160	0.237	0.155	0.246	0.207	0.244	0.209	0.202	0.245	0.209	0.246	0.209	0.246
150	0.140	0.133	0.129	0.094	0.095	0.135	0.094	0.137	0.121	0.137	0.122	0.125	0.136	0.127	0.136	0.128	0.136
ρ=0.95																	
50	5.320	4.801	4.517	3.212	3.299	1.910	2.191	2.708	3.174	2.506	3.264	0.567	4.595	0.787	4.842	0.778	4.810
75	1.504	1.310	1.206	0.702	0.722	0.938	0.541	1.176	0.880	1.124	0.906	0.455	1.354	0.539	1.380	0.551	1.374
100	0.779	0.686	0.630	0.371	0.390	0.575	0.310	0.665	0.496	0.647	0.511	0.354	0.710	0.394	0.721	0.399	0.719
150	0.414	0.375	0.351	0.206	0.209	0.359	0.191	0.389	0.295	0.384	0.300	0.279	0.387	0.298	0.391	0.301	0.389
ρ=0.99																	
50	31.19	26.21	27.80	17.96	17.25	1.156	6.368	2.928	11.63	2.246	11.65	0.166	18.60	0.325	24.56	0.322	23.23
75	8.526	6.288	6.999	3.279	3.287	1.256	1.325	2.364	2.750	2.034	2.845	0.301	5.407	0.487	6.336	0.505	6.168
100	4.854	3.437	3.913	1.768	1.835	1.188	0.916	1.995	1.762	1.762	1.843	0.378	3.287	0.547	3.682	0.564	3.617
150	2.338	1.664	1.893	0.797	0.830	0.977	0.505	1.427	0.988	1.312	1.027	0.447	1.692	0.598	1.840	0.612	1.813
Estimated MSE when the intercept equals 1																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
100	1.360	1.324	1.303	1.103	1.101	1.149	0.959	1.254	1.206	1.235	1.211	0.883	1.301	1.036	1.328	1.036	1.325
150	0.427	0.415	0.406	0.334	0.337	0.408	0.313	0.418	0.385	0.416	0.387	0.377	0.416	0.389	0.418	0.389	0.418
200	0.230	0.224	0.220	0.180	0.182	0.224	0.173	0.227	0.211	0.227	0.212	0.216	0.225	0.219	0.226	0.219	0.226
300	0.118	0.116	0.115	0.098	0.099	0.117	0.095	0.118	0.112	0.118	0.112	0.115	0.117	0.116	0.117	0.116	0.117
ρ=0.95																	
100	4.772	4.594	4.486	3.720	3.677	2.639	2.884	3.316	3.784	3.160	3.794	1.339	4.386	1.873	4.564	1.833	4.533
150	1.310	1.241	1.193	0.892	0.907	1.077	0.746	1.189	1.044	1.166	1.057	0.807	1.229	0.909	1.251	0.912	1.248
200	0.783	0.744	0.721	0.526	0.529	0.687	0.456	0.738	0.640	0.729	0.648	0.566	0.742	0.617	0.751	0.621	0.749
300	0.373	0.357	0.347	0.255	0.260	0.354	0.239	0.365	0.322	0.363	0.325	0.329	0.359	0.339	0.361	0.340	0.361
ρ=0.99																	
100	21.93	20.72	20.12	15.99	16.02	3.081	9.616	5.526	13.64	4.723	13.86	0.779	17.25	1.496	19.50	1.459	19.19
150	7.506	6.869	6.515	4.457	4.503	2.427	2.504	3.830	4.362	3.426	4.490	0.951	5.717	1.548	6.445	1.545	6.346
200	4.471	4.097	3.886	2.612	2.677	1.997	1.605	2.837	2.758	2.608	2.845	1.019	3.544	1.485	3.897	1.498	3.841
300	2.215	2.024	1.902	1.244	1.298	1.410	0.909	1.741	1.465	1.659	1.512	0.913	1.840	1.163	1.956	1.167	1.941
Estimated MSE when the intercept equals 2																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
200	2.568	2.551	2.541	2.425	2.424	2.438	2.297	2.504	2.489	2.491	2.494	2.216	2.528	2.394	2.549	2.372	2.546
300	0.583	0.576	0.572	0.520	0.517	0.570	0.496	0.577	0.559	0.575	0.560	0.551	0.574	0.563	0.577	0.562	0.576
400	0.305	0.301	0.299	0.272	0.273	0.302	0.262	0.304	0.295	0.303	0.295	0.298	0.302	0.300	0.303	0.300	0.302
500	0.201	0.200	0.198	0.182	0.181	0.200	0.179	0.201	0.196	0.201	0.196	0.199	0.200	0.199	0.200	0.199	0.200
ρ=0.95																	
200	20.37	20.30	20.25	19.84	19.81	17.88	19.07	18.70	19.85	18.58	19.85	15.46	20.04	17.08	20.24	16.94	20.21
300	1.711	1.682	1.663	1.454	1.452	1.560	1.280	1.644	1.571	1.628	1.577	1.354	1.652	1.503	1.674	1.503	1.670
400	0.941	0.924	0.912	0.790	0.794	0.893	0.711	0.920	0.871	0.915	0.876	0.831	0.914	0.873	0.923	0.873	0.921
500	0.680	0.668	0.659	0.571	0.574	0.656	0.527	0.669	0.634	0.666	0.637	0.626	0.663	0.645	0.667	0.646	0.667
ρ=0.99																	
200	52.55	51.66	52.00	48.39	48.01	21.45	39.22	27.65	46.24	25.45	46.43	12.43	47.66	16.63	50.91	15.83	50.53
300	9.154	8.691	8.877	7.319	7.371	4.631	5.313	6.168	7.330	5.764	7.447	2.342	7.927	3.709	8.556	3.600	8.483
400	5.416	5.100	5.229	4.132	4.182	3.483	3.113	4.287	4.338	4.102	4.411	2.166	4.746	3.038	5.049	3.001	5.011
500	3.452	3.246	3.325	2.583	2.609	2.534	1.976	2.971	2.805	2.873	2.847	1.791	3.053	2.347	3.227	2.326	3.203

Table 4: Estimated MAE of the proposed estimators when p=4

Estimated MSE when the intercept equals 0																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
50	1.663	1.567	1.514	1.182	1.198	1.421	1.079	1.528	1.357	1.509	1.372	1.047	1.611	1.110	1.621	1.116	1.619
75	0.957	0.909	0.880	0.688	0.700	0.910	0.671	0.936	0.826	0.932	0.832	0.796	0.938	0.814	0.940	0.817	0.939
100	0.750	0.723	0.705	0.578	0.582	0.729	0.576	0.741	0.677	0.739	0.680	0.678	0.738	0.687	0.739	0.689	0.739
150	0.569	0.556	0.547	0.464	0.465	0.561	0.465	0.566	0.531	0.565	0.532	0.542	0.562	0.546	0.563	0.547	0.563
ρ=0.95																	
50	2.983	2.722	2.570	1.856	1.892	1.916	1.472	2.288	2.055	2.201	2.093	1.049	2.776	1.181	2.823	1.189	2.815
75	1.801	1.651	1.566	1.103	1.120	1.478	0.984	1.630	1.343	1.601	1.362	1.023	1.714	1.095	1.726	1.105	1.723
100	1.312	1.218	1.160	0.836	0.857	1.165	0.781	1.237	1.036	1.224	1.050	0.923	1.259	0.964	1.266	0.969	1.264
150	0.976	0.925	0.892	0.656	0.660	0.919	0.634	0.951	0.820	0.945	0.828	0.820	0.945	0.842	0.949	0.845	0.948
ρ=0.99																	
50	7.142	5.958	6.327	4.183	4.239	1.607	2.277	2.506	3.615	2.212	3.720	0.561	5.509	0.757	6.095	0.755	6.000
75	4.242	3.413	3.678	2.207	2.223	1.762	1.435	2.416	2.271	2.245	2.320	0.809	3.438	1.021	3.648	1.037	3.611
100	3.297	2.629	2.856	1.678	1.718	1.745	1.230	2.245	1.880	2.119	1.929	0.947	2.741	1.140	2.865	1.155	2.843
150	2.296	1.864	2.014	1.179	1.198	1.575	0.956	1.870	1.451	1.805	1.477	1.066	1.974	1.221	2.039	1.234	2.026
Estimated MSE when the intercept equals 1																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
100	1.511	1.478	1.458	1.272	1.273	1.436	1.202	1.475	1.399	1.468	1.405	1.304	1.485	1.358	1.492	1.360	1.490
150	0.955	0.939	0.928	0.819	0.818	0.938	0.799	0.947	0.904	0.946	0.905	0.907	0.944	0.917	0.946	0.918	0.945
200	0.729	0.719	0.712	0.633	0.636	0.722	0.623	0.726	0.698	0.725	0.700	0.710	0.723	0.714	0.724	0.714	0.724
300	0.526	0.521	0.517	0.474	0.475	0.523	0.471	0.525	0.511	0.525	0.512	0.519	0.522	0.520	0.523	0.521	0.523
ρ=0.95																	
100	2.768	2.670	2.610	2.177	2.167	2.257	1.906	2.479	2.366	2.432	2.378	1.683	2.660	1.897	2.692	1.897	2.686
150	1.690	1.634	1.594	1.310	1.322	1.565	1.203	1.629	1.490	1.616	1.500	1.375	1.641	1.439	1.651	1.443	1.650
200	1.301	1.263	1.239	1.016	1.020	1.238	0.952	1.273	1.170	1.267	1.176	1.142	1.269	1.178	1.275	1.182	1.274
300	0.935	0.914	0.900	0.755	0.756	0.915	0.729	0.927	0.868	0.924	0.871	0.885	0.919	0.896	0.922	0.897	0.921
ρ=0.99																	
100	6.082	5.725	5.558	4.373	4.400	2.606	3.082	3.532	4.276	3.270	4.349	1.258	5.256	1.745	5.583	1.733	5.533
150	3.988	3.730	3.584	2.711	2.736	2.436	2.027	3.011	2.886	2.870	2.934	1.509	3.487	1.913	3.658	1.917	3.632
200	3.070	2.878	2.768	2.086	2.100	2.213	1.676	2.573	2.326	2.493	2.357	1.595	2.743	1.894	2.845	1.905	2.828
300	2.230	2.102	2.020	1.514	1.538	1.854	1.300	2.030	1.769	1.992	1.795	1.522	2.039	1.688	2.090	1.693	2.082
Estimated MSE when the intercept equals 2																	
	ML	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12	K13	K14	K15	K16
ρ=0.85																	
200	1.865	1.852	1.844	1.742	1.737	1.825	1.677	1.846	1.816	1.842	1.817	1.759	1.848	1.806	1.854	1.803	1.853
300	1.107	1.100	1.094	1.029	1.024	1.098	1.007	1.103	1.083	1.102	1.083	1.086	1.100	1.093	1.102	1.093	1.101
400	0.827	0.822	0.818	0.774	0.772	0.823	0.761	0.826	0.813	0.825	0.813	0.819	0.823	0.821	0.824	0.821	0.824
500	0.684	0.681	0.678	0.646	0.641	0.682	0.641	0.683	0.674	0.683	0.674	0.679	0.681	0.680	0.681	0.681	0.681
ρ=0.95																	
200	3.367	3.334	3.313	3.102	3.097	3.011	2.886	3.165	3.200	3.139	3.203	2.551	3.297	2.834	3.327	2.813	3.323
300	1.877	1.853	1.838	1.666	1.661	1.812	1.560	1.848	1.784	1.842	1.787	1.718	1.845	1.779	1.854	1.780	1.853
400	1.415	1.398	1.386	1.258	1.255	1.386	1.195	1.402	1.354	1.399	1.357	1.349	1.395	1.371	1.400	1.372	1.399
500	1.222	1.209	1.200	1.095	1.093	1.204	1.053	1.214	1.176	1.212	1.179	1.183	1.207	1.195	1.211	1.196	1.210
ρ=0.99																	
200	8.066	7.841	7.923	7.099	7.062	4.591	5.692	5.711	7.025	5.386	7.069	2.546	7.410	3.659	7.779	3.500	7.733
300	4.347	4.165	4.237	3.620	3.638	3.312	3.015	3.749	3.774	3.647	3.811	2.431	4.026	2.996	4.166	2.963	4.148
400	3.452	3.307	3.365	2.834	2.854	2.889	2.425	3.153	3.030	3.097	3.058	2.346	3.230	2.717	3.318	2.707	3.306
500	2.766	2.651	2.695	2.256	2.266	2.446	1.973	2.609	2.459	2.575	2.479	2.120	2.609	2.361	2.668	2.357	2.659

Acknowledgements

This paper was written while the first author was on sabbatical leave (2010-2011). He is grateful to the Florida International University for awarding him the sabbatical leave which, gave him excellent research facilities. The second author would gratefully like to acknowledge the financial support provided by Sparbankernas Forskningsstiftelse.

References

Frisch, R. (1934). Statistical Confluence Analysis by Means of Complete Regression Systems. Publication 5 (Oslo: University Institute of Economics, 1934).

Greene, W., H. (1994) Accounting for excess of zeros and sample selection in Poisson and negative binomial regression model. Discussion Paper EC-94-10, Department of Economics, New York University.

Gråsjö, U. (2005). Accessibility to R&D and patent production. 37, Royal Institute of Technology, CESIS - Centre of Excellence for Science and Innovation Studies.

Hoerl, A.E. and Kennard, R.W. (1970a). Ridge regression: biased estimation for non-orthogonal Problems. Technometrics 12, 55-67.

Hoerl, A. E. and Kennard, R. W. (1970b). Ridge Regression: Application to Non-Orthogonal Problems. Technometrics 12, 69-82.

Kibria, B.M.G. (2003). Performance of some new ridge regression estimators. Communications in Statistics- Theory and Methods, 32, 419-435.

Lambert, D. (1992). Zero-Inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 34, 1-14.

Månsson, K. and Shukur, G. (2011a)**.** On Ridge Parameters in Logistic Regression", to appear in *Communications in Statistics, Theory and Methods.*

Månsson, K. and Shukur, (2011b). A Poisson Ridge Regression estimator. to appear in Economic Modelling.

Muniz, G., and Kibria, B. M. G. (2009). On some ridge regression estimators: An Empirical Comparisons. Communications in Statistics-Simulation and Computation, 38, 621- 630.

Kibria, B. M. Golam, Kristofer Månsson and Ghazi Shukur (2011). Performance of some logistic ridge regression parameters. To appear in *Computational Economics*.

Nelder, J. A. and Mead, R. 81965). A Simplex Method for Function Minimization. *Comput. J.* **7**, 308-313,

Schaefer, R.L., Roi, L. D. and Wolfe, R. A. (1984). A ridge logistic estimator. Communications in Statistics- Theory and Methods, 13, 99-113.