

New Liu Estimators for the Poisson Regression Model: Method and Application

By

Kristofer Månsson¹ B. M. Golam Kibria², Pär Sjölander¹ and Ghazi Shukur^{1,3}

¹ Department of Economics, Finance and Statistics, Jönköping University
Jönköping, Sweden

² Department of Mathematics and Statistics, Florida International University,
Miami, Florida, USA

³ Department of Economics and Statistics, Linnaeus University,
Växjö Sweden.

Abstract

A new shrinkage estimator for the Poisson model is introduced in this paper. This method is a generalization of the Liu (1993) estimator originally developed for the linear regression model and will be generalised here to be used instead of the classical maximum likelihood (ML) method in the presence of multicollinearity since the mean squared error (MSE) of ML becomes inflated in that situation. Furthermore, this paper derives the optimal value of the shrinkage parameter and based on this value some methods of how the shrinkage parameter should be estimated are suggested. Using Monte Carlo simulation where the MSE and mean absolute error (MAE) are calculated it is shown that when the Liu estimator is applied with these proposed estimators of the shrinkage parameter it always outperforms the ML. Finally, an empirical application has been considered to illustrate the usefulness of the new Liu estimators.

Key words: Estimation; MSE; MAE; Multicollinearity; Poisson; Liu; Simulation.

AMS Subject classification: Primary 62J07, Secondary 62F10.

1. Introduction

In the field of health, social, economics and physical sciences, the dependent variable often comes in the form of a non-negative integers or counts. In that situation one often apply the Poisson regression model which is usually estimated by maximum likelihood (ML) where the solution to a non-linear equation is found by applying iterative weighted least square (IWLS). This method has been shown in Månsson and Shukur (2011a,b) to be sensitive to multicollinearity and it becomes difficult to make a valid statistical inference since the mean squared error (MSE) becomes inflated. In those papers, a ridge regression estimator (RRE) was presented which was a generalization of that proposed for linear regression by Hoerl and Kennard (1970). In both papers it was shown that the RRE outperformed the ML.

The RRE is effective but as Liu (1993) pointed out it has the disadvantage that the estimated parameters are complicated non-linear functions of the ridge parameter k . Therefore, in this paper another shrinkage estimator for the Poisson model will be proposed which is a generalization of the method proposed for linear regression by Liu (1993). The advantage of this method is that the estimators are a linear function of the shrinkage parameter d . For this reason, this shrinkage estimator has become more popular during recent years (see for examples, Akdeneiz and Kaciranlar (1995), Kaciranlar (2003) and Alheety and Kibria (2009) among others).

The purpose of this paper is to solve the problem of an inflated MSE of the ML estimator by applying a Liu estimator. Furthermore, we derive the optimal value of the shrinkage parameter and based on this value we suggest some methods of how the shrinkage parameter should be estimated. In a Monte Carlo study we evaluate the performance of the ML and the Liu estimator applied with the suggested estimators of the shrinkage parameter. The performance criteria used in the simulation study is the MSE and mean absolute error (MAE). In our simulation, factors including the degree of correlation, the sample size and the number of explanatory variables are varied. Finally, an empirical example has been considered to illustrate the benefit of the Liu estimator. In this application, the effect of the usage of cars and trucks on the number of killed pedestrians in different counties in Sweden is investigated

This paper is organized as follows: In Section 2, the statistical methodology is described. In Section 3, the design of the Monte Carlo experiment is presented and the result from the simulation study is discussed. An application is presented in Section 4. Finally, a brief summary and conclusions is given in section 5.

2. Methodology

2.1 Poisson regression

The Poisson regression model is a benchmark model when the dependent variable (y_i) comes in the form of counts data and distributed as $Po \mu_i$, where $\mu_i = \exp x_i \beta$, x_i is the i th row of X which is a $n \times p+1$ data matrix with p explanatory variables and β is a $p+1 \times 1$ vector of coefficients. The log likelihood of this model may be written as:

$$l(\mu; y) = \sum_{i=1}^n \exp x_i \beta + \sum_{i=1}^n y_i \log \exp x_i \beta + \log \left(\prod_{i=1}^n y_i! \right). \quad (2.1)$$

The most common method to maximize the likelihood function is to apply the IWLS algorithm:

$$\hat{\beta}_{ML} = (X' \hat{W} X)^{-1} X' \hat{W} \hat{z}, \quad (2.2)$$

where $\hat{W} = \text{diag } \hat{\mu}_i$ and \hat{z} is a vector where the i th element equals $\hat{z}_i = \log \hat{\mu}_i + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i}$.

The MSE of this estimator equals:

$$E L_{ML}^2 = E (\hat{\beta}_{ML} - \beta)' (\hat{\beta}_{ML} - \beta) = \text{tr} (X' W X)^{-1} = \sum_{j=1}^J \frac{1}{\lambda_j}, \quad (2.3)$$

where λ_j is the j th eigenvalue of the $X' W X$ matrix. When the explanatory variables are highly correlated the weighted matrix of cross-products, $X' W X$, is ill-conditioned which leads to instability and high variance of the ML estimator. In that situation it is very hard to interpret the estimated parameters since the vector of estimated coefficients is on average too long. By noting that the IWLS algorithm approximately minimizes the weighted sum of squared error (WSSE) one may apply a generalization of the Liu (1993) estimator for linear regression instead:

$$\hat{\beta}_d = (X' \hat{W} X + dI)^{-1} X' \hat{W} X + dI \hat{\beta}_{ML} \quad (2.4)$$

For this estimator we have replaced the matrix of cross-products used in the Liu (1993) estimator with the weighted matrix of cross-products and the ordinary least square estimator (OLS) of β with the ML estimator. The MSE of the Liu estimator equals:

$$\begin{aligned}
MSE \hat{\beta}_d &= E L_d^2 = E \hat{\beta}_d - \beta' \hat{\beta}_d - \beta = \\
&E \left[\hat{\beta}_{ML} - \beta' Z' Z \hat{\beta}_{ML} - \beta \right] + Z \beta - \beta' Z \beta - \beta = \\
&tr \left[\hat{\beta}_{ML} - \beta' \hat{\beta}_{ML} - \beta Z' Z \right] + k^2 \beta' X' W X + k I^{-2} \beta = \\
&\sum_{j=1}^J \frac{\lambda_j + d^2}{\lambda_j \lambda_j + 1} + d - 1 \sum_{j=1}^J \frac{\alpha_j^2}{\lambda_j + 1} =
\end{aligned} \tag{2.5}$$

where α_j^2 is defined as the j th element of $\gamma\beta$ and γ is the eigenvector defined such that $X'WX = \gamma'\Lambda\gamma$, where Λ equals $diag \lambda_j$. In order to show that there exist a value of d bounded between zero and one so that $MSE \hat{\beta}_d < MSE \hat{\beta}_{ML}$, we start by taking the first derivative of equation (2.5) with respect to d :

$$g' d = 2 \sum_{j=1}^J \frac{\lambda_j + d}{\lambda_j \lambda_j + 1} + 2 d - 1 \sum_{j=1}^J \frac{\alpha_j^2}{\lambda_j + 1} \tag{2.6}$$

and then by inserting the value one in equation (2.6) we get:

$$g' d = 2 \sum_{j=1}^J \frac{1}{\lambda_j \lambda_j + 1}, \tag{2.7}$$

which is greater than zero since $\lambda_j > 0$. Hence, there exists a value of d that lies between zero and one so that $MSE \hat{\beta}_d < MSE \hat{\beta}_{ML}$. Furthermore, the optimal value of any individual parameter d_j can be found by setting equation (2.6) to zero and solve for d_j . Then it may be shown that

$$d_j = \frac{\alpha_j^2 - 1}{\frac{1}{\lambda_j} + \alpha_j^2}, \tag{2.8}$$

corresponds to the optimal value of the shrinkage parameter. Hence, the optimal value of d_j is negative when α_j^2 is less than one and positive when it is greater than one. However, just as in Liu (1993) the shrinkage parameter will be limited to take on values only between zero and one.

2.2 Estimating the shrinkage parameter

The value of d may only take on values between zero and one and there does not exist a definite rule of how to estimate it. However, in this paper some methods will be proposed that are based on the work for linear ridge regression by for instance Hoerl and Kennard (1970), Kibria (2003) and Khalaf and Shukur (2005). As in those papers, the shrinkage parameter, d_j , will be estimated by a single value \hat{d} . The first estimator is the following:

$$D1 = \max \left(0, \frac{\hat{\alpha}_{\max}^2 - 1}{\frac{1}{\hat{\lambda}_{\max}} + \hat{\alpha}_{\max}^2} \right),$$

where we define $\hat{\alpha}_{\max}^2$ and $\hat{\lambda}_{\max}$ to be the maximum element of $\hat{\alpha}_j^2$ and $X'WX$, respectively.

Replacing the values of the unknown parameters with the maximum value of the unbiased estimators is an idea taken from Hoerl and Kennard (1970). However, for the Liu estimator another maximum operator is also used that will ensure that the estimated value of the shrinkage parameter is not negative. Furthermore, the following estimators will be used:

$$D2 = \max \left(0, \text{median} \left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2} \right) \right) \quad D3 = \max \left(0, \frac{1}{p} \sum_j \frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2} \right)$$

$$D4 = \max \left(0, \max \left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2} \right) \right) \quad D5 = \max \left(0, \min \left(\frac{\hat{\alpha}_j^2 - 1}{\frac{1}{\hat{\lambda}_j} + \hat{\alpha}_j^2} \right) \right).$$

Using the average value and the median is very common when estimating the shrinkage parameter for ridge parameter and the D2 and D3 estimators has direct counterparts in equation (13) and (15) in Kibria (2003). Using other quantiles such as the maximum value, was successfully applied in Khalaf and Shukur (2005) and the idea behind the D4 and D5 estimator are taken from those papers.

2.3 Judging the performance of the estimators

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

$$MSE = \frac{\sum_{i=1}^R (\hat{\beta} - \beta)' (\hat{\beta} - \beta)}{R}, \quad (2.9)$$

and the MAE as:

$$MAE = \frac{\sum_{i=1}^R |\hat{\beta}_i - \beta_i|}{R} \quad (2.10)$$

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

3. The Monte Carlo simulation

This section consists of a brief description of how the data is generated together with a discussion of our findings.

3.1 The Design of the Experiment

The dependent variable of the Poisson regression model is generated using pseudo-random numbers from the $Po \mu_i$ distribution where

$$\mu_i = \exp \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p. \quad (3.1)$$

The parameter values in equation (3.1) are chosen so that $\sum_{j=1}^p \beta_j^2 = 1$ and $\beta_1 = \dots = \beta_p$. To be able to generate data with different degrees of correlation we use the following formula to obtain the regressors:

$$x_{ij} = 1 - \rho^2 \quad z_{ij} + \rho z_{ip}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p \quad (3.2)$$

where z_{ij} are pseudo-random numbers generated using the standard normal distribution and ρ^2 represents the degree of correlation (see, Kibria 2003 and Muniz and Kibria 2009 among others). In the design of the experiment three different values of ρ^2 corresponding to 0.85, 0.95 and 0.99 are considered. To reduce eventual start-up value effects we discard the first 200 observations.

In the design of the experiment the factors n and p are also varied. Since the ML estimators are consistent, increasing the sample size is assumed to lower MSE and MAE while p is assumed to increase the instability of $X'WX$ and lead to an increase of both measures of performance. We use sample sizes corresponding to 15, 20, 30, 50 and 100 degrees of freedoms ($df=n-p$) and number of regressors p equals to 2 and 4.

3.2 Results Discussion

The estimated MSE and MAE for $p=2$ and 4 are presented in Tables 1 and 2 respectively. It is evident from these tables that the degree of correlation and the number of explanatory variables inflate both the MSE and MAE while increasing the sample size leads to a decrease of both measures of performance. We can also see that the MSE increases more when considering the MSE instead of the MAE criteria. Hence, the gain of applying Liu is larger in terms of MSE than MAE. Furthermore, when looking at both measures of performance we can see that the estimator D5 is always either the shrinkage parameter that minimizes the MSE and MAE or it is very close to the shrinkage parameter that minimizes these loss functions.

Table 1: Estimated MSE and MAE of the estimators when $p=2$

Estimated MSE							Estimated MAE					
	ML	D1	D2	D3	D4	D5	ML	D1	D2	D3	D4	D5
$\rho^2=0.85$												
50	0.514	0.244	0.216	0.216	0.251	0.210	0.742	0.517	0.499	0.499	0.520	0.495
75	0.350	0.198	0.183	0.183	0.201	0.179	0.626	0.478	0.468	0.468	0.479	0.465
100	0.181	0.125	0.122	0.122	0.125	0.122	0.459	0.387	0.385	0.385	0.387	0.385
150	0.098	0.080	0.080	0.080	0.080	0.080	0.339	0.308	0.308	0.308	0.308	0.308
200	0.038	0.035	0.035	0.035	0.035	0.035	0.215	0.207	0.207	0.207	0.207	0.207
$\rho^2=0.95$												
50	1.755	0.581	0.442	0.442	0.656	0.375	1.391	0.730	0.655	0.655	0.761	0.621
75	1.044	0.379	0.302	0.302	0.396	0.279	1.104	0.633	0.589	0.589	0.644	0.575
100	0.621	0.273	0.250	0.250	0.279	0.239	0.847	0.552	0.538	0.538	0.556	0.531
150	0.324	0.186	0.183	0.183	0.187	0.181	0.625	0.477	0.474	0.474	0.477	0.473
200	0.135	0.103	0.103	0.103	0.103	0.103	0.409	0.359	0.359	0.359	0.359	0.359
$\rho^2=0.99$												
50	10.49	3.366	3.277	3.277	4.887	2.453	3.345	1.439	1.305	1.305	1.748	1.094
75	6.396	1.922	1.764	1.764	2.679	1.356	2.685	1.150	1.009	1.009	1.338	0.887
100	3.443	1.010	0.809	0.809	1.221	0.670	2.001	0.870	0.770	0.770	0.951	0.712
150	1.767	0.541	0.445	0.445	0.582	0.414	1.473	0.718	0.667	0.667	0.742	0.648
200	0.774	0.310	0.298	0.298	0.318	0.290	0.984	0.597	0.590	0.590	0.602	0.585

Table 2: Estimated MSE and MAE of the estimators when $p=4$

Estimated MSE							Estimated MAE					
	ML	D1	D2	D3	D4	D5	ML	D1	D2	D3	D4	D5
$\rho^2=0.85$												
50	1.810	0.861	0.645	0.572	1.079	0.465	1.846	1.260	1.109	1.082	1.341	1.049
75	0.916	0.483	0.386	0.374	0.533	0.366	1.383	1.025	0.956	0.948	1.052	0.943
100	0.424	0.286	0.268	0.268	0.290	0.267	0.978	0.815	0.800	0.800	0.818	0.800
150	0.194	0.158	0.157	0.157	0.158	0.157	0.671	0.613	0.612	0.612	0.613	0.612
200	0.072	0.067	0.067	0.067	0.067	0.067	0.417	0.404	0.404	0.404	0.404	0.404
$\rho^2=0.95$												
50	5.658	2.191	1.627	1.231	3.592	0.695	3.253	1.878	1.521	1.395	2.260	1.225
75	3.165	1.324	0.877	0.753	1.780	0.606	2.571	1.607	1.337	1.276	1.786	1.207
100	1.513	0.756	0.544	0.520	0.832	0.501	1.818	1.270	1.133	1.119	1.312	1.109
150	0.618	0.387	0.349	0.349	0.394	0.348	1.187	0.949	0.920	0.920	0.953	0.920
200	0.226	0.181	0.180	0.180	0.181	0.180	0.733	0.662	0.661	0.661	0.662	0.661
$\rho^2=0.99$												
50	27.75	9.273	8.294	7.180	20.554	2.621	7.329	3.530	3.052	2.810	5.577	1.553
75	17.44	6.077	4.952	3.854	11.940	1.424	5.880	2.988	2.446	2.168	4.274	1.390
100	7.538	2.631	1.740	1.370	4.320	0.750	4.050	2.173	1.699	1.555	2.725	1.279
150	3.471	1.435	0.870	0.730	1.795	0.616	2.791	1.687	1.362	1.297	1.852	1.236
200	1.321	0.693	0.520	0.510	0.728	0.508	1.739	1.241	1.127	1.122	1.263	1.121

4. Empirical Application

To illustrate the performance of different estimators we consider data that are taken from the Department of Transport Analysis in Sweden¹. The data is for different counties (the total number is 22) in Sweden for the year 2010. The number of pedestrians were killed is used as a dependent variable and the number of kilometers driven by cars (x_1) and trucks (x_2) respectively are considered as independent variables. A likelihood ratio test has been used to determine whether the dependent variable follows a Poisson distribution or not. We found that the test statistic cannot be rejected at the one percent level of significance. Furthermore, the bivariate correlation between the two regressors is 0.92, so there is a problem of multicollinearity.

The proposed different Liu estimators are estimated using IWLS algorithms in R. Furthermore, bootstrap technique was applied in order to calculate the standard errors of the estimated parameters. The results are presented in Table 3. From this table it is clear that the bootstrapped standard errors are the highest for ML and the lowest for D5. This supported the simulation results in section 3. Moreover, we can also see a positive relationship between the explanatory variables and the number of pedestrians killed. This is expected since both regressors show the usage of cars and trucks respectively which are supposed to increase the number of killed pedestrians. But due to the multicollinearity problem the vector of coefficients becomes too long and the result obtained from the ML estimation method exaggerates this positive effect. Instead, we should look at the much lower estimated values of the coefficients obtained from D5 since the simulations shows that this estimator has the lowest estimated MSE and the empirical application shows that it has the lowest bootstrapped standard errors.

Table 3: The estimated parameters and the standard errors of the different estimators

ML		D1		D2		D3		D4		D5	
x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
3.461	10.237	1.886	5.349	1.994	5.683	1.994	5.683	3.068	9.018	0.919	2.347
(15.75)	(8.79)	(10.45)	(6.03)	(9.72)	(5.84)	(9.72)	(5.84)	(13.66)	(7.45)	(7.16)	(4.85)

Note: The standard errors are in parenthesis.

¹ The data is publically available on the webpage of the Department of Transport Analysis, www.trafa.se. The data is available from the authors upon request.

5. Conclusions

In this paper, the shrinkage estimator developed by Liu (1993) for the linear regression model has been extended for the Poisson regression model. This estimator is proposed in order to reduce the inflation of the variance of the ML estimator caused by multicollinearity. The Liu and the ML estimators are evaluated by means of Monte Carlo simulations. Both MSE and MAE are used as a performance criteria and factors including the degree of correlation, the sample size and the number of explanatory variables are varied. Both measures of performance show that the proposed Liu estimators are better than ML in the sense of smaller MSE and MAE. We also observed that the estimator D5 is often the shrinkage parameter that minimizes the estimates MSE and MAE. The benefit of proposed Liu estimators is shown by an example. Both the results from the simulation study and the empirical application shows that the proposed D5 should be the estimator of the Liu parameter to recommend for practitioners.

Acknowledgements

This paper was written while Dr. Kibria was visiting Professor Ghazi Shukur in the Department of Economics and Statistics, Linnaeus University, Växjö, Sweden during May-June 2011. He acknowledges the excellent research support provided by Linnaeus University, Växjö, Sweden.

References

- Akdeniz, F. and S. Kaciranlar (1995). On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. *Communications in Statistics-Theory and Methods*, **24**, 1789-1797.
- Alheety, M. I. and B. M. G. Kibria (2009). On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors. *Surveys in Mathematics and its Applications*, **4**, 155-167.
- Alkhamisi, M. and G. Shukur (2007). A Monte-Carlo Study of Recent Ridge Parameters. *Communications in Statistics- Theory and Methods*, **36**, 535-547.
- Hoerl, A.E. and R.W. Kennard (1970). Ridge regression: biased estimation for non-orthogonal Problems. *Technometrics*, **12**, 55-67.
- Liu, K. (1993). A new class of biased estimate in linear regression. *Communications in Statistics-Theory and Methods*, **22**, 393-402.
- Kaciranlar, S. (2003). Liu estimator in the general linear regression model. *Journal of Applied Statistical Science*, **13**, 229-234.
- Khalaf, G. and Shukur, G. (2005). Choosing Ridge Parameter for Regression Problems. *Communications in Statistics, Theory and Methods*, **34**, Issue 5, 1177-1182.
- Kibria B.M.G. (2003). Performance of some new ridge regression estimators *Communications in Statistics, Theory and Methods*, **32**, 419-435.
- Muniz, G. and Kibria, B. M. G. (2009). On some ridge regression estimators: An Empirical Comparison. *Communications in Statistics-Simulation and Computation* 38:621-630.
- Månsson, K. and Shukur, G. (2011a). A Poisson Ridge Regression estimator. to appear in *Economic Modeling*.
- Månsson, K. and Shukur, G. (2011b). On Ridge Parameters in Logistic Regression, to appear in *Communications in Statistics, Theory and Methods*.