

11-1-2003

Variable Selection for Poisson Regression Model

Felix Famoye

Central Michigan University, felix.famoye@cmich.edu

Daniel E. Rothe

Alpena Community College, rothed@alpenacc.edu

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Famoye, Felix and Rothe, Daniel E. (2003) "Variable Selection for Poisson Regression Model," *Journal of Modern Applied Statistical Methods*: Vol. 2 : Iss. 2 , Article 11.

DOI: 10.22237/jmasm/1067645460

Variable Selection for Poisson Regression Model

Cover Page Footnote

The first author acknowledges the support received from Central Michigan University FRCE Committee under Grant #48136.

Variable Selection for Poisson Regression Model

Felix Famoye
Department of Mathematics
Central Michigan University

Daniel E. Rothe
Alpena Community College

Poisson regression is useful in modeling count data. In a study with many independent variables, it is desirable to reduce the number of variables while maintaining a model that is useful for prediction. This article presents a variable selection technique for Poisson regression models. The data used is log-linear, but the methods could be adapted to other relationships. The model parameters are estimated by the method of maximum likelihood. The use of measures of goodness-of-fit to select appropriate variables is discussed. A forward selection algorithm is presented and illustrated on a numerical data set. This algorithm performs as well if not better than the method of transformation proposed by Nordberg (1982).

Key words: Transformation, goodness-of-fit, forward selection, R-square

Introduction

Regression models using count data have a wide range of applications in engineering, medicine, and social sciences. Other forms of regression such as logistic regression are well established in various social science and medical fields. For example, in epidemiology, researchers study the relationship between the chance of occurrence of a disease and various suspected risk factors. However, when the outcomes are counts, Signorini (1991) and others point out that Poisson regression gives adequate results.

The social sciences often perform studies that involve count data. Sociology, psychology, demography, and economics all perform studies using the type of data that can make use of the Poisson regression model. Sociology applications involve situations where researchers wish to predict an individual's

behavior based on a particular group of observed characteristics and experiences. D'Unger et al. (1998) examined categories of criminal careers using Poisson latent class regression models. They assert that Poisson regression models are appropriate for modeling delinquent behavior and criminal careers.

Gourieroux et al. (1984) and Cameron and Trivedi (1986) described the use of Poisson regression in economics applications such as the daily number of oil tankers' arrivals in a port, the number of accidents at work by factory, the number of purchases per period, the number of spells of unemployment, the number of strikes in a month, or the number of patents applied for and received by firms. Gourieroux et al. (1984) concluded that the use of Poisson regression model is justified in a situation where the dependent variable consists of counts of the occurrence of an event during a fixed time period.

Christiansen and Morris (1997) listed applications of Poisson regression in a variety of fields. Poisson regression has been used in literary analysis of Shakespeare's works and the Federalist Papers, Efron and Thisted (1976). Home run data has been analyzed using these types of methods, Albert (1992). Poisson regression and count data in general are very important in a wide range of fields and thus deserve special attention. Often these models

Felix Famoye, Department of Mathematics, Central Michigan University, Mt. Pleasant, MI, 48859. E-mail: felix.famoye@cmich.edu. Daniel E. Rothe, Alpena Community College, 666 Johnson Street, Alpena, MI, 49707. Email: rothed@alpenacc.edu. The first author acknowledges the support received from Central Michigan University FRCE Committee under Grant #48136.

involve many independent variables. Hence there is a need to consider variable selection for the Poisson regression model.

Variable selection techniques are well known for linear regression. See for example Efroymson (1960). Beale (1970) summarizes the various familiar methods: forward, backward, stepwise, and several other methods. Krall et al. (1975) discussed a forward selection technique for exponential survival data. They used the likelihood ratio as the criterion for adding significant variables. Greenberg et al. (1974) discussed a backward selection and use a log likelihood ratio step-down procedure for elimination of variables. For other nonlinear regressions and Poisson regression in particular, little is available in the literature.

Nordberg (1982) considered a certain data transformation in order to change the variable selection problem for a general linear model including the Poisson regression model into a variable selection problem in an ordinary unweighted linear regression model. Thus, ordinary linear regression variable selection software can be used.

In this article, we provide the Poisson regression model and describe some goodness-of-fit statistics. These statistics will be used as selection criteria for the variable selection method. A variable selection algorithm is described. We present the results of a simulation study to compare the variable selection algorithm with the method suggested by Nordberg (1982). The algorithm is illustrated with a numerical example and it is compared with the method suggested by Nordberg. Finally, we give some concluding remarks.

Poisson Regression Model and Goodness-of-fit Measures

The Poisson regression model assumes the response variable y_i , which is a count, has a Poisson distribution given by

$$P(y_i; \mu_i) = \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!}, y_i = 0, 1, 2, \dots \quad (1)$$

and $\mu_i = \mu_i(x_{ij}) = \exp\left(\sum_{j=0}^k \beta_j x_{ij}\right)$, where

x_{ij} ($j = 0, 1, \dots, k$ and $x_{i0} = 1$) are independent variables, and β_j ($j = 0, 1, 2, \dots, k$) are regression parameters. The mean and variance of y_i are equal and this is given by

$$E(y_i | x_{ij}) = V(y_i | x_{ij}) = \mu_i, \quad (2)$$

$$i = 1, 2, \dots, n, \text{ and } j = 0, 1, \dots, k.$$

Throughout this article, a log linear relationship

$$\mu_i = \exp\left(\sum_{j=0}^k \beta_j x_{ij}\right) \text{ will be considered.}$$

However, the results can be modified to accommodate other types of relationships. Frome et al. (1973) described the use of the maximum likelihood (ML) method to estimate the unknown parameters for the Poisson regression model.

Several measures of goodness-of-fit for the Poisson regression model have been proposed in the literature. The Akaike information criterion (AIC) is a commonly used measure (Akaike, 1973). It is defined as

$$AIC = -\log L + (k + 1) \quad (3)$$

where $k + 1$ is the number of estimated parameters and L is the likelihood function. The smaller the value of the AIC statistic, the better the fit of the model. The log likelihood could be used as a measure of goodness-of-fit. However, the AIC criterion also includes k as an adjustment for the number of independent variables, so that a model with many variables included is not necessarily better using this statistic.

Merkle and Zimmermann (1992) suggested some measures similar to the R^2 statistic for linear regression. They define

$$R_D^2 = \frac{l(\hat{\mu}_i) - l(\bar{y})}{l(y_i) - l(\bar{y})} \quad (4)$$

where

$$l(\hat{\mu}_i) = \sum_{i=1}^n (y_i \log \hat{\mu}_i - \hat{\mu}_i - \log y_i!),$$

$$l(\bar{y}) = \sum_{i=1}^n (y_i \log \bar{y} - \bar{y} - \log y_i!)$$

and

$$l(y_i) = \sum_{i=1}^n (y_i \log y_i - y_i - \log y_i!).$$

The quantity R_D^2 measures the goodness-of-fit by relating the explained increase in the log-likelihood to the maximum increase possible. The interpretation is that higher R_D^2 indicates a better fit from the model. The numerator of R_D^2 is the deviance statistic. Cameron and Windmeijer (1996) analyzed R^2 measures for count data. They establish five criteria for judging various R^2 measures. Among all R^2 measures considered, only the R_D^2 defined by Merkle and Zimmermann (1992) satisfies all the five criteria.

Selection Criteria Statistics

Variable selection procedures need criteria for adding significant variables. We propose two selection criteria statistics (SCS). The first SCS is the Akaike information criterion (AIC) defined earlier. The smaller the value of the AIC statistic, the better the fit of the model.

The second SCS is a modification of the R_D^2 suggested by Cameron and Windmeijer (1996) by taking the number of parameters into account. We define R_{adj}^2 as

$$R_{adj}^2 = \frac{\sum_{i=1}^n [y_i \log(\hat{\mu}_i / \bar{y}) - (\hat{\mu}_i - \bar{y})]}{\sum_{i=1}^n y_i \log(y_i / \bar{y})} \cdot \frac{(n-1)}{(n-k-1)} \quad (5)$$

where n is the sample size and k is the number of independent variables. Either of the selection criteria statistics in (3) and (5) can be used to

determine which variable to add in the selection procedure. These variable selection criteria measures are adjusted to include the number of parameters. In this way, an additional variable being added to the model may not necessarily result in an improvement to the measure.

Selection Algorithm

The transformation suggested by Nordberg (1982) for log-linear Poisson regression model takes the form

$$u_{ij} = x_{ij} \sqrt{\hat{\mu}_i}, \text{ where } j = 0, 1, 2, \dots, k, \text{ and } i = 1, 2, \dots, n \quad (6)$$

$$z_i = \left(\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \right) + \sum_{j=0}^k \hat{\beta}_j u_{ij} \quad (7)$$

where $\hat{\mu}_i$'s are the estimates of the predicted values from the full Poisson regression model. The variable selection procedure is as follows. Compute the ML estimate of β in the full Poisson regression model. Transform the data using (6) and (7). Perform variable selection on the linear model with z_i as the dependent variable and u_{ij} as the independent variables. Identify the subset of the u_{ij} variables that is selected and choose the corresponding x_{ij} variables. This gives the Poisson regression sub-model. Compute the maximum likelihood estimate for the Poisson regression on the chosen x_{ij} variables. This gives the final result of variable selection through transformation.

Nordberg (1982) indicated that the success of this technique depends on the accuracy of the approximation of the log-likelihood function given by

$$\log L(\underline{\beta}) \approx \log L(\underline{\hat{\beta}}) - (Q(\underline{\beta}) - Q(\underline{\hat{\beta}})) / 2, \quad (8)$$

where $Q(\underline{\beta})$ is given by

$$Q(\underline{\beta}) = \sum_{i=1}^n \left(z_i - \sum_{j=0}^k \beta_j u_{ij} \right)^2.$$

The error in (8) is given by

$$E = \frac{1}{6} \sum_{i=1}^n \frac{1}{\sqrt{\hat{\mu}_i}} \left(\sum_{j=0}^k (\beta_j - \hat{\beta}_j) u_{ij} \right)^3. \quad (9)$$

Nordberg (1982) concludes that the approximation is adequate even when 30% of the $\hat{\mu}_i$ are less than or equal to 4. However it is not clear what would happen to a case with say 70% of the $\hat{\mu}_i$ are less than or equal to 4. We note here that Nordberg did not run simulations on such cases.

Forward Selection Algorithm

The forward selection program begins by finding all possible regression models with one variable. The one with the best selection criteria statistic is chosen as the best one variable model. Once the best one variable model has been chosen, all models with the first variable and one additional variable are calculated and the one with the best selection criteria statistic is chosen. In this way, a two variable model is chosen. The process continues to add variables until the asymptotically normal Wald type “t”-value for an added variable is not significant. The process then stops and returns the previous acceptable model.

The selection criteria statistics (SCS) and a test of significance of each variable are used to determine which variable to enter. The following is the algorithm:

[Initialize: k = number of independent variables, α = significance level]

1. $\nu \leftarrow 1$
2. Fit k Poisson regression models with the intercept and ν independent variable
3. Select the model with the optimal SCS. Let x_i be the independent variable chosen and β_i be its parameter.
4. If the asymptotically normal Wald type “t”-value associated with β_i is significant at level α ,
 - Retain Poisson regression model with independent variable x_i and go to 5.
 - else

- Return “No variables are significant” and Stop.
5. Do while ($k \geq 2$)
 - $\nu \leftarrow \nu + 1$
 - $k \leftarrow k - 1$
 - Fit k Poisson regression models each with the intercept and ν independent variables. [The model includes all previously selected x_i ’s and one new $x_j, j = 1, 2, 3, \dots, k$]
 - Select the model with the optimal SCS. Let x_{new} be the independent variable added and β_{new} be its parameter.
 - If the asymptotically normal Wald type “t”-value for β_{new} is not significant at level α ,
 - $\nu \leftarrow \nu - 1$
 - go to 6, else
 - add x_{new} to the Poisson regression model
 - Continue
 6. The forward selection selects ν independent variables. Deliver the parameter estimates, t -values, and goodness-of-fit statistics for the selected model.

Simulation Study

In order to compare the proposed method with the method proposed by Nordberg (1982), we conduct a simulation study. The Poisson regression model in (1) is generated and both methods were used for variable selection.

We generated a set of x -data consisting of n ($n = 100, 250, 500,$ and 1000) observations on eight explanatory variables $x_{ij}, i = 1, 2, \dots, n$ and $j = 0, 1, 2, \dots, 7$, where $x_{i0} = 1$ (a constant term). The variables $x_{i1}, x_{i2}, \dots, x_{i7}$ were generated as uncorrelated standard normal variates. All simulations were done using computer programs written in Fortran codes and the Institute of Mathematical Statistics Library (IMSL) is used.

The parameter vector $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_7)$ used in the simulation study is chosen in such a way that $\beta_5 = \beta_6 = \beta_7 = 0$, while $\beta_0, \beta_1, \beta_2, \beta_3,$ and β_4 are non-zero. For all simulations, we

chose $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0.2$ and six different values of β_0 . We consider the six values $\beta_0 = -1.0, -0.5, 1.5, 1.7, 2.0,$ and 3.0 . These values were chosen so that certain percentages of fitted values $\hat{\mu}_i$ will be less than or equal to 4.0. When $\beta_0 = -1.0$ or -0.5 , all fitted values $\hat{\mu}_i$ from the Poisson regression model are less than or equal to 4.0. For $\beta_0 = 1.5$, about 40% of the fitted values $\hat{\mu}_i$ are less than or equal to 4.0. When $\beta_0 = 1.7$, about 20% of the fitted values $\hat{\mu}_i$ are less than or equal to 4.0, and for $\beta_0 = 3.0$, almost all fitted values $\hat{\mu}_i$ exceed 4.0.

Using the β -vector and $x_{i0}, x_{i1}, x_{i2}, \dots, x_{i7}$ as explanatory variables, the observations $y_i, i = 1, 2, \dots, n$, were generated from the Poisson regression model in (1). Thus, the y -variates are Poisson distributed with mean

$$\mu_i = \exp\left(\sum_{j=0}^7 \beta_j x_{ij}\right).$$

The Nordberg method is used to perform variable selection on each set of data generated. The forward selection algorithm developed in this article is also used for variable selection. The result from using AIC selection criterion is presented in this article. The result from using R_{adj}^2 selection criterion is the same as that of using AIC, and hence the result is not given.

Each simulation was repeated 1000 times by generating new y -variates keeping the x -data and the $\beta_1, \beta_2, \dots, \beta_7$ constant. Since the parameters $\beta_5 = \beta_6 = \beta_7 = 0$, we expect $x_5, x_6,$ and

x_7 not to enter into the selected model. Whenever any or all of these three variables enter a selected model, it is considered an error. The error rate from the 1000 simulations was recorded in Table 1 for both selection methods. In each simulation, the percentage of fitted values $\hat{\mu}_i$ less than or equal to 4 is recorded. These percentage values are averaged over the 1000 simulations and the results are presented in Table 1.

From Table 1, we notice some differences between the error rates from the forward selection method and the transformation method proposed by Nordberg. In general, the error rates from the forward selection method are smaller than the error rates from the Nordberg method. The error rates are much larger when the sample size is small, say $n = 100$ or $n = 250$. As the sample size increases to $n = 500$ or $n = 1000$, the two methods are closer in performance. However, the forward selection method seems to have a slight advantage over the Nordberg method. When the percentage of the fitted values $\hat{\mu}_i$ less than or equal to 4.0 is high, the error rates from the Nordberg method seem to be high, especially when the sample size n is small.

From the simulation study, the difference between the two selection methods is not only due to whether the percentage of fitted values $\hat{\mu}_i$ less than or equal to 4.0 is high, it also depends on the sample size n . For small sample size, the Nordberg method tends to select variables $x_5, x_6,$ and/or x_7 more often than the forward selection algorithm presented earlier. As the sample size increases to 1000, the Nordberg method tends to perform as well as the forward selection algorithm.

Table 1. Error Rates For Nordberg And Forward Selection Algorithms.

N	β_0	Nordberg Method	Forward selection	Percentage of $\hat{\mu}_i \leq 4.0$
100	-1.0	0.188	0.166	100.0
	-0.5	0.189	0.170	100.0
	1.5	0.164	0.140	42.3
	1.7	0.159	0.129	24.1
	2.0	0.145	0.129	6.9
	3.0	0.147	0.130	0.0
250	-1.0	0.191	0.168	100.0
	-0.5	0.171	0.155	100.0
	1.5	0.155	0.136	43.0
	1.7	0.155	0.142	24.0
	2.0	0.155	0.143	7.8
	3.0	0.152	0.142	0.4
500	-1.0	0.159	0.151	100.0
	-0.5	0.147	0.139	100.0
	1.5	0.133	0.136	41.6
	1.7	0.139	0.139	23.4
	2.0	0.149	0.147	6.9
	3.0	0.143	0.138	0.2
1000	-1.0	0.144	0.144	100.0
	-0.5	0.154	0.146	100.0
	1.5	0.162	0.160	38.9
	1.7	0.153	0.147	20.5
	2.0	0.159	0.156	5.7
	3.0	0.148	0.145	0.1

Numerical Example

We applied the forward selection algorithm and the transformation method suggested by Nordberg (1982) to several data sets. The forward selection algorithm was implemented using AIC and R^2_{adj} as selection criteria statistics. When the percentage of the $\hat{\mu}_i$ less than or equal to 4 satisfied the cases considered by Nordberg (1982), both methods yielded the same sub-model. However, when the data has a much larger percentage of $\hat{\mu}_i$ less than or equal to 4, we tend to obtain different results. We now present the results of a data set.

Wang and Famoye (1997) modeled fertility data using Poisson and generalized Poisson regression models. The data was from the Michigan Panel Study of Income Dynamics (PSID), a large national longitudinal data set. The particular portion of the data used in this paper was from 1989 and consisted of data from

2936 married women who were not head of households and with nonnegative total family income. The dependent variable was the number of children. Of the families, 1029 (35.05%) had no children under age 17. The response variable had a mean of 1.29 and a variance of 1.50. The predicted values under full Poisson regression model were small with 54.26% less than or equal to 1. Thus the data set was much more extreme than any of the cases considered by Nordberg (1982).

The Poisson regression model was fitted to the data using 12 covariates. The results are presented in Table 2. The forward selection algorithm was run on the data and the variables chosen are $x_9, x_1, x_4, x_5, x_2,$ and x_{10} . The variables chosen are exactly the same variables that are significant in the full model. The transformation method proposed by Nordberg (1982) was applied to the data. The variables selected were

Table 2. Poisson Regression Model.

Parameter	Full Model		Forward Selection Sub-model		
	Estimate \pm s.e.	t-value	Estimate \pm s.e.	t-value	Step added
Intercept	2.0686 \pm 0.1511	13.69*	2.1226 \pm .0744	28.53*	--
x_1	-0.2657 \pm 0.0356	-7.46*	-0.2674 \pm .0351	-7.61*	2
x_2	-0.0193 \pm 0.0041	-4.71*	0.0196 \pm .0041	-4.82*	5
x_3	-0.1226 \pm 0.0651	-1.88	--	--	--
x_4	-0.2811 \pm 0.0379	-7.42*	-0.2629 \pm .0368	-7.15*	3
x_5	0.3057 \pm 0.0575	5.32*	0.3002 \pm .0567	5.29*	4
x_6	-0.0050 \pm 0.0087	-0.57	--	--	--
x_7	0.0035 \pm 0.0071	0.49	--	--	--
x_8	-0.0143 \pm 0.0187	-0.76	--	--	--
x_9	-0.0211 \pm 0.0038	-5.55*	-0.0217 \pm .0038	-5.76*	1
x_{10}	-0.0147 \pm 0.0066	-2.23*	-0.0132 \pm .0059	-2.25*	6
x_{11}	0.0118 \pm 0.0078	1.51	--	--	--
X_{12}	-0.0545 \pm 0.0340	-1.60	--	--	--

*Significant at 5% level.

x_6 , x_7 , x_8 , x_{11} , x_9 , and x_{10} . These are not the same variables chosen by the forward selection procedure. Only two of the variables are chosen by both methods. The results from the transformation method are shown in Table 3. The parameters corresponding to x_6 , x_7 , x_8 , and x_{11} are not significant in the full Poisson regression model (see Table 2), causing concerns about the accuracy of the transformation method.

Goodness-of-fit statistics for the models are provided in Table 4. The goodness-of-fit statistics for the forward selection sub-model are close to those for the full model even though the number of independent variables is now six. This is not the case for the transformation sub-model. All these results are in support of the simulation study reported earlier.

Table 3. Nordberg’s Transformation Method.

Parameter	Transformation sub-model		
	Estimate ± s.e.	t-value	Step added
Intercept	1.8062±.1484	12.17*	--
x_6	-0.0140±.0086	-1.63	1
x_7	0.0075±.0070	-1.07	2
x_8	-0.0063±.0186	-0.34	3
x_9	-0.0376±.0017	-22.12*	5
x_{10}	-0.0027±.0008	-3.04*	6
x_{11}	0.0124±.0077	1.61	4

* Significant at 5% level.

Table 4. Goodness-of-fit For The Poisson Model.

Statistic	Full Model	Forward Selection	Nordberg
Deviance	3277.84	3286.14	3430.81
d.f.	2923.0	2929.0	2929.0
Log-likelihood	-2410.09	-2414.24	-2486.58
AIC	2423.09	2421.24	2493.58
R^2_{adj}	0.2014	0.1994	0.1639

Conclusion

The size of the predicted values affected the usefulness of the transformation method. In the data set, the predicted values are relatively small (54.3% less than or equal to 1). Since the approximation error E in (9) for the transformation method involves division by the square root of the predicted value, one should be concerned when many predicted values are small. Dividing by small values may cause this error term to become large and make the approximation inaccurate. Although many other data sets analyzed indicate that the transformation method can be useful when the

predicted values are large, it may run into problems when predicted values are small. Real world data may not necessarily have large predicted values. It would be ideal to have an algorithm that is not dependent on the size of the predicted values. The forward selection method presented performed well regardless of the size of the predicted values.

The forward selection algorithm may take much more computer time than the transformation method proposed by Nordberg (1982). In these days of better computer technology, more computer time should not be a reason for using a method that may not always produce an adequate result. From our simulation study, the forward selection algorithm performs as well if not better than the transformation method.

In this article, a forward selection algorithm was developed. Similar methods could be developed using backward or stepwise selection for the class of generalized linear models. In addition, other selection criteria statistics could be used. Count data occur very frequently in real world applications. The size of the predicted values cannot be controlled within a particular study. Thus a selection method that can deal with any size of predicted values is desirable.

References

Akaike, H. (1973). Information theory and an extension of the likelihood principle, in *Proceedings of the Second International Symposium of Information Theory*, Eds. Petrov, B.N., & Csaki, F. Budapest: Akademiai Kiado.

Albert, J. (1992), A Bayesian analysis of a Poisson random-effects model for Home Run hitters, *The American Statistician*, 46, 246-253.

Beale, E. M. L. (1970). Note on procedures for variable selection in multiple regression, *Technometrics*, 12, 909-914.

Cameron, A. C., & Trivedi, P. (1986). Econometric models based on count data: comparisons and applications of some estimators and tests, *Journal of Applied Econometrics*, 1, 29-53.

- Cameron, A. C., & Windmeijer, F. A. G. (1996). R-squared measures for count data regression models with applications to health-care utilization, *Journal of Business and Economic Statistics*, 14, 209-220.
- Christiansen, C. L., & Morris, C. N. (1997). Hierarchical Poisson regression modeling, *Journal of the American Statistical Association*, 92, 618-632.
- D'Unger, D. V., Land, K. C., McCall, P. L., & Nagin, D. S. (1998). How many latent classes of delinquent/criminal careers? Results from mixed Poisson regression analyses, *American Journal of Sociology*, 103, 1593-1630.
- Efron, B., & Thisted, R. (1976). Estimating the number of unseen species: how many words did Shakespeare know? *Biometrika*, 63, 435-447.
- Efroymson, M. A. (1960). Multiple regression analysis, *Mathematical Methods for Digital Computers*. Eds. A. Ralston & H.S. Wilf, New York: John Wiley & Sons, Inc.
- Frome, E. L., Kutner, M. H., & Beauchamp, J. J. (1973). Regression analysis of Poisson-distributed data, *Journal of the American Statistical Association*, 68, 935-940.
- Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: applications to Poisson models, *Econometrica*, 52, 701-720.
- Greenberg, R. A., Bayard, S., & Byar, D. (1974). Selecting concomitant variables using a likelihood ratio step-down procedure and a method of testing goodness-of-fit in an exponential survival model, *Biometrics*, 30, 601-608.
- Institute for Social Research, The University of Michigan (1992). *A Panel Study of Income Dynamics*. Ann Arbor, MI.
- Krall, J. M., Uthoff, V. A., & Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival, *Biometrics*, 31, 49-57.
- Lawless, J. F. (1987). Negative binomial and mixed Poisson regression, *The Canadian Journal of Statistics*, 15, 209-225.
- Merkle, L., & Zimmermann, K. F. (1992). The demographics of labor turnover: a comparison of ordinal probit and censored count data models, *Recherches Economiques de Louvain*, 58, 283-307.
- Nordberg, L. (1982). On variable selection in generalized linear and related regression models, *Communications in Statistics Theory and Method*, 11, 2427-2449.
- Signorini, D. F. (1991). Sample size for Poisson regression, *Biometrika*, 78, 446-450.
- Wang, W., & Famoye, F. (1997). Modeling household fertility decisions with generalized Poisson regression, *Journal of Population Economics*, 10, 273-283.