Economics Letters 107 (2010) 310–312

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

On the existence of the maximum likelihood estimates in Poisson regression

J.M.C. Santos Silva [a,*], Silvana Tenreyro [b,1]

[a] University of Essex and CEMAPRE. Wivenhoe Park, Colchester CO4 3SQ, United Kingdom
[b] London School of Economics, CREI, CEP, and CEPR. Department of Economics, s.600. St. Clement's Building. Houghton St., London WC2A 2AE, United Kingdom

ARTICLE INFO

Article history:
Received 11 June 2009
Received in revised form 24 February 2010
Accepted 25 February 2010
Available online 2 March 2010

JEL classification:
C13
C25

Keywords:
Collinearity
Complete separation
Perfect prediction

ABSTRACT

We note that the existence of the maximum likelihood estimates in Poisson regression depends on the data configuration, and propose a strategy to identify the existence of the problem and to single out the regressors causing it.

1. Introduction

The Poisson regression model is defined by

$$\Pr(y_i = j | x_i) = \frac{\exp(-\lambda)\lambda^j}{j!}, \qquad j = 0, 1, 2, \ldots$$

where λ is generally specified as $\lambda = \exp(x_i'\beta) = \exp(\beta_0 + \beta_1 x_{1i} + \ldots)$.[2] With this formulation, β, the vector of parameters of interest, can be estimated my maximizing the log-likelihood function given by

$$\ln L(\beta) = \sum_{i=1}^{n} \left[-\exp(x_i'\beta) + (x_i'\beta)y_i - \ln(y_i!) \right]. \quad (1)$$

Poisson regression is not only the most widely used model for count data (see Winkelmann, 2008; Cameron and Trivedi, 1998), but it is also becoming increasingly popular to estimate multiplicative models for other kinds of data (see, among others, Manning and Mullahy, 2001, and Santos Silva and Tenreyro, 2006).

The reasons that make this estimator popular can be clearly understood by inspecting the corresponding score vector and Hessian matrix, given respectively by

$$s(\beta) = \sum_{i=1}^{n} \left[y_i - \exp(x_i'\beta) \right] x_i, \quad (2)$$

and

$$H(\beta) = -\sum_{i=1}^{n} \exp(x_i'\beta) x_i x_i'.$$

The form of the score vector makes clear that β will be consistently estimated as long as $E(y_i|x_i) = \exp(x_i'\beta)$, i.e., the only condition required for consistency is the correct specification of the conditional mean. This is the well known pseudo-maximum likelihood result of Gourieroux et al. (1984).

Besides this robustness property, the estimator also has the advantage of being very well behaved. Indeed, it is easy to see that the Hessian is negative definite for all x and β, which facilitates the estimation and ensures the uniqueness of the maximum, if it exists. Consequently, estimation of β is relatively simple and generally the estimation algorithm converges in a handful of iterations, even for relatively large problems.

In spite of this general result, for certain data configurations, some of the parameters in β are not identified by the (pseudo) maximum likelihood estimator described above. That is, for certain data

* Corresponding author. Fax: +44 1206 872724.
E-mail addresses: jmcss@essex.ac.uk (J.M.C. Santos Silva), s.tenreyro@lse.ac.uk (S. Tenreyro).
1 Fax: +44 20 78311840.

2 See Winkelmann (2008) and Cameron and Trivedi (1998) for further details and background on the Poisson regression model and its properties.

0165-1765/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2010.02.020

configurations, the maximum likelihood estimates of β, say $\hat{\beta}$, do not exist. Because this type of identification failure has not been widely recognized as a problem in count data models, standard software does not check for its presence and therefore the practitioner may be surprised to find that estimation of the Poisson regression is unusually difficult, even in some apparently simple problems. This letter provides details on when this problem arises, on how it can be detected, and on how it can be overcome.

2. The problem

To better see the nature of the problem, it is useful to start by considering the case where a regressor, say x_{i2}, is zero when y_i is positive, otherwise being non-negative with at least one positive observation. The leading example of a regressor with these characteristics is a dummy variable that is equal to zero for all observations with positive y_i, having some positive values for $y_i = 0$. From Eq. (2), the first order condition for a maximum of Eq. (1) corresponding to the parameter associated with x_{i2} can be written as

$$s(\beta_2) = -\sum_{x_{2i}>0} \exp\left(x_i'\hat{\beta}\right)x_{2i} = 0,$$

which can never be satisfied. Therefore, when regressors such as x_2 are present, the (pseudo) maximum likelihood estimate of β does not exist.

More generally, this problem can occur whenever two regressors are perfectly collinear for the sub-sample with positive observations of y_i.[3] To see this, write Eq. (2) as

$$s(\beta) = \sum_{y_i>0}[y_i - \exp(x_i'\beta)]x_i - \sum_{y_i=0}\exp(x_i'\beta)x_i,$$

and notice that the first order conditions for a maximum corresponding to β_0, β_1 and β_2 imply

$$\sum_{y_i>0}\left[y_i - \exp\left(x_i'\hat{\beta}\right)\right] = \sum_{y_i=0}\exp\left(x_i'\hat{\beta}\right), \tag{3a}$$

$$\sum_{y_i>0}\left[y_i - \exp\left(x_i'\hat{\beta}\right)\right]x_{1i} = \sum_{y_i=0}\exp\left(x_i'\hat{\beta}\right)x_{1i}, \tag{3b}$$

$$\sum_{y_i>0}\left[y_i - \exp\left(x_i'\hat{\beta}\right)\right]x_{2i} = \sum_{y_i=0}\exp\left(x_i'\hat{\beta}\right)x_{2i}. \tag{3c}$$

Suppose now that x_1 and x_2 are perfectly collinear for the sub-sample with positive observations of y_i. In particular, let $x_{2i} = \alpha_0 + \alpha_1 x_{1i}$ for $y_i > 0$. Then, writing x_{2i} as a function of x_{1i} on the left hand side of Eq. (3c) and using equalities Eqs. (3a) and (3b), it is possible to obtain

$$\alpha_0 \sum_{y_i=0}\exp\left(x_i'\hat{\beta}\right) + \alpha_1 \sum_{y_i=0}\exp\left(x_i'\hat{\beta}\right)x_{1i} = \sum_{y_i=0}\exp\left(x_i'\hat{\beta}\right)x_{2i}. \tag{4}$$

Whether or not Eq. (4) has a solution depends on the values of α_0 and α_1, and on the ranges of x_1 and x_2 for the observations with $y_i = 0$. For instance, in the illustrative example presented before where x_{i2} is zero when y_i is positive, we have that $\alpha_0 = \alpha_1 = 0$, and for Eq. (4) to have a solution it is necessary, but not sufficient, that x_2 has positive and negative values for $y_i = 0$. Heuristically, Eq. (4) will have a solution when there is a reasonable overlap between the ranges of x_{2i} for $y_i = 0$ and $y_i > 0$. However, it is not possible to provide an easy way to determine the existence of a $\hat{\beta}$ that solves Eq. (4).[4] Therefore, the existence of this sort of identification problem has to be investigated on a case-by-case basis.

Of course, Newton-type algorithms used to maximize the likelihood function may achieve convergence even when Eqs. (3a), (3b), (3c) have no solution, leading to spurious maximum likelihood estimates, say b. It is easy to see that for b to provide an approximate solution for Eq. (4) it has to be such that $\exp(x_i'b)$ is close to zero for the observations with $y_i = 0$. Therefore, these spurious solutions can be easily identified because they are characterized by a "perfect" fit for the observations with $y_i = 0$.

This situation is analogous to what happens in binary choice models when there is complete separation or quasi-complete separation, as described by Albert and Anderson (1984) and Santner and Duffy (1986). Moreover, it is clear that it can also occur in any other regression model where the conditional mean is specified in such a way that its image does not include all the points in the support of the dependent variable. Therefore, this problem can occur not only in the Poisson regression model but whenever y is non-negative and the conditional mean is specified as $E(y_i|x_i) = \exp(x_i'\beta)$.

3. Discussion

The results of the previous section make clear that the non-existence of the (pseudo) maximum likelihood estimates of the Poisson regression models is more likely when the data has a large number of zeros.[5] For example, this problem is likely to arise when modelling the number of crimes committed, the number of instances of substance abuse, or the volume of trade between pairs of countries. Therefore, in these cases, even if the estimation algorithm converges and delivers a set of estimates, it is recommended that the researcher checks whether or not the results obtained actually correspond to a maximum of the (pseudo) log-likelihood function. This can be easily done by checking for the overfitting of the observations with $y_i = 0$, for example by computing descriptive statistics for the fitted values of y for the relevant sub-sample.

When the researcher identifies a situation where the (pseudo) maximum likelihood estimates do not exist, either because the algorithm does not converge or because convergence is achieved by overfitting the zeros, it is useful to have a simple strategy to single out the regressors causing the problem. Because these regressors are characterized by their perfect collinearity with the others for the sub-sample with $y_i > 0$, they can be identified using the following procedure, which explores the fact that statistical software generally drop perfectly collinear explanatory variables in ordinary least squares regression:[6]

Step 1. For the observations with $y_i > 0$, estimate the ordinary least squares regression of $ln(y_i)$ on x_i;[7]
Step 2. Construct a subset of explanatory variables, say $\tilde{x}_i$, comprising only the regressors whose coefficients were estimated in Step 1;
Step 3. Using the full sample, run the Poisson regression of y_i on $\tilde{x}_i$.

This procedure eliminates all potentially problematic regressors, even those that actually do not lead to the non-existence of the maximum likelihood estimates. Therefore, the researcher should then investigate one-by-one all the variables that were excluded, to see if any of them can be included in the model.

[3] Notice that the problem identified here is very different from the one resulting from perfect collinearity between regressors. Perfect collinearity leads to the existence on an infinite number of solutions to the likelihood equations, whereas here we are concerned with the situation where the likelihood equations have no solution.

[4] A referee has pointed out to us that Haberman (1973, p. 624) provides a necessary and sufficient condition for the existence of the maximum likelihood estimates of β. However, as the referee noted, the condition is not easily verifiable.

[5] Indeed, if y_i is strictly positive, the maximum likelihood estimate of β always exists (see Corollary 3.1 in Haberman, 1973, p. 624).

[6] Stata code to implement this algorithm is available in the authors' web page. We are grateful to Markus Baldauf for the help with the development of this code.

[7] We suggest using $ln(y_i)$ as the dependent variable in this regression so that the estimates obtained here can be used as starting values for the Poisson regression in the third step.

Acknowledgements

We are grateful to Ines Buono, Virginia Di Nino, Dave Donaldson, Doireann Fitzgerald, Lissandra Flach and Randi Hjalmarsson for noting the problem and for providing examples of data sets where it occurs. We also thank J.M. Andrade e Silva and an anonymous referee for the valuable comments. The usual disclaimer applies. Santos Silva also gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010). Tenreyro gratefully acknowledges the support of Banco de España through the Banco de España Associate Professorship.

References

Albert, A., Anderson, J.A., 1984. On the existence of maximum likelihood estimates in logistic models. Biometrika 71, 1–10.

Cameron, A.C., Trivedi, P.K., 1998. Regression Analysis of Count Data. Cambridge University Press, Cambridge, MA.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: applications to Poisson models. Econometrica 52, 701–720.

Haberman, S.J., 1973. Log-linear models for frequency data: sufficient statistics and likelihood equations. Annals of Statistics 1, 617–632.

Manning, W.G., Mullahy, J., 2001. Estimating log models: to transform or not to transform? Journal of Health Economics 20, 461–494.

Santner, T.J., Duffy, E.D., 1986. A note on A. Albert and J.A. Anderson's conditions for the existence of maximum likelihood estimates in logistic regression models. Biometrika 73, 755–758.

Santos Silva, J.M.C., Tenreyro, S., 2006. The log of gravity. The Review of Economics and Statistics 88, 641–658.

Winkelmann, R., 2008. Econometric Analysis of Count Data, 5th ed. Springer-Verlag, Berlin.