CENTRE for ECONOMIC PERFORMANCE

CEP Discussion Paper No 932

May 2009

On the Existence of the Maximum Likelihood Estimates for Poisson Regression

J. M. C. Santos-Silva and Silvana Tenreyro

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Abstract
We note that the existence of the maximum likelihood estimates for Poisson regression depends on the data configuration. Because standard software does not check for this problem, the practitioner may be surprised to find that in some applications estimation of the Poisson regression is unusually difficult or even impossible. More seriously, the estimation algorithm may lead to spurious maximum likelihood estimates. We identify the signs of the non-existence of the maximum likelihood estimates and propose a simple empirical strategy to single out the regressors causing this type of identification failure.

Keywords: Poisson estimation, gravity equation
JEL Classifications: C13; C50; F10

This paper was produced as part of the Centre's Globalisation Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements
We are grateful to Ines Buono, Virginia Di Nino, Dave Donaldson, Doireann Fitzgerald, Lissandra Flach and Randi Hjalmarsson for posing the problem and for providing examples of data sets where it occurs. Special thanks are due to Markus Baldauf for help with the code used to bypass the problem. Santos Silva also gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

João M. C. Santos-Silva is Professor of Economics at the University of Essex. Silvana Tenreyro is an Associate at the Centre for Economic Performance and Reader in the Economics Department, London School of Economics.

Published by
Centre for Economic Performance
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

ISBN 978-0-85328-385-0

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

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

Poisson regression is not only the most widely used model for count data (see Winkelmann, 2008, and Cameron and Trivedi, 1997), 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

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

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 $\mathrm{E}(y_i|x_i) = \exp(x_i'\beta)$, 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, Monfort and Trognon (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.

[1] See Winkelmann (2008) and Cameron and Trivedi (1997) for further details and background on the Poisson regression model and its properties.

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 configurations, the maximum likelihood estimates do not exist. Because this type of identification failure has not been 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. The next section provides details on when this problem arises and on how it can be detected.

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 equation (2), the first order condition for a maximum of (1) corresponding to the parameter associated with x_{i2} can be written as

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

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.[2] To see this, write (2) as

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

[2]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.

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

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

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

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

where $\hat{\beta}$ denotes the maximum likelihood estimates of β.

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 (3c) and using equalities (3a) and (3b), it is possible to obtain

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

Whether or not (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, $\alpha = \beta = 0$ and for (4) to have a solution it is necessary, but not sufficient, that x_2 has positive and negative values for $y_i = 0$. Heuristically, (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 a sharp criterion determining the existence of a $\hat{\beta}$ that solves (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 (3) has no solution, leading to spurious maximum likelihood estimates, say b. It is easy to see that for b to provide an approximate solution for (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 $\mathrm{E}(y_i|x_i) = \exp(x_i'\beta)$.

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. 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:

Step 1: For the observations with $y_i > 0$, estimate the ordinary least squares regression of $\ln(y_i)$ on x_i;

Step 2: Construct a subset of explanatory variable, say $\tilde{x}_i$, comprising only the regressors whose coefficients were estimated in Step 1;

Step 3: Using the observations with $y_i \geq 0$, i.e., the full sample, run the Poisson regression of y_i on $\tilde{x}_i$.[3]

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.

REFERENCES

Albert, A. and Anderson, J.A. (1984). "On the existence of maximum likelihood estimates in logistic models," *Biometrika*, 71, 1-10.

Cameron, A.C. and Trivedi, P.K. (1998). *Regression analysis of count data*, Cambridge, MA: Cambridge University Press.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). "Pseudo maximum likelihood methods: Applications to Poisson models," *Econometrica*, 52, 701-720.

Manning, W.G. and Mullahy, J. (2001). "Estimating Log Models: To Transform or Not to Transform?," *Journal of Health Economics* 20, 461-494.

Santner T.J. and 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. and Tenreyro, S. (2006), "The log of gravity," *The Review of Economics and Statistics*, 88, 641-658.

Winkelmann, R. (2008). *Econometric analysis of count data*, 5$^{\text{th}}$ ed., Berlin: Springer-Verlag.

[3] The parameters estimated by the OLS regression in Step 1 can be used as starting values for this Poisson regression.

CENTRE FOR ECONOMIC PERFORMANCE
Recent Discussion Papers

931	Richard Freeman John Van Reenen	What If Congress Doubled R&D Spending on the Physical Sciences?
930	Hector Calvo-Pardo Caroline Freund Emanuel Ornelas	The ASEAN Free Trade Agreement: Impact on Trade Flows and External Trade Barriers
929	Dan Anderberg Arnaud Chevalier Jonathan Wadsworth	Anatomy of a Health Scare: Education, Income and the MMR Controversy in the UK
928	Christos Genakos Mario Pagliero	Risk Taking and Performance in Multistage Tournaments: Evidence from Weightlifting Competitions
927	Nick Bloom Luis Garicano Raffaella Sadun John Van Reenen	The Distinct Effects of Information Technology and Communication Technology on Firm Organization
926	Reyn van Ewijk	Long-term health effects on the next generation of Ramadan fasting during pregnancy
925	Stephen J. Redding	The Empirics of New Economic Geography
924	Rafael Gomez Alex Bryson Tobias Kretschmer Paul Willman	Employee Voice and Private Sector Workplace Outcomes in Britain, 1980-2004
923	Bianca De Paoli	Monetary Policy Under Alterative Asset Market Structures: the Case of a Small Open Economy
922	L. Rachel Ngai Silvana Tenreyro	Hot and Cold Seasons in the Housing Market
921	Kosuke Aoki Gianluca Benigno Nobuhiro Kiyotaki	Capital Flows and Asset Prices
920	Alex Bryson John Forth Patrice Laroche	Unions and Workplace Performance in Britain and France
919	David Marsden Simone Moriconi	‘The Value of Rude Health’: Employees’ Well Being, Absence and Workplace Performance
918	Richard Layard Guy Mayraz Stephen Nickell	Does Relative Income Matter? Are the Critics Right?

917	Ralf Martin Laure B. de Preux Ulrich J. Wagner	The Impacts of the Climate Change Levy on Business: Evidence from Microdata
916	Paul-Antoine Chevalier Rémy Lecat Nicholas Oulton	Convergence of Firm-Level Productivity, Globalisation, Information Technology and Competition: Evidence from France
915	Ghazala Azmat Nagore Iriberri	The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment using High School Students
914	L Rachel Ngai Robert M. Samaniego	Accounting for Research and Productivity Growth Across Industries
913	Francesco Caselli Tom Cunningham	Leader Behavior and the Natural Resource Curse
912	Marco Manacorda Edward Miguel Andrea Vigorito	Government Transfers and Political Support
911	Philippe Aghion John Van Reenen Luigi Zingales	Innovation and Institutional Ownership
910	Fabian Waldinger	Peer Effects in Science – Evidence from the Dismissal of Scientists in Nazi Germany
909	Tomer Blumkin Yossi Hadar Eran Yashiv	The Macroeconomic Role of Unemployment Compensation
908	Natalie Chen Dennis Novy	International Trade Integration: A Disaggregated Approach
907	Dongshu Ou	To Leave or Not to Leave? A Regression Discontinuity Analysis of the Impact of Failing the High School Exit Exam
906	Andrew B. Bernard J. Bradford Jensen Stephen J. Redding Peter K. Schott	The Margins of US Trade
905	Gianluca Benigno Bianca De Paoli	On the International Dimension of Fiscal Policy
904	Stephen J. Redding	Economic Geography: A Review of the Theoretical and Empirical Literature

The Centre for Economic Performance Publications Unit
Tel 020 7955 7284 Fax 020 7955 7595 Email info@cep.lse.ac.uk
Web site http://cep.lse.ac.uk