Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator*

J.M.C. Santos Silva† Silvana Tenreyro‡

January 27, 2009

Abstract

We extend the simulation results given in Santos Silva and Tenreyro (2006, "The log of gravity," *The Review of Economics and Statistics*, 88, 641-658) by considering data generated as a finite mixture of gamma variates. Data generated in this way can naturally have a large proportion of zeros and is fully compatible with constant elasticity models such as the gravity equation. Our results confirm that the Poisson pseudo maximum likelihood estimator is generally well behaved.

1. INTRODUCTION

Santos Silva and Tenreyro (2006) suggested that the Poisson pseudo-maximum likelihood (PPML) estimator introduced by Gourieroux Monfort and Trognon (1984) has all the characteristics needed to make it a promising workhorse for the estimation of gravity equations and, more generally, constant elasticity models. Santos Silva and Tenreyro

*Santos Silva gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (FEDER/POCI 2010).

†University of Essex and CEMAPRE. Wivenhoe Park, Colchester CO1 1ED, United Kingdom. Fax: +44 (0)1206 872769. E-mail: jmcss@essex.ac.uk.

‡London School of Economics, CEP, and CEPR. Department of Economics, s.600. St. Clement's Building. Houghton St., London WC2A 2AE, United Kingdom. Fax: +44 (0)20 78311840. E-mail: s.tenreyro@lse.ac.uk.

(2006) provided simulation evidence that the PPML is well behaved in a wide range of situations and is resilient to the presence of a specific type of measurement error of the dependent variable.

However, in the simulations performed by Santos Silva and Tenreyro (2006), the dependent variable was necessarily positive, except in the case where the dependent variable was contaminated by measurement error. This lack of zeros of the dependent variable in the main set of experiments presented in Santos Silva and Tenreyro (2006) has raised some questions about the performance of the estimator in situations where the dependent variable is frequently equal to zero. Although there is no theoretical justification to expect any significant difference in the performance of the PPML estimator when the dependent variable is non-negative rather that positive, it is interesting to investigate the issue with an appropriate Monte Carlo study.

This issue has been addressed by Martínez-Zarzoso, Nowak-Lehmann and Vollmer (2007) and by Martin and Pham (2008). However, the simulations performed by these authors are flawed in that the data is not generated by a constant elasticity model. Therefore, these simulations provide no information at all on the performance of the PPML estimator of constant elasticity models. In this paper we present simulation evidence on the performance of the PPML estimator when the data is generated by a constant elasticity model and the dependent variable has a large proportion of zeros, as is typical of the trade data used in the estimation of gravity equations.

2. SIMULATION DESIGN

In these simulations, the non-negative dependent variable y_i is generated so that $\Pr(y_i = 0)$ is substantial and

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

where x_i is a vector of regressors.[1] In particular, y_i is generated as a finite mixture model of the form

$$y_i = \sum_{j=1}^{m_i} z_{ij},$$

where $m_i \geq 0$ is the number of components of the mixture, and z_{ij} is a continuous random variable with support in $\mathbb{R}^+$ and distributed independently of m_i.

Besides being computationally convenient, this data generation scheme has a natural interpretation in the context of trade data. Indeed, m_i can be understood as the number of exporters and z_{ij} the quantity exported by firm j.

It is easy to see that

$$\mathrm{E}(y_i|x_i) = \mathrm{E}(m_i|x_i)\,\mathrm{E}(z_{ij}|x_i).$$

Therefore, if $\mathrm{E}(m_i|x_i) = \exp(x_i'\gamma)$ and $\mathrm{E}(z_{ij}|x_i) = \exp(x_i'\delta)$, we have that $\mathrm{E}(y_i|x_i) = \exp(x_i'\beta)$ with $\beta = \gamma + \delta$.

Draws of z_{ij} can be obtained from any continuous distribution with support in $\mathbb{R}^+$, like the gamma, lognormal or exponential distributions. However, due to its additivity, the gamma distribution is particularly suited for simulations and it is used here. The number of components of the mixture can be generated by any standard distribution for counts and in these experiments m_i will be generated as a negative-binomial random variable, with conditional mean $\exp(x_i'\gamma)$ and a variance to be specified below.

In order to simplify the simulation design, we set $\delta = 0$ and z_{ij} will be generated by a gamma distribution with mean 1 and variance 2. Specifically, z_{ij} is generated as a $\chi^2_{(1)}$ random variable, implying that conditionally on m_i, y_i follows a $\chi^2_{(m_i)}$ distribution. Integrating out m_i, we obtain $\mathrm{E}(y_i|x_i) = \mathrm{E}(m_i|x_i)$ and $\mathrm{Var}(y_i|x_i) = \mathrm{E}(m_i|x_i) + 2\mathrm{Var}(m_i|x_i)$.

As in Santos Silva and Tenreyro (2006), the conditional mean $\mathrm{E}(y_i|x_i)$ was specified as:

$$\mathrm{E}(y_i|x_i) = \mathrm{E}(m_i|x_i) = \mu(x_i\beta) = \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}), \qquad (1)$$

[1] The vector x_i can be interpreted as containing the logs of the elements of a vector of regressors X_i, assumed to be positive. Therefore, β can be interpreted as the elasticity of the conditional expectation of y_i with respect to X_i.

where, x_{1i} is drawn from a standard normal and x_2 is a binary dummy variable that equals 1 with a probability of 0.4. The two covariates are independent and a new set of observations of all variables is generated in each replication using $\beta_0 = 0$, $\beta_1 = \beta_2 = 1$.

To complete the design of the experiments it is necessary to define the conditional variance of m_i. We considered the following quadratic specification:

$$\text{Var}\left(m_i|x_i\right) = a\text{E}\left(m_i|x_i\right) + b\text{E}\left(m_i|x_i\right)^2,$$

which implies $\text{Var}\left(y_i|x_i\right) = \left(1 + 2a\right)\text{E}\left(m_i|x_i\right) + 2b\text{E}\left(m_i|x_i\right)^2$. Therefore, by varying the values of a and b, it is possible to generate a rich set of patterns of heteroskedasticity. The combinations of a and b used in the experiments are presented in Table 1, which also displays the approximate probability of observing $y_i = 0$ in each case..

Table 1: Values of $\Pr\left(y_i = 0\right)$ for different combinations of the parameters

Case number	1	2	3	4
a	10	50	1	1
b	0	0	5	15
$\Pr\left(y_i = 0\right)$	0.62	0.83	0.65	0.81

In cases 1 and 2, m_i has a NegBin1 distribution, with conditional variance proportional to the conditional mean. Therefore, in these cases the PPML estimator is optimal in the sense that its implicit assumption about the conditional variance is valid. For cases 3 and 4, the conditional variance is a quadratic function of the conditional mean and therefore m_i follows a NegBin2 distribution (see Cameron and Trivedi, 1997, or Wikelmann, 2008, for details on the NegBin1 and NegBin2 distributions). For cases 2 and 4, none of the estimators considered in these experiments will be optimal in the sense used above. However, as the importance of the quadratic term in the variance increases, the gamma pseudo-maximum likelihood estimator (GPML) will become approximately optimal.

In these experiments we analysed the performance of two consistent pseudo-maximum likelihood estimators of the multiplicative model: GPML and the PPML. The non-linear least squares considered by Santos Silva and Tenreyro (2006) was not included in these simulations because it revealed a dismal performance in preliminary trials. We also considered different estimators of the log-linearized model, namely, the truncated-at-zero OLS estimator, denoted OLS $(y>0)$; the OLS estimator using as dependent variable $\ln(y_i+1)$, denoted OLS $(y+1)$; and the threshold Tobit of Eaton and Tamura (1994), denoted ET-Tobit.[2]

In view of the claims of Martínez-Zarzoso, Nowak-Lehmann and Vollmer (2007), we also tried a FGLS estimator version of OLS $(y>0)$. In particular, we implemented the FGLS as described in Wooldridge (2009, p. 283). However, the results obtained with this estimator did not dominate those obtained with the simpler OLS $(y>0)$ and therefore will not be presented.

3. SIMULATION RESULTS

The results presented in this section where obtained with $10,000$ replicas of the simulation procedure described above, for samples of size $1,000$ and $10,000$. The results of these experiments are summarized in Table 2, which displays the biases and standard errors of the different estimators of β. Only results for β_1 and β_2 are presented, as these are generally the parameters of interest.

The results in Table 2 fully confirm the findings of Santos Silva and Tenreyro (2006). In particular, the PPML estimator is well behaved in all the cases considered, even when it is far from being optimal. The maximum bias of the PPML estimator over all the cases considered is smaller than 3.5%, for Case 4 and $N = 1,000$. The performance of the GPML is also generally very good, but its has reasonably large biases for Cases 1 and 2 when the smaller sample is considered. Indeed, for Case 2 and $N = 1,000$, the bias of the GPML is almost 17%. For $N = 10,000$ both the PPML and GPML have much lower biases, but the bias of the GPML is still above 2.5% for case 2.

[2] We also studied the performance of other variants of the Tobit model, finding very poor results.

Table 2: Simulation results when y_i is generated as a finite mixture of gamma variates

	$N = 1,000$				$N = 10,000$			
	β_1		β_2		β_1		β_2	
Estimator:	Bias	S.E.	Bias	S.E.	Bias	S.E.	Bias	S.E.
Case 1: $\text{Var}(y_i\|x_i) = 21\text{E}(m_i\|x_i)$								
PPML	−0.00066	0.066	0.00389	0.139	−0.00014	0.021	0.00062	0.043
GPML	0.04561	0.156	0.02440	0.224	0.00547	0.052	0.00330	0.071
ET-Tobit	−0.26013	0.085	−0.25741	0.109	−0.25971	0.027	−0.25813	0.034
OLS ($y > 0$)	−0.41440	0.105	−0.42796	0.199	−0.41453	0.033	−0.42952	0.062
OLS ($y + 1$)	−0.53477	0.029	−0.51048	0.057	−0.53468	0.009	−0.51135	0.018
Case 2: $\text{Var}(y_i\|x_i) = 101\text{E}(m_i\|x_i)$								
PPML	0.00038	0.139	0.00762	0.291	−0.00011	0.044	0.00228	0.091
GPML	0.16789	0.329	0.08616	0.483	0.02517	0.103	0.01338	0.147
ET-Tobit	0.11603	0.177	0.11511	0.234	0.11741	0.056	0.11702	0.074
OLS ($y > 0$)	−0.69422	0.178	−0.71706	0.337	−0.69202	0.055	−0.71717	0.105
OLS ($y + 1$)	−0.70096	0.033	−0.68840	0.060	−0.70065	0.011	−0.68832	0.019
Case 3: $\text{Var}(y_i\|x_i) = 3\text{E}(m_i\|x_i) + 10\text{E}(m_i\|x_i)^2$								
PPML	−0.01552	0.156	−0.00516	0.237	−0.00222	0.057	−0.00076	0.078
GPML	0.01453	0.110	0.00575	0.187	0.00187	0.035	0.00085	0.058
ET-Tobit	−0.36969	0.086	−0.37120	0.142	−0.36781	0.027	−0.36930	0.045
OLS ($y > 0$)	−0.35931	0.118	−0.35291	0.220	−0.35766	0.037	−0.35478	0.070
OLS ($y + 1$)	−0.71959	0.033	−0.70878	0.062	−0.71952	0.010	−0.70907	0.019
Case 4: $\text{Var}(y_i\|x_i) = 3\text{E}(m_i\|x_i) + 30\text{E}(m_i\|x_i)^2$								
PPML	−0.03480	0.242	−0.00594	0.390	−0.00546	0.095	−0.00272	0.129
GPML	0.01557	0.156	0.00650	0.284	0.00174	0.047	0.00034	0.087
ET-Tobit	−0.45051	0.124	−0.45489	0.224	−0.44949	0.039	−0.45262	0.070
OLS ($y > 0$)	−0.41138	0.167	−0.40497	0.318	−0.41339	0.052	−0.41382	0.100
OLS ($y + 1$)	−0.84074	0.031	−0.83526	0.060	−0.84077	0.010	−0.83597	0.018

Therefore, although both the PPML and the GPML are both consistent and generally well behaved, the PPML appears to be more robust to departures from the implicit heteroskedasticity assumptions.

As for the results in of the estimators based on the log-linear model, the results in Table 2 also fully confirm the findings of Santos Silva and Tenreyro (2006). Indeed, the ET-Tobit, the OLS $(y > 0)$ and the OLS $(y + 1)$ have very large biases that do not vanish as the sample size increases, confirming the inconsistency of these estimators.

4. CONCLUSIONS

The results presented in this study confirm that the Poisson pseudo maximum likelihood estimator is generally well behaved, even when the conditional variance is far from being proportional to the conditional mean. Moreover, as expected, the fact that the dependent variable has a large proportion of zeros does not affect the performance of the estimator. On the contrary, the presence of the zeros is an additional motive to use the Poisson pseudo maximum likelihood because in this case all estimators based on the log-linearization of the gravity equation have to use unreasonable solutions to deal with these observations.

Hence, like before, we conclude that the Poisson pseudo maximum likelihood estimator is a promising workhorse for the estimation of constant elasticity models such as the gravity equations.

REFERENCES

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.

Martin, W. and Pham, C.S. (2008). *Estimating the Gravity Equation when Zero Trade Flows are Frequent*. Available at: `http://mpra.ub.uni-muenchen.de/9453/`.

Martínez-Zarzoso, I., Nowak-Lehmann D., F. and Vollmer, S. (2007). "The Log of Gravity Revisited," CeGE Discussion Papers 64, University of Goettingen, available at: `http://www.etsg.org/ETSG2007/papers/zarzoso.pdf`.

Santos Silva, J.M.C. and Tenreyro, Silvana (2006), "The log of gravity," *The Review of Economics and Statistics*, 88, 641-658.

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

Wooldridge, J.M. (2009). *Introductory Econometrics, A Modern Approach*, 4^{th} ed., Cincinnati (OH): South-Western.