Scale-dependence of the Negative Binomial Pseudo-Maximum Likelihood Estimator

Clément Bosquet, Hervé Boulhol

▶ To cite this version:

Clément Bosquet, Hervé Boulhol. Scale-dependence of the Negative Binomial Pseudo-Maximum Likelihood Estimator. 2010. halshs-00544129

Scale-dependence of the Negative Binomial
Pseudo-Maximum Likelihood Estimator

Clément BOSQUET, Hervé BOULHOL

Scale-dependence of the Negative Binomial Pseudo-Maximum Likelihood Estimator[*]

Clément Bosquet[†]   Hervé Boulhol[‡]

November 9, 2010

Abstract

Following Santos Silva and Tenreyro (2006), various studies have used the Poisson Pseudo-Maximum Likelihood to estimate gravity specifications of trade flows and non-count data models more generally. Some papers also report results based on the Negative Binomial estimator, which is more general and encompasses the Poisson assumption as a special case. This note shows that the Negative Binomial estimator is inappropriate when applied to a continuous dependent variable which unit choice is arbitrary, because estimates artificially depend on that choice.

JEL Codes: C13, C21, F10

Keywords: pseudo-maximum likelihood methods, negative binomial estimator, Poisson regression, gamma PML

1

Résumé

Depuis l'article de Santos Silva et Tenreyro (2006), plusieurs études ont utilisé le pseudo-maximum de vraisemblance (PMV) avec loi de Poisson pour estimer les équations de gravité des flux de commerce et, plus généralement, des modèles avec données continues (c'est-à-dire non discrètes). Certains papiers rapportent aussi des résultats obtenus à l'aide du PMV avec loi binomiale négative, estimateur plus général dont la loi de Poisson est un cas particulier. Cette note montre que l'estimateur du PMV avec loi binomiale négative est inapproprié lorsque la variable à expliquer est mesurée avec une unité dont le choix est arbitraire, parce que les estimations dépendent artificiellement de ce choix.

Mots clés : méthodes du pseudo-maximum de vraisemblance, estimateur avec loi binomiale négative, régression de Poisson, pseudo-maximum de vraisemblance avec loi gamma

2

1   Introduction

Pseudo-Maximum Likelihood (PML) methods were introduced and then derived for Poisson models by Gourieroux, Monfort and Trognon (1984a,b). Following these seminal works, the Poisson PML (PPML) estimator, which assumes proportionality between the conditional variance and the conditional expectancy of the dependent variable, has often been used for count data models. However, beyond count data, Gourieroux *et al.* (1984b) note that "the pseudo-maximum likelihood method with Poisson family may be applied even if the dependent variable is any real number".

Santos Silva and Tenreyro (2006) highlight the advantages of this estimator for gravity equations of bilateral trade flows specified in levels, relative to the common practice of estimating these equations in log-levels by Ordinary Least Squares. Indeed, these authors show that the log-linear specification leads to biased estimates following Jensen's inequality, due to heteroskedasticity in trade levels.[1] Moreover, they provide some evidence that the PPML estimator is more efficient than the nonlinear least squares estimator of the trade specification in level.

As a result, a number of empirical studies of trade flows apply the PPML estimator. As an extension, some researchers consider other PML estimators based on non-Poisson distributions such as gamma according to which the variance is proportional to the square of the conditional mean. The Negative Binomial (NB) PML estimator has also been increasingly used recently in trade as well as mergers and acquisitions studies, including Head, Mayer and Ries (2009), Burger, van Oort and Linders (2009), Briant, Combes and Lafourcade (2009), Westerlund and Wilhelmsson (2009) and Garita and van Marrewijk (2008). The NB distribution assumes that the conditional variance is a linear combination, to be estimated, of the conditional mean and of its square. The NB PML estimator is appealing because it encompasses both PPML and gamma PML as special cases.

This note shows that the NB PML estimator is inappropriate when applied to continuous dependent variables, such as trade or M&A flows, for which the choice of the unit measure is arbitrary. For example, in the case of trade equations, the NB PML estimated parameters

---

[1]Because $E(Log\ x) \neq Log\ E(x)$ the expected value of the logarithm of trade flows depends on higher moments, including the variance. Since the variance of the residuals is likely to depend on explanatory variables, estimators using the log specification are biased.

depend artificially on whether trade flows are measured in thousands of dollars, in billions of dollars or in millions of euros. More precisely, when flows are measured in small units (*e.g.* thousands of dollars), the NB PML converges towards the gamma PML estimator. In contrast, when flows are measured in large units (*e.g.* trillions of dollars), the NB PML converges towards the Poisson PML estimator. This scale dependence has been unnoticed so far.

As an interesting case, Garita and van Marrewijk (2008) use the NB PML estimator with either the value or the number of mergers and acquisitions as the dependent variable. According to this note, the estimator based on the value will artificially depend on the choice of unit, while in principle the estimator based on the number is immune to this problem. However, even for count data, the NB estimator is sensitive to whether the dependent variable is measure in the actual number, in hundreds, in thousands, etc.

Section 2 provides the proof of the scale-dependence of the NB estimator, and section 3 illustrates this proposition with an application based on the trade gravity equation.

2 Proof

The specification is $y_i = exp(X_i \ \beta + u_i)$ where u_i is the residual. The first-order conditions for PPML, NB PML and gamma PML are, respectively (Gourieroux, Monfort and Trognon ; 1984a,b):

$$\text{PPML:} \qquad \sum_i \big(y_i - exp(X_i \ \beta)\big) \ X_i = 0 \qquad (1)$$

$$\text{NB PML:} \qquad \sum_i \big(1 + \alpha \ exp(X_i \ \beta)\big)^{-1} \big(y_i - exp(X_i \ \beta)\big) \ X_i = 0 \qquad (2)$$

$$\text{gamma PML:} \quad \sum_i exp(-X_i \ \beta) \big(y_i - exp(X_i \ \beta)\big) \ X_i = 0 \qquad (3)$$

Whereas the underlying assumption of the PPML and gamma PML is that the conditional variance is proportional to the conditional expectancy and to its square, respectively, the NB PML assumes that $Var(y|X) = E(y|X) + \alpha \ E^2(y|X)$, where α is a constant, generally considered to be positive. Eq. (1-3) confirm that when $\alpha \to 0$, NB PML $\to$ PPML, while when $\alpha \to \infty$, NB PML $\to$ gamma PML.

4

This note focuses on the impact of using $\tilde{y} = \lambda \, y$ as the dependent variable instead of y where λ is a scalar that can be either very small or very large depending on the unit choice. The first-order conditions indicate that both the Poisson and gamma estimator are independent of scale, as only the constant, denoted β_0, is affected by the linear transformation according to $\tilde{\beta}_0(\lambda) = \beta_0 + Log \, \lambda$, such that $exp(\tilde{\beta}_0(\lambda)) = \lambda \, exp(\beta_0)$. This implies that $exp(X_i \, \tilde{\beta}(\lambda)) = \lambda \, exp(X_i \, \beta) \quad \forall i$, and the FOC (1) and (3) are unaffected by scale[2]:

$$\begin{cases} \tilde{\beta}_0(\lambda) = \beta_0 + Log \, \lambda \\ \tilde{\beta}_k(\lambda) = \beta_k \quad \forall k \neq 0 \end{cases} \Rightarrow \tag{4}$$

$$\text{PPML:} \quad \sum_i \big(\tilde{y}_i - exp(X_i \, \tilde{\beta}) \big) \, X_i = 0 \Leftrightarrow \lambda \sum_i \big(y_i - exp(X_i \, \beta) \big) \, X_i = 0 \tag{5}$$

$$\text{gamma PML:} \quad \sum_i exp(-X_i \, \tilde{\beta}) \, \big(\tilde{y}_i - exp(X_i \, \tilde{\beta}) \big) \, X_i = 0 \Leftrightarrow \sum_i exp(-X_i \, \beta) \, \big(y_i - exp(X_i \, \beta) \big) \, X_i = 0 \tag{6}$$

In contrast, the first-order condition for NB PML (eq. 2) is sensitive to λ. When $\tilde{y}$ is the dependent variable, that condition is:

$$\sum_i \big(1 + \tilde{\alpha} \, exp(X_i \, \tilde{\beta}) \big)^{-1} \, \big(\tilde{y}_i - exp(X_i \, \tilde{\beta}) \big) \, X_i = 0 \tag{7}$$

For $\tilde{\beta}$ to be unaffected (except the intercept) by the transformation, the FOC must be:

$$\sum_i \big(1 + \tilde{\alpha} \, \lambda \, exp(X_i \, \beta) \big)^{-1} \, \lambda \big(y_i - exp(X_i \, \beta) \big) \, X_i = 0 \tag{8}$$

The comparison of (2) and (8) implies that the condition under which the NB PML estimator is independent of λ is $\tilde{\alpha}(\lambda) = \alpha/\lambda$, with $\alpha \equiv \tilde{\alpha}(\lambda = 1)$.[3]

However, this condition is violated in general. Let us first consider the two-step estimator implemented in various econometric softwares. The first step consists in computing a consistent

---

[2] $\beta \equiv \tilde{\beta}(\lambda = 1)$.

[3] Another way to see this is as follows. The NB PML assumption is $Var(\tilde{y}(\lambda)|X) = E[\tilde{y}(\lambda)|X] + \tilde{\alpha}(\lambda) \, E^2[\tilde{y}(\lambda)|X]$. Under the condition that $\tilde{\beta}$ is independent from λ (except $\tilde{\beta}_0$), this becomes: $\lambda^2 \, Var[y|X] = \lambda \, E[y|X] + \lambda^2 \, \tilde{\alpha}(\lambda) \, E^2[y|X] \Leftrightarrow Var[y|X] = 1/\lambda \big(E[y|X] + \lambda \, \tilde{\alpha}(\lambda) \, E^2[y|X] \big)$. Independence of the estimator with respect to λ implies that $\lambda \, \tilde{\alpha}(\lambda) = \alpha$.

estimator, *e.g.* PPML which is used in most softwares (Stata, SAS), and $\hat{y}_i$ denotes the first-step estimated observations. In a second step, α is estimated by OLS from the following regression:

$$(y_i - \hat{y}_i)^2 - \hat{y}_i = \alpha \ \hat{y}_i^{\ 2} + \epsilon_i \tag{9}$$

where ϵ_i is a residual, which yields:

$$\hat{\alpha} = \frac{\sum_i \left[(y_i - \hat{y}_i)^2 - \hat{y}_i \right] \hat{y}_i^{\ 2}}{\sum_i \hat{y}_i^{\ 4}} \tag{10}$$

This corresponds to the Quasi-generalized PML estimator for NB proposed by Gourieroux *et al.* (1984b), renamed "two-step NB" by Wooldridge (1999), and for which Head *et al.* (2009) provide a Stata code.

What happens to this estimator when the linear transformation $\tilde{y} = \lambda \ y$ is used as the dependent variable? As shown before, the first-step PPML estimator is unaffected by scale, hence $\hat{\tilde{y}}_i(\lambda) = \lambda \ \hat{y}_i$. It follows that:

$$\begin{aligned} \lambda \ \hat{\tilde{\alpha}}(\lambda) & = \lambda \frac{\sum_i \left[(\tilde{y}_i - \hat{\tilde{y}}_i)^2 - \hat{\tilde{y}}_i \right] \hat{\tilde{y}}_i^2}{\sum_i \hat{\tilde{y}}_i^4} = \lambda \frac{\sum_i \left[\lambda^2 \ (y_i - \hat{y}_i)^2 - \lambda \ \hat{y}_i \right] \lambda^2 \ \hat{y}_i^{\ 2}}{\lambda^4 \ \sum_i \hat{y}_i^{\ 4}} \\ & = \frac{\sum_i \left[\lambda \ (y_i - \hat{y}_i)^2 - \hat{y}_i \right] \hat{y}_i^{\ 2}}{\sum_i \hat{y}_i^{\ 4}} = \hat{\alpha} + (\lambda - 1) \frac{\sum_i (y_i - \hat{y}_i)^2 \ \hat{y}_i^{\ 2}}{\sum_i \hat{y}_i^{\ 4}} \end{aligned} \tag{11}$$

which proves that:

- $\lambda \ \hat{\tilde{\alpha}}(\lambda) \neq \hat{\alpha}$ as soon as $\lambda \neq 1$;

- when $\lambda \to +\infty$, $\lambda \ \hat{\tilde{\alpha}}(\lambda) \to +\infty$ and NB PML $\to$ gamma PML ;

- when $\lambda \to 0$, $\lambda \ \hat{\tilde{\alpha}}(\lambda) \to \hat{\alpha} - \frac{\sum_i (y_i - \hat{y}_i) \ \hat{y}_i^{\ 2}}{\sum_i \hat{y}_i^{\ 4}} = -\frac{\sum_i \hat{y}_i^{\ 3}}{\sum_i \hat{y}_i^{\ 4}}$. When the software constrains the estimated value to be positive (*e.g.* Stata), $\lambda \ \hat{\tilde{\alpha}}(\lambda) \to 0$, and NB PML $\to$ PPML.

Section 3 provides an empirical example illustrating these results.

This problem is not an artefact of using a two-step estimator. Even the theoretical NB PML estimator exhibits scale dependence. To see this, calculate instead of eq. (9) the FOC on

6

α. The full log-likelihood is:[4]

$$
\begin{aligned}
LogL \;=\; & \sum_i Log\Big(\frac{\Gamma(\alpha^{-1}+y)}{\Gamma(\alpha^{-1})\Gamma(y+1)}\Big) - \alpha^{-1}Log[\alpha] + y_i\,X_i\,\beta \\
& - \big(\alpha^{-1}+y_i\big)\Big(Log[1+\alpha\,exp(X_i\,\beta)] - Log[\alpha]\Big)
\end{aligned}
\tag{12}
$$

where Γ is the standard Gamma function.[5]

While differentiating this expression with respect to β leads to (2), differentiating with respect to α entails:[6]

$$
\alpha^{-2}\sum_i\left(Log[1+\alpha\,exp(X_i\,\beta)] + \frac{y_i - exp(X_i\,\beta)}{\alpha^{-1}+exp(X_i\,\beta)} - \sum_{k=0}^{+\infty}\frac{y_i}{(k+\alpha^{-1})(k+\alpha^{-1}+y_i)}\right) = 0 \tag{13}
$$

With the linear transformation, the FOC becomes:

$$
\tilde{\alpha}^{-2}\sum_i\left(Log[1+\tilde{\alpha}\,exp(X_i\,\tilde{\beta})] + \frac{\tilde{y}_i - exp(X_i\,\tilde{\beta})}{\tilde{\alpha}^{-1}+exp(X_i\,\tilde{\beta})} - \sum_{k=0}^{+\infty}\frac{\tilde{y}_i}{(k+\tilde{\alpha}^{-1})(k+\tilde{\alpha}^{-1}+\tilde{y}_i)}\right) = 0 \tag{14}
$$

As seen above, the non-scale-dependence of β (except the intercept) is equivalent to $exp(X_i\,\tilde{\beta}(\lambda)) = \lambda\,exp(X_i\,\beta) \quad \forall i$ and $\tilde{\alpha}(\lambda) = \alpha/\lambda$. By absurd reasoning, the FOC is then written:

$$
\lambda^2\,\alpha^{-2}\sum_i\left(Log[1+\alpha\,exp(X_i\,\beta)] + \frac{\lambda\,y_i - \lambda\,exp(X_i\,\beta)}{\lambda\,\alpha^{-1}+\lambda\,exp(X_i\,\beta)} - \sum_{k=0}^{+\infty}\frac{\lambda\,y_i}{(k+\lambda\,\alpha^{-1})(k+\lambda\,\alpha^{-1}+\lambda\,y_i)}\right) = 0
\tag{15}
$$

$$
\Leftrightarrow \sum_i\left(Log[1+\alpha\,exp(X_i\,\beta)] + \frac{y_i - exp(X_i\,\beta)}{\alpha^{-1}+exp(X_i\,\beta)}\right) = \sum_i\sum_{k=0}^{+\infty}\frac{y_i}{(\frac{k}{\lambda}+\alpha^{-1})(\frac{k}{\lambda}+\alpha^{-1}+y_i)} \tag{16}
$$

which is absurd given that α is defined according to (13) by:

$$
\Leftrightarrow \sum_i\left(Log[1+\alpha\,exp(X_i\,\beta)] + \frac{y_i - exp(X_i\,\beta)}{\alpha^{-1}+exp(X_i\,\beta)}\right) = \sum_i\sum_{k=0}^{+\infty}\frac{y_i}{(k+\alpha^{-1})(k+\alpha^{-1}+y_i)} \tag{17}
$$

That is, the FOC on α is scale-dependent as the right hand side term $\sum_i\sum_{k=0}^{+\infty}\frac{y_i}{(\frac{k}{\lambda}+\alpha^{-1})(\frac{k}{\lambda}+\alpha^{-1}+y_i)}$ in eq. (16) depends on λ. The empirical example developed in the following section illustrates

---

[4]Calculations details are provided in Appendix A.

[5]$\Gamma(x) = \int_0^{+\infty} t^{x-1}\,e^{-t}\,dt.$

[6]Calculations details are provided in Appendix B.

that the direct estimation of $\tilde{\alpha}(\lambda)$ based on eq. (14) violates the condition that $\lambda \ \tilde{\alpha}(\lambda)$ is independent of scale.

3 Application to the trade gravity equation

3.1 Data

Trade flow data are taken from the IMF Direction Of Trade Statistics (DOTS) database. For the year 2000, there are 21,543 non-zero flows between 196 trading partners.[7] The geographical variables (distance between countries, common border, common language and colonial linkage dummies) are provided by the CEPII database[8], and the FTA data are based on Fontagné and Zignago (2007) who improve those used by Baier and Bergstrand (2007).

3.2 Specification

The bilateral trade equation is estimated following Anderson and van Wincoop (2003) according to:

$$x_{ij} = exp(\beta_0 + \beta_1 \ Log \ d_{ij} + \beta_2 \ B_{ij} + \beta_3 \ L_{ij} + \beta_4 \ C_{ij} + \beta_5 \ FTA_{ij} + FX_i + FM_j)u_{ij} \qquad (18)$$

where x_{ij} is the value of export from country i to country j, FX_i and FM_j are exporting and importing countries fixed effects, respectively. B_{ij}, L_{ij} and C_{ij} are the traditionnal control covariates: common border, common official language and colonial linkage dummies, respectively. The u_{ij} are the multiplicative error terms of the nonlinear estimates. Following Anderson and van Wincoop (2003), importer and exporter fixed effects are used to control for multilateral resistance terms as well as for the income levels of both importers and exporters.

---

[7]Focusing on non-zero flows is sufficient for illustration purposes. Including zero flows or focusing on other years unsurprisingly leads to the same conclusion as the proof in section 2 is general.

[8]http://www.cepii.fr/anglaisgraph/bdd/distances.htm, Centre d'Etudes Prospectives et d'Informations Internationales.

8

3.3 Results

Table 1 shows the estimates of equation (18) from the PPML and gamma PML estimators, in the first and last columns, respectively. The columns in between report the NB PML estimates based on different unit values for trade flows. Is is computed either by the Stata *nbreg* (or *glm* with *nbinomial* family) estimator, the SAS *proc genmod* procedure or the two-step Head *et al.* (2009)'s code, which yield identical estimated values. When flows are measured with a very large unit (trillions of US$), which means that flow values are very small, NB PML and PPML estimates are visually identical. At the other extreme, when flows take very large values (*i.e.* when the unit is small such as thousands of US$), the NB PML are very close to the gamma PML estimates. This illustrates that the NB PML estimator is inappropriate as the estimates depend arbitrarily on the unit choice of the dependent variable.

Table 2 compares the dispersion parameter $\tilde{\alpha}(\lambda)$ estimated by various NB PML estimators, and shows that for all of them the condition under which these estimators are not sensitive to scale (*i.e.* $\lambda \, \tilde{\alpha}(\lambda)$ does not depend on λ) is violated. The NB PML estimators that are compared are: that computed by Stata, that by SAS, the two-step estimator using PPML in the first step according to eq. (9-11), the two-step estimator using the geometric estimator in the first step according to Head *et al.* (2009), and the one-step estimator computing $\tilde{\alpha}$ such that likelihood is maximized (eq. 14, using Newton algorithm and PPML for β). Both Stata and SAS compute an iterated estimator, reestimating eq. (9-11) at each step and starting with PPML, with $\tilde{\alpha}$ being the final iterated value, and yield similar results. For all these NB PML estimators, $\lambda \, \tilde{\alpha}$ depend on λ, converging towards zero when λ becomes small and infinity when λ becomes large. That is, all NB PML estimators converge towards PPML and GPML, respectively.

4 Conclusion

Although it is being increasingly used, the NB PML is not appropriate when the unit choice of the dependent variable is arbitrary.

Appendix

A   Log-likelihood of the Negative Binomial estimator

Negative Binomial density:

$$Pr[Y = y] = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(y+1)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu + \alpha^{-1}}\right)^{y} \tag{19}$$

Log-likelihood:

$$LogL = \sum_i A_i + \alpha^{-1}\Big(Log[\alpha^{-1}] - Log[\alpha^{-1} + \mu_i]\Big) + y_i\Big(Log[\mu_i] - Log[\mu_i + \alpha^{-1}]\Big) \tag{20}$$

with:

$$A_i = Log\left(\frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})\Gamma(y_i + 1)}\right) \tag{21}$$

Replacing μ_i by $exp(X_i\ \beta)$ leads to:

$$LogL = \sum_i A_i - \alpha^{-1}\Big(Log[\alpha] + Log[\alpha^{-1} + exp(X_i\ \beta)]\Big) + y_i\Big(Log[exp(X_i\ \beta)] - Log[exp(X_i\ \beta) + \alpha^{-1}]\Big) \tag{22}$$

$$\Leftrightarrow LogL = \sum_i A_i - \alpha^{-1}Log[\alpha] - \alpha^{-1}Log[\alpha^{-1} + exp(X_i\ \beta)] + y_i\ X_i\ \beta - y_i\ Log[exp(X_i\ \beta) + \alpha^{-1}] \tag{23}$$

$$\Leftrightarrow LogL = \sum_i A_i - \alpha^{-1}Log[\alpha] + y_i\ X_i\ \beta - \big(\alpha^{-1} + y_i\big)Log[\alpha^{-1} + exp(X_i\ \beta)] \tag{24}$$

Rewriting:

$$Log[\alpha^{-1} + exp(X_i\ \beta)] = Log[\alpha^{-1}\big(1 + \alpha\ exp(X_i\ \beta)\big)] = Log[\alpha^{-1}] + Log[1 + \alpha\ exp(X_i\ \beta)] \tag{25}$$

yields:

$$LogL = \sum_i A_i - \alpha^{-1}Log[\alpha] + y_i\ X_i\ \beta - \big(\alpha^{-1} + y_i\big)\Big(Log[1 + \alpha\ exp(X_i\ \beta)] - Log[\alpha]\Big) \tag{26}$$

Rewriting:

$$B_i = y_i \, X_i \, \beta - \left(\alpha^{-1} + y_i\right) \, Log[1 + \alpha \, exp(X_i \, \beta)] \qquad (27)$$

leads to:

$$LogL = \sum_i A_i + B_i - \alpha^{-1} Log[\alpha] + \left(\alpha^{-1} + y_i\right) \, Log[\alpha] = \sum_i A_i + B_i + y_i \, Log[\alpha] \qquad (28)$$

Consistent with $\sum_i B_i$ as the objective function with respect to β in Gourieroux, Monfort and Trognon (1984a).

B   First-order condition with respect to α

For derivating the Gamma function, the digamma function, denoted ψ is used:[9]

$$\psi(x) = D \, Log[\Gamma(x)] = \frac{\Gamma'(x)}{\Gamma(x)} = -\gamma - \frac{1}{x} + x \sum_{k=1}^{+\infty} \frac{1}{k(k+x)} \qquad (29)$$

where D is the differential operator and γ the Euler-Mascheroni' constant.

Dropping terms without α from (28) leads to the objective function with respect to α:

$$\sum_i Log\left(\frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})}\right) - \left(\alpha^{-1} + y_i\right) \, Log[1 + \alpha \, exp(X_i \, \beta)] + y_i \, Log[\alpha] \qquad (30)$$

Rewriting:

$$C_i = Log\left(\frac{\Gamma(\alpha^{-1} + y_i)}{\Gamma(\alpha^{-1})}\right) \qquad (31)$$

and:

$$D_i = -\left(\alpha^{-1} + y_i\right) \, Log[1 + \alpha \, exp(X_i \, \beta)] + y_i \, Log[\alpha] \qquad (32)$$

the FOC would be $\sum_i C_i' + D_i' = 0$. Differentiating C_i leads to:

$$C_i' = -\alpha^{-2}\left(-\gamma - \frac{1}{\alpha^{-1} + y_i} + (\alpha^{-1} + y_i)\sum_{k=1}^{+\infty}\frac{1}{k(k+\alpha^{-1}+y_i)} - \left(-\gamma - \frac{1}{\alpha^{-1}} + \alpha^{-1}\sum_{k=1}^{+\infty}\frac{1}{k(k+\alpha^{-1})}\right)\right) \qquad (33)$$

---

[9]Andrews, Askey and Roy (1999)

$$C_i' = -\alpha^{-2} \left(\frac{1}{\alpha^{-1}} - \frac{1}{\alpha^{-1} + y_i} + (\alpha^{-1} + y_i) \sum_{k=1}^{+\infty} \frac{1}{k(k + \alpha^{-1} + y_i)} - \alpha^{-1} \sum_{k=1}^{+\infty} \frac{1}{k(k + \alpha^{-1})} \right) \quad (34)$$

$$C_i' = -\alpha^{-2} \left(\frac{y_i}{\alpha^{-1}(\alpha^{-1} + y_i)} + y_i \sum_{k=1}^{+\infty} \frac{1}{k(k + \alpha^{-1} + y_i)} + \alpha^{-1} \sum_{k=1}^{+\infty} \left(\frac{1}{k(k + \alpha^{-1} + y_i)} - \frac{1}{k(k + \alpha^{-1})} \right) \right)$$
$$(35)$$

Rewriting E_i, the last term within the parenthesis:

$$E_i = \alpha^{-1} \sum_{k=1}^{+\infty} \left(\frac{1}{k(k + \alpha^{-1} + y_i)} - \frac{1}{k(k + \alpha^{-1})} \right) = \alpha^{-1} \sum_{k=1}^{+\infty} \frac{-y_i}{k(k + \alpha^{-1} + y_i)(k + \alpha^{-1})} \quad (36)$$

$$= -y_i \sum_{k=1}^{+\infty} \frac{\alpha^{-1}}{k(k + \alpha^{-1} + y_i)(k + \alpha^{-1})} \quad (37)$$

leads to:

$$C_i' = -\alpha^{-2} \left(\frac{y_i}{\alpha^{-1}(\alpha^{-1} + y_i)} + y_i \sum_{k=1}^{+\infty} \frac{1}{k(k + \alpha^{-1} + y_i)} - y_i \sum_{k=1}^{+\infty} \frac{\alpha^{-1}}{k(k + \alpha^{-1} + y_i)(k + \alpha^{-1})} \right) \quad (38)$$

$$\Leftrightarrow C_i' = -\alpha^{-2} \left(\frac{y_i}{\alpha^{-1}(\alpha^{-1} + y_i)} + y_i \sum_{k=1}^{+\infty} \frac{1}{(k + \alpha^{-1})(k + \alpha^{-1} + y_i)} \right) \quad (39)$$

$$\Leftrightarrow C_i' = -\alpha^{-2} \sum_{k=0}^{+\infty} \frac{y_i}{(k + \alpha^{-1})(k + \alpha^{-1} + y_i)} \quad (40)$$

Differentiating D_i leads to:

$$D_i' = \alpha^{-2} \, Log[1 + \alpha \, exp(X_i \, \beta)] - \left(\alpha^{-1} + y_i \right) \frac{exp(X_i \, \beta)}{1 + \alpha \, exp(X_i \, \beta)} + \frac{y_i}{\alpha} \quad (41)$$

$$\Leftrightarrow D_i' = \alpha^{-2} \, Log[1 + \alpha \, exp(X_i \, \beta)] + \frac{y_i - exp(X_i \, \beta)}{\alpha \left(1 + \alpha \, exp(X_i \, \beta) \right)} \quad (42)$$

$$\Leftrightarrow D_i' = \alpha^{-2} \left(Log[1 + \alpha \, exp(X_i \, \beta)] + \frac{y_i - exp(X_i \, \beta)}{\alpha^{-1} + exp(X_i \, \beta)} \right) \quad (43)$$

which yields the FOC with respect to α:

$$\alpha^{-2} \sum_i \left(Log[1 + \alpha \, exp(X_i \, \beta)] + \frac{y_i - exp(X_i \, \beta)}{\alpha^{-1} + exp(X_i \, \beta)} - \sum_{k=0}^{+\infty} \frac{y_i}{(k + \alpha^{-1})(k + \alpha^{-1} + y_i)} \right) = 0 \quad (44)$$

12

References

Anderson, J. and van Wincoop, E. (2003), 'Gravity with Gravitas: A Solution to the Border Puzzle', *American Economic Review* **93**(1), 170–192.

Andrews, G., Askey, R. and Roy, R. (1999), *Special Functions*, Cambridge: Cambridge University Press.

Baier, S. L. and Bergstrand, J. H. (2007), 'Do Free Trade Agreements Actually Increase Members International Trade?', *Journal of International Economics* **71**(1), 72–95.

Briant, A., Combes, P.-P. and Lafourcade, M. (2009), 'Product Complexity, Quality of Institutions and the Pro-Trade Effect of Immigrants', *CEPR Discussion Paper* (7192).

Burger, M., van Oort, F. and Linders, G.-J. (2009), 'On the Specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation', *Spatial Economic Analysis* **4**(2), 167–190.

Fontagné, L. and Zignago, S. (2007), 'A Re-evaluation of the Impact of Regional Agreements on Trade Patterns', *Economie Internationale* **109**(1), 31–51.

Garita, G. and van Marrewijk, C. (2008), 'Countries of a Feather Flock Together – Mergers and Acquisitions in the Global Economy', *Tinbergen Institute Discussion Paper* **067**(2).

Gourieroux, C., Monfort, A. and Trognon, A. (1984*a*), 'Pseudo Maximum Likelihood Methods: Applications to Poisson Models', *Econometrica* **52**(3), 701–720.

Gourieroux, C., Monfort, A. and Trognon, A. (1984*b*), 'Pseudo Maximum Likelihood Methods: Theory', *Econometrica* **52**(3), 681–700.

Head, K., Mayer, T. and Ries, J. (2009), 'How Remote is the Offshoring Threat?', *European Economic Review* (53), 429–444.

Santos Silva, J. and Tenreyro, S. (2006), 'The Log of Gravity', *The Review of Economics and Statistics* **88**(4), 641–658.

Westerlund, J. and Wilhelmsson, F. (2009), 'Estimating the Gravity Model Without Gravity Using Panel Data', *forthcoming in Applied Economics* .

Wooldridge, J. M. (1999), Quasi-likelihood methods for count data, *in* M. H. Pesaran and P. Schmidt, eds, 'Handbook of Applied Econometrics Volume II: Microeconomics', Oxford: Blackwell, pp. 352–406.

Tables

<div align="center">

Table 1: Scale-dependence of the Negative-Binomial Estimator
Gravity equation ; 2000

</div>

	PPML	Negative Binomial PML				GPML
	PPML b/se	Tr. USD b/se	B. USD b/se	M. USD b/se	Th. USD b/se	GPML b/se
Distance	-0.606*** (0.032)	-0.606*** (0.032)	-0.758*** (0.035)	-1.259*** (0.023)	-1.232*** (0.024)	-1.231*** (0.024)
Contiguity dummy	0.644*** (0.099)	0.644*** (0.099)	0.500*** (0.094)	0.880*** (0.086)	0.959*** (0.090)	0.959*** (0.090)
Common-language dummy	0.154* (0.082)	0.154* (0.082)	0.167** (0.072)	0.513*** (0.047)	0.541*** (0.051)	0.541*** (0.051)
Colonial-tie dummy	0.175* (0.103)	0.175* (0.103)	0.536*** (0.125)	1.296*** (0.089)	1.290*** (0.100)	1.289*** (0.100)
Free-trade agreement dummy	0.479*** (0.088)	0.479*** (0.088)	0.322*** (0.073)	0.173*** (0.060)	0.249*** (0.061)	0.250*** (0.061)
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21543	21543	21543	21543	21543	21543

Notes: * p<0.1, ** p<0.05, *** p<0.01 ; PML = Pseudo-Maximum Likelihood, PPML = Poisson PML, NB = Negative Binomial, GPML = gamma PML ; USD = United States Dollars, Tr.=Trillions, B.=Billions, M.=Millions, Th.=Thousands. Fixed effects are importer and exporter country fixed effects.

<div align="center">

Table 2: Estimation of the dispersion parameter $\alpha = \lambda\ \tilde{\alpha}$

</div>

unit	Tr. USD	B. USD	M. USD	Th. USD
Stata	0	0.9e-4	1.75	2,510
SAS	0	0.9e-4	1.75	18,186
two-step (first step=PPML, eq. 9-11)	0	0.2e-4	0.02	22
two-step (first step=geometric, HMR)	0	3.6e-4	0.88	900
one-step (eq. 14)	1.4e-8	0.6e-4	2.50	3,734

Notes: Stata = *nbreg* procedure, SAS = *genmod* procedure, HMR = Head *et al.* (2009) ; USD = United States Dollars, Tr.=Trillions, B.=Billions, M.=Millions, Th.=Thousands.