



AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
<http://ageconsearch.umn.edu>
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Estimation Issues in Single Commodity Gravity Trade Models

Sören Prehn & Bernhard Brümmer

**Georg August Universität Göttingen
Department für Agrarökonomie und Rurale Entwicklung
sprehn@gwdg.de**



**Paper prepared for presentation at the EAAE 2011 Congress
Change and Uncertainty
Challenges for Agriculture,
Food and Natural Resources**

August 30 to September 2, 2011
ETH Zurich, Zurich, Switzerland

Copyright 2011 by [Sören Prehn & Bernhard Brümmer]. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

Estimation Issues in Single Commodity Gravity Trade Models

Abstract

Recently gravity trade models are applied to disaggregated trade data. Here many zeros are characteristic. In the presence of excess zeros usual Poisson Pseudo Maximum Likelihood (PPML) is still consistent, the variance covariance matrix however is invalid. Correct economic interpretation however requires also the last. So alternative estimators are looked for. [Staub & Winkelmann \[2010\]](#) argue that zero-inflated count data models (i.e. zero-inflated Poisson / Negative Binomial Pseudo Maximum Likelihood (ZIPPM / ZINBPML)) are no alternative since under model misspecification these estimators are inconsistent. Yet zero-inflated Poisson Quasi-Likelihood (PQL) is a reliable alternative. It is consistent even under model misspecifications and beyond that robust against unobserved heterogeneity. Another alternative is a log-skew-normal Two-Part Model (G2PM) which generalizes the standard log-normal Two-Part Model (2PM). It is insofar advantageous as it adjusts for (negative) skewness and regression coefficients retain usual interpretations as in log-normal models. PQL is useful for multiplicative gravity model estimation and G2PM for log-linear gravity model estimation. Exemplarily the estimators are applied to intra-European piglet trade to assess their empirical performance and applicability for single commodity trade flow analysis. The empirical part favours PQL but G2PM is a reliable alternative for other trade flow analyses. PQL and G2PM should become standard tools for single commodity trade flow analysis.

Keywords: Gravity Model, Homogeneous Firm Trade Model, Excess Zeros, Overdispersion, Negatively Skewed Distribution, Poisson Quasi Likelihood, Two Part Model

Introduction

Recently research focus in gravity trade model analysis shifted from an aggregated macro-level to a disaggregated micro-level. This shift is not only statistically motivated but also empirically. Statistically it is more appropriate to disaggregate data and then to reaggregate the corresponding micro-level results than to do the whole analysis at an aggregated macro-level. The latter approach potentially suffers under an aggregation bias. Empirically the research interest becomes more influenced by policy. It is more meaningful to get policy results at micro-level than at macro-level. The economic effects of single politics are often neglectable at macro-level but not so at micro-level. Single commodity gravity trade models are now in the focus of interest in gravity trade model analysis.

Although there are applications of gravity trade models to disaggregated data there are only few papers which explicitly deal with estimation issues of single commodity gravity trade models¹. An exemption is the paper by [Burger et al. \[2009\]](#). Here the authors seize on the dispute of methods between multiplicative and log-linear gravity models. In their argumentation they partly follow [Santos Silva & Tenreyro \[2006\]](#) that logarithmising leads to biased gravity estimates but they claim that for single commodity trade flow analysis the Poisson Pseudo Maximum Likelihood (PPML) estimator is inappropriate. The PPML suffers under the problems of excess zeros and overdispersion; the former is a consequence of disaggregation which naturally

¹This paper focuses on estimation issues of single commodity gravity trade models its statements however should also apply for sector gravity trade models.

increases the number of zero trade flows the latter is caused by unobserved heterogeneity.² Instead they recommend to apply zero-inflated Poisson Pseudo Maximum Likelihood / Negative Binomial Pseudo Maximum Likelihood (ZIPPML / ZINBPML). The latter approach not only deals with excess zeros but also with overdispersion.

In a recent paper [Staub & Winkelmann \[2010\]](#) however show that ZIPPML / ZINBPML are only consistent as long as the underlying model is not misspecified. If the model is misspecified the estimators are biased. The authors further show that PPML is still consistent even in the presence of excess zeros, but the variance covariance matrices are invalid. In their paper the authors instead recommend a zero-inflated Poisson Quasi-Likelihood (PQL). This estimator is not only consistent in the presence of excess zeros but also practically unaffected by unobserved heterogeneity. [Staub & Winkelmann's \[2010\]](#) findings are important since they question [Burger et al.'s \[2009\]](#) statements!

The problem of zero trade flows is also tackled in log-linear gravity models. Different estimators are applied here. In the presence of excess zeros Two-Part Models seem statistically to be the most reliable.³ [Hillberry \[2002\]](#) proposes the use of standard log-normal Two-Part Models (2PM). Since logarithmising sometimes leads to (negatively) skewed distributions, a more general Two-Part Model is preferable. Recently [Chai & Bailey \[2008\]](#) develop a log-skew-normal 2PM (G2PM) which not only adjusts for asymmetries but also the regression coefficients still have the usual interpretations as in the standard log-normal 2PM.

Consistent estimation is always important for economic interpretation. This applies not only but especially for single commodity trade flow analysis where estimates suffer beside unobserved heterogeneity also under excess zeros. New estimators are looked for. This paper therefore proposes PQL and G2PM as reliable alternatives. Both estimators are applied PQL and G2PM to disaggregated data (i.e. intra-European piglet trade) to evaluate their empirical applicability and performance. Further the paper recommends statistical testing procedures for model selection; overdispersion test for Poisson Models and Likelihood Ratio (LR) tests for Two-Part Models. Both estimators and model specification tests should become standard tools for single commodity gravity trade flow analysis!

The focus of this paper is just methodological. So for the theoretical model homogeneous firms trade models [[Felbermayr & Kohler, 2010](#), [Egger & Larch, 2011](#)] are recommended, but not further discussed. Homogeneous firms trade models have the advantage that they share the same properties as heterogeneous firms trade models [[Helpman et al., 2008](#)], but assume a homogeneous firm structure. This often fits better with market characteristics especially for primary sectors (i.e. agricultural sectors). Theoretical aspects however are not further discussed here.

This paper is organised as follows. The first section deals with the appropriate specification of the theoretical model; homogeneous firms trade models vs. heterogeneous firms trade models and multiplicative gravity models vs. log-linear gravity models. In the second section methodological aspects of single commodity gravity estimation are discussed. The advantage of PQL and G2PM are sketched. The next section then applies both new estimators to intra-European piglet trade. The paper concludes with some recommendations for future single commodity gravity trade model estimation.

Model Specification

Starting point of every gravity trade model analysis is the specification of an appropriate gravity model. Here the [Anderson & van Wincoop \[2003\]](#) (AvW) Model is standard for aggregated data. For disaggregated data however the AvW Model does not fit perfectly. The AvW Model

²For statistical detail on excess zeros and overdispersion see below.

³Tobit estimators are biased in the presence of excess zeros.

neither deals with zero trade flows which are manifold at micro-level nor it deals with asymmetries in trade flows which are caused by different degrees of specialisation [Helpman et al., 2008]. Helpman et al. propose a heterogeneous firms trade model which simultaneously deals with zero and asymmetric trade flows. A heterogeneous firm structure however does not always describe market structure adequately. Especially for primary sector (i.e. agriculture) the firm structure is more homogeneous than heterogeneous. Usually farm sector is characterised by many small and rather homogeneous farms. Recently two papers one by Felbermayr & Kohler [2010] and the other one by Egger & Larch [2011] develop homogeneous firms trade models (see Box 1). These models have the advantage that they deal with zero and asymmetric trade flows but also restrict firm structure to be homogeneous.

The homogeneous firms trade model depends on two parts. The first part deals with the extensive trade margin whereas the second part deals with the intensive trade margin. Practically spoken the first part asks the question whether trade occurs yes or no and the second part discusses the question to which extent trade takes place. The two parts can either be estimated together in a multiplicative gravity model

$$[1] \quad X_{ij} = \exp(\beta_0 + \lambda_i + \chi_j + \gamma d_{ij}) + u_{ij}$$

or they can be estimated separately. Here the extensive trade margin is estimated via a limited dependent variable model (i.e. Probit or Logit) and the intensive trade margin usually via a log-linear gravity model⁴

$$[2] \quad \log(X_{ij}) = \beta_0 + \lambda_i + \chi_j + \gamma d_{ij} + u_{ij}.$$

In both specifications X_{ij} is the import value of importer j from exporter i and β_0 is a constant which captures the effect of total sector production Y . $\lambda_i = y_i + (\sigma - 1) \pi_i$ and $\chi_j = e_j + (\sigma - 1) p_j$ are exporter and importer fixed effects which capture the effects of exporter i 's production Y_i and outward multilateral resistance Π_i and importer j 's expenditure E_j and inward multilateral resistance P_j respectively [Anderson, 2010]. d_{ij} is the gravitational distance and u_{ij} is an i.i.d. error term.⁵

Model Implementation and Estimation Issues

The second step in gravity trade model analysis is the econometric implementation of the theoretical model and its appropriate estimation. Here caution should be applied since not every functional form and estimation method is (always) appropriate. This especially applies for single commodity trade flow analysis where the problem of excess zeros and overdispersion is often immanent. Excess zeros hereby means that there are more zeros in the data than predicted by the theoretical model. Overdispersion is immanent when the observed variance is higher than the variance of the theoretical model. Non-consideration of both would lead to biased estimates.

In their seminal paper Santos Silva & Tenreyro [2006] argue gravity models not to estimate in log-linear form but in multiplicative form. Logarithmising could lead to inconsistent estimates. If heterogeneity is present Jensen's Inequality would apply (i.e. $\ln[E(x)] \neq E[\ln(x)]$) which then would render estimates inconsistent. They instead propose a PPML. This estimator is consistent

⁴Indeed Egger & Larch [2011] also propose to estimate the intensive trade margin in multiplicative form via PPML.

⁵All small letters indicate logarithms.

BOX 1: Homogeneous Firms Trade Models

Model 1^a (Felbermayr & Kohler [2010])

$$X_{ij} = \begin{cases} \frac{(\tau_{ij})^{1-\sigma}}{Q_j \Pi_i} \frac{Y_i E_j}{Y} \\ 0 \end{cases}$$
$$\Pi_i = \frac{Y_i}{Y} Q_i + \sum_j \frac{E_j}{Y} (\tau_{ij})^{1-\sigma} Q_j$$
$$Q_j = \frac{E_j}{Y} \Pi_j + \sum_i \frac{Y_i}{Y} (\tau_{ij})^{1-\sigma} \Pi_i$$

Model 2 (Egger & Larch [2011])

$$X_{ij} = I_{ij} \left(\frac{\tau_{ij}}{P_j \Pi_i} \right)^{1-\sigma} \frac{Y_i E_j}{Y}$$
$$\Pi_i^{1-\sigma} = \sum_j I_{ij} \left(\frac{\tau_{ij}}{P_j} \right)^{1-\sigma} \frac{E_j}{Y}$$
$$P_j^{1-\sigma} = \sum_i I_{ij} \left(\frac{\tau_{ij}}{\Pi_i} \right)^{1-\sigma} \frac{Y_i}{Y}$$

^aModel slightly modified to make it more Anderson & van Wincoop like.

even in the presence of heterogeneity and it has the appeal to deal with zero trade flows. Burger et al. [2009] extend this framework for more disaggregated data. The authors partly follow in their argumentation Santos Silva & Tenreyro [2006] that logarithmising leads to biased gravity estimates but they claim that for single commodity trade flow analysis PPML is not appropriate. PPML suffers under the problems of excess zeros and overdispersion. The two problems have to be separated from each other. So excess zeros is a consequence of disaggregation which naturally increases the number of zero trade flows and overdispersion is a consequence of unobserved heterogeneity caused by a omitted variable problem. The authors instead recommend to apply zero-inflated Poisson / Negative Binomial Pseudo Maximum Likelihood (ZIPML / ZINBPM). The latter approach not only deals with excess zeros but also with overdispersion. It is important to mention here that Burger et al. do not develop an asymptotic theory!

An asymptotic theory is just recently developed by Staub & Winkelmann [2010]. Utilising the framework of Gourieroux et al.'s [1984a, 1984b] seminal papers on Pseudo Maximum Likelihood (PML) Staub & Winkelmann can show that ZIPML / ZINBPM are only consistent as long as the underlying model is not misspecified. If the model is misspecified the estimators are biased. The authors further show that PPML is still consistent even in the presence of excess zeros but the variance covariance matrices are invalid. In their paper the authors instead recommend a zero-inflated Poisson Quasi-Likelihood (PQL). This estimator is not only consistent in the presence of excess zeros but also practically unaffected by unobserved heterogeneity. Staub & Winkelmann's [2010] findings are important since they question Burger et al.'s [2009] statements!

If the focus is on log-linear gravity model estimation then the importance of excess zeros also

increases with the degree of disaggregation. For moderately disaggregated data usual 2-step estimators (i.e. Tobit estimators) are consistent as long as 'the sample proportion of zeros is roughly equivalent to the left tail area of the assumed parametric distribution' [Chai & Bailey, 2008, pg. 3644]. If the tail-probability constraint is not fulfilled then estimators are biased. Since the tail-probability constraint is obviously hurted in the presence of excess zeros 2-step estimators are inappropriate for single commodity trade flow analysis. In the presence of excess zeros Two-Part Models [Cragg, 1971] which relax the tail-probability constraint seem statistically to be more reliable. Hillberry [2002] proposes the use of standard log-normal Two-Part Models (2PM). Since logarithmising sometimes leads to (negatively) skewed distributions a more general Two-Part Model is preferable. Recently Chai & Bailey [2008] develop a log-skew-normal Two-Part Model (G2PM) which not only adjusts for asymmetries but also regression coefficients still have the usual interpretation as in standard log-normal models. Beside model specification and estimation model selection is another important task in gravity trade model analysis. So far there are only few papers which deal with testing procedures which allow to judge between different estimation methods. For Poisson Models overdispersion test are recommended [Burger et al., 2009] and for Two-Part Models Likelihood Ratio (LR) tests.⁶

Application: Intra-European Piglet Trade

The previous section discusses the statistical superiority of PQL over ZIPPM / ZINBPML and G2PM over 2PM respectively. PQL and G2PM are now applied to intra-European piglet trade to illustrate their empirical applicability.⁷ Estimation results are presented in Table 1. The homogeneous firms trade model is once estimated in multiplicative form via PPML and PQL and once in log-linear form via 2PM and G2PM. The benchmark model (i.e. traditional gravity model [Anderson & van Wincoop, 2003]) is estimated via Ordinary Least Squares (OLS). A fixed effects structure with importer, exporter and time fixed effects is assumed for each model. Import data are extracted from the Statistical Office of the European Union (Eurostat); physical distance data from the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII); all other data are self-constructed. The data frequency is annual, starting from 2000 to 2009. Following Marquer [2010] the trade flow analysis concentrates on the eight most important exporters and importers i.e. Austria, Germany, Denmark, Spain, Italy, The Netherlands, Poland and Romania.

Gravitational distance is approximated by physical distance (Dist), presence of a common border (CONTIG), presence of a common language (COMLANG), a binary indicator for trade between member states and new member states before EU enlargement (NMS_inter) and a binary indicator for trade between new member states before EU enlargement (NMS_intra).

The signs of all models are in accordance. Market characteristics and market developments are adequately represented by estimation results. A common border favours piglet trade as well as a common language. The trade reversal indicated by the indicator NMS_inter fits well with market developments. After the EU enlargement large commercial hog producers as Smithfield opened up commercial pig plants in East Europe. This decreases exports to member states of the EU15 as indicated. The other indicator NMS_intra indicates that piglet trade between new member states decreases after EU enlargement. This decrease can be explained by a large decrease in sow stocks in East European member states after 2005 [Marquer, 2010]. The time fixed effects also seem reasonably to capture global market developments. The time fixed effects fol-

⁶Generals test to decide between different model specifications (e.g. Poisson vs. Two-Part Model) are at the moment at the research frontier [Santos Silva et al., 2010].

⁷For details on European pig farming and recent developments see Marquer [2010].

Table 1: Overview Gravity Model Estimation Results

	LOGIT	OLS	PPML	PQL	2PM	G2PM
INTERCEPT	-10.473*** (3.470)	11.915*** (4.329)	-4.450 (5.499)	-4.450 (5.428)	11.915*** (4.329)	10.724*** (4.205)
log(DIST)	0.841** (0.375)	-0.305 (0.566)	0.869 (0.709)	0.869 (0.702)	-0.305 (0.566)	-0.152 (0.547)
CONTIG	0.844** (0.390)	0.393 (0.897)	2.377*** (0.904)	2.377*** (0.894)	0.393 (0.897)	0.372 (0.858)
COMLANG	-0.381 (0.385)	5.473*** (1.094)	2.774*** (0.999)	2.774*** (0.989)	5.473*** (1.094)	5.398*** (1.126)
NMS_inter	2.382*** (0.807)	1.222 (1.297)	4.420*** (1.032)	4.420*** (1.030)	1.222 (1.297)	1.382 (1.301)
NMS_intra	3.042*** (1.031)	0.238 (1.301)	5.529*** (0.963)	5.529*** (0.955)	0.238 (1.301)	0.410 (1.280)
FE_2000	1.041 (0.646)	0.965 (1.059)	3.215*** (0.743)	3.215*** (0.732)	0.965 (1.059)	1.110 (1.052)
FE_2001	1.322** (0.590)	2.965*** (1.112)	4.297*** (0.671)	4.297*** (0.647)	2.965*** (1.112)	2.964*** (1.088)
FE_2002	1.586*** (0.577)	1.688 (1.140)	6.480*** (0.883)	6.480*** (0.874)	1.688 (1.140)	1.839* (1.131)
FE_2003	0.249 (0.592)	1.112 (1.178)	1.771** (0.724)	1.771** (0.712)	1.112 (1.178)	1.049 (1.159)
FE_2004	0.164 (0.560)	0.764 (1.636)	1.102 (0.741)	1.102 (0.726)	0.764 (1.636)	0.784 (1.650)
FE_2005	0.612 (0.545)	1.844* (1.039)	3.536*** (0.621)	3.536*** (0.604)	1.844* (1.039)	1.908** (1.020)
FE_2006	-0.018 (0.611)	0.339 (0.975)	1.949** (0.814)	1.949** (0.807)	0.339 (0.975)	0.389 (0.991)
FE_2007	0.857 (0.698)	4.549*** (1.237)	6.921*** (0.767)	6.921*** (0.757)	4.549*** (1.237)	4.809*** (1.255)
FE_2008	-0.253 (0.698)	-1.319 (1.088)	-1.274 (0.921)	-1.274 (0.892)	-1.319 (1.088)	-1.269 (1.071)
No. of Obs.	560	114	560	560	114	114

Notes: Importer, exporter, and year fixed effects. Robust standard errors (clustering by country pair).

Signif. levels: 0 '***' 0.01 '**' 0.05 '*' 0.1 ' ' 1

low hog price developments lagged by one year.⁸ That piglet producers can not immediately adjust their production to hog price developments seems obvious. Also the 2007 price bubble is immanent. The time fixed effect of 2008 has even a negative sign. In consequence of the high feed prices in 2007 many piglet producers closed business which decreases supply.

The only striking feature among the different models is the changing sign of the physical distance variable. However this directly leads over to model comparison. Here an overdispersion test indicates overdispersion which then requires a PQL estimation for the multiplicative gravity model. Overdispersion is also indicated by the change in standard errors between PPML and PQL. Also the change here is neglectable, this has not always to be the case. The other comparison is between the 2PM and the G2PM. Here a Likelihood Ratio (LR) Test does not indicate skewness. So for the log-linear gravity model a 2PM estimation is appropriate. The important statistical question is then which model specification is to be favoured. Here a comparison between the significance levels of PQL and 2PM favours PQL. So intra-European piglet trade is best modeled by means of a gravity model in multiplicative form estimated by PQL.

Conclusions

Recently research focus in gravity estimation shifted from an aggregated macro-level to a disaggregated micro-level. Gravity trade model analysis now focuses on single commodity trade flows. Consistent estimation however is as important as ever before. This paper therefore deals with estimation issues of single commodity gravity trade models.

This paper argues that in the presence of excess zeros usual PPML is not appropriate since variance covariance matrices are invalid. Zero-inflated count data models (i.e. ZIPPMML / ZINBPML) are no alternative since under model misspecification these estimators are inconsistent. A reliable alternative is PQL. PQL is consistent even in the presence of model misspecifications and also robust against unobserved heterogeneity [Staub & Winkelmann, 2010]. For log-linear gravity models G2PM is another alternative. G2PM is insofar advantageous as it adjusts for (negative) skewness and regression coefficients still retain usual interpretations as in log-normal models [Chai & Bailey, 2008].

The empirical application done in this paper favours PQL over G2PM. This however is not to be generalized rather research should always follow statistical testing and exclude step by step different alternatives. Overdispersion tests should become standard for multiplicative gravity model analysis as well as LR tests for Two-Part Models. PQL and G2PM should become standard tools for single commodity gravity estimation!

⁸Price charts can be found under <http://www.bordbia.ie/industryservices/pig/pages/prices.aspx>

References

- Anderson, James E. (2010). "The incidence of gravity." *The Gravity Model of International Trade: Advances and Applications*, Brakman, Steven and Bergeijk, eds., Cambridge University Press.
- Anderson, James E. and Wincoop, Eric van (2003). "Gravity with Gravititas: A Solution to the Border Puzzle." *The American Economic Review*, 93, 1, 170-192.
- Burger, Martijn, Oort, Frank van and Linders, Gert-Jan (2009). "On the specification of the Gravity Model of Trade: Zeros, Excess Zeros and Zero-inflated Estimation." *Spatial Economic Analysis*, 4, 2, 167-190.
- Chai, High S. and Bailey, Kent R. (2008). "Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero." *Statistics in Medicine*, 27, 18, 3643-3655.
- Cragg, John G. (1971). "Some Statistical Model for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica*, 39, 5, 829-844.
- Egger, Peter and Larch, Mario (2011). "An assessment of the Europe agreements' effects on bilateral trade, GDP, and welfare." *European Economic Review*, 55, 2, 263-279.
- Felbermayr, Gabriel J. and Kohler, Wilhelm (2010). "Modelling the Extensive Margin of World Trade: New Evidence on GATT and WTO Membership." *The World Economy*, 33, 11, 1430-1469.
- Gourieroux, Christian, Monfort, Alain, and Trognon, Alain (1984a). "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52, 3, 681-700.
- Gourieroux, Christian, Monfort Alain, and Trognon, Alain (1984b). "Pseudo Maximum Likelihood Methods: Applications to Poisson models." *Econometrica*, 52, 3, 701-721.
- Helpman, Elhanan, Melitz, Marc and Rubinstein, Yona (2008). "Estimating Trade Flows: Trading Partners and Trading Volumes." *Quarterly Journal Of Economics*, 123, 2, 441-487.
- Hillberry, Russell H. (2002). "Aggregation Bias, Compositional Change, and the Border Effect." *The Canadian Journal of Economics*, 35, 3, 517-530.
- Marquer, Pol (2010). "Pig farming in the EU, a changing sector." *eurostat Statistics in focus* 8/2010.
- Santos Silva, João M. C. and Tenreyro, Silvana (2006). "The Log of Gravity." *The Review of Economics and Statistics*, 88, 4, 641-658.
- Santos Silva, João M. C., Tenreyro, Silvana and Windmeijer, Frank (2010). "Is it different for zeros? Discriminating between models for non-negative data with many zeros." *CEMMAP Working Paper CWP20/10*.
- Staub, Kevin E. and Winkelmann, Rainer (2010). "Quasi-likelihood estimation of zero-inflated count models." *Working Paper*.