Practical Aspects of Implementing the Multinomial PML Estimator

Sebastian Sotelo

University of Michigan[*]

June 2019

Please click here for most recent version

Abstract

This note expands on the Multinomial Pseudo Maximum Likelihood (MPML) gravity model developed by Eaton, Kortum, and Sotelo (2012) and its relation with Poisson PML, a widely used estimator for gravity equations in the trade literature. It exploits the relation between Multinomial and Poisson random variables to show that (i) MPML and Poisson PML produce the same estimates if the latter includes destination fixed effects; (ii) MPML is easily implemented based on widely available software, and (iii) when using PML methods to estimate gravity models, specifying the dependent variable as shares or as levels amounts to assigning different weights to each importer country. Finally, I illustrate these findings using trade data.

JEL Codes: F14, C13

1 Introduction

Eaton, Kortum, and Sotelo (2012) – EKS, henceforth – develop a model of firm participation in international trade which features a finite number of firms. The model in EKS delivers a gravity-type relation for aggregate bilateral trade *shares*, which they estimate using a Pseudo Maximum Likelihood estimator based on the Multinomial distribution (MPML). This note explains three aspects of the implementation of MPML and illustrates them using trade data. First, I show that MPML and Poisson PML, widely used in the trade literature to model bilateral flows, provide the exact same estimates of the parameters of interest, if the PPML specification includes a full set of destination dummies. This result is important because, in the international trade context, including destination dummies is often the correct specification. To show this equivalence I exploit the relation between the Multinomial and Poisson PDFs. These relation between these two distributions is well known in the literature but, to the best of my knowledge, it has not been used to provide this explicit link between PPML and MPML.[1]

Second, building on this equivalence, I show an easy way to implement the MPML estimator using readily available routines. When used on trade data, this implementation boils down to using the Poisson PML estimator with trade shares as dependent variables and including a full set of destination dummies among the regressors. The MPML estimation routine is usually not available out of the box, while the PPML routine is. Hence, the approach I explain here can be quite useful to practitioners. [2]

Third, I argue that the difference between estimating a gravity model for trade *shares* (as

---

[1]Guimaraes, Figueiredo, and Woodward, 2003 establishes the equivalence between the Poisson likelihood and the likelihood of a conditional logit (up to a constant). In a panel setting, Hausman, Hall, and Griliches, 1984 show that counts over time, for a given observation, are jointly distributed multinomial if each period the counts are distributed Poisson. Cameron and Trivedi, 1998, using the approach I use later, show that the Poisson model with fixed effects does not suffer from an incidental parameters problem. I thank two anonymous referees for pointing me to this literature.

[2]To the best of my knowledge, MPML is not yet available in Stata or R. Poisson PML, however, can be obtained in many ways: either using the poisson and ppml commands in Stata (see Silva and Tenreyro, 2011), or using Generalized Linear Models in Stata or R. Two recent new commands poi2hdfe (Guimaraes, 2016) and ppmlhdfe(Correia, Guimaraes, and Zylkin, 2019), offer substantial spped gains and behave well in the presence of a large number of fixed effects.

in EKS) and estimating it on *levels* (while including fixed effects) can be seen as a difference in how observations are weighted in their contribution to the score vector. In thinking about bilateral trade, it turns out that both estimators compute the contribution of each destination country to the score in the same way; the difference is that a levels specification gives more importance to large countries, while a shares specification treats them all equally. This fact helps explain why the corresponding estimates are different in finite samples, where countries differ substantially in their total expenditures. Finally, using trade data, I illustrate that these different weighting schemes can lead to sizable differences in parameters of interest.

Recently, the quantitative trade literature has devoted much attention to the correct specification of errors in gravity-type regressions and to the performance of PML estimators compared to traditional alternatives. In part, this interest is due to an influential paper by Santos Silva and Tenreyro (2006), who warn against log-linearizing non-linear models to estimate them by OLS, because this approach can produce inconsistent estimates even under relatively mild conditions. They propose the Poisson PML estimator as an alternative with many desirable properties. Santos Silva and Tenreyro (2011) provides further evidence on those properties. Fally (2015) shows that Poisson PML automatically satisfies the adding up constraints imposed by general equilibrium and that, furthermore, with the fixed effects one recovers theory-consistent multilateral further evidence on those properties. Head and Mayer (2014) present a detailed survey of the literature on new estimation techniques for gravity models, including MPML and other work that Santos Silva and Tenreyro spurred. As shown by Gourieroux, Monfort, and Trognon (1984), both PPML and MPML will yield consistent estimates of the first-moment parameters, provided that the conditional expectation is correctly specified. In fact, as I show in this note, they are identical for certain specifications.

2 How to Implement the Multinomial PML Estimator

In this section I present the well-known link between the densities of Multinomial and Poisson random variables. I then show that, a PML estimator based on the Multinomial likelihood will be identical to one based on the Poisson likelihood, if the latter includes a full set of dummies. Next I show that, evaluating with trade shares the first order conditions arising from either likelihood, one can estimate the model in EKS. This equivalence is especially useful if the Poisson routine is readily available, and the MPML routine is not. To avoid programming the latter directly, we instead "trick" a Poisson estimator into doing the job, by adding destination fixed effects and treating trade shares as the dependent variable.

2.1 The Trade Model

In the EKS model, the expected share of imports in destination n from source i, conditional on observables, is

$$\mathbb{E}\left[s_{ni}\right] = \frac{\varphi_{ni}}{\sum_{k=1}^{N} \varphi_{nk}}, \tag{1}$$

for a set of N countries. In equation (1), $s_{ni} = X_{ni}/X_n$, i.e., imports in n from i, divided by total spending in n; and φ_{ni} contains observables and parameters related to the ease with which i exports to n. This expectation considers both the randomness induced by the existence of a finite number of exporters from i selling in n and by the unobservable components of trade costs.

2.2 The Multinomial Likelihood Approach

Consider a vector of observations $\{y_{ni}\}_{i=1,\ldots,N}$, for a given destination n. If this vector were drawn from a multinomial distribution with parameters $\{\varphi_{ni}/\sum_k \varphi_{nk}\}_{i=1,\ldots,N}$, its density would be:

$$f^M\left(\{y_{ni}\}_i ; \{\varphi_{ni}\}_i\right) = \frac{I!}{y_{n1}! \ldots y_{nN}!} \prod_{i=1}^{N} \left(\frac{\varphi_{ni}}{\sum_k \varphi_{nk}}\right)^{y_{ni}} \tag{2}$$

where $\sum_i y_{ni} = I$. The temporary change in notation to y_{ni} highlights that the vector of trade shares s_{ni} cannot be distributed as a multinomial variable, because trade shares are not integers.[3]

The log-likelihood for the complete sample is

$$L^M\left(\{\varphi_{ni}\}_{n,i}\,;\{y_{ni}\}_{n,i}\right) = \sum_{n=1}^{N}\sum_{i=1}^{N} y_{ni}\log\left(\frac{\varphi_{ni}}{\sum_j \varphi_{nj}}\right) + k_1 \tag{3}$$

where k_1 is a constant that does not depend on parameters. Thus one can ignore the constant k_1 when maximizing L^M to estimate the parameters contained in φ_{ni}. Moreover, ignoring k_1 one can evaluate this objective function using shares s_{ni} as a dependent variable, since the objective function no longer restricts the data to be counts.

2.3   The Poisson Likelihood Approach

In practice, we exploit a Poisson (P) likelihood, with destination fixed effects (D_n). Suppose we treat a single observation for n and i, y_{ni}, as drawn from a Poisson distribution with mean $D_n\varphi_{ni}$. The probability of observing y_{ni} is

$$f^P\left(y_{ni};\varphi_{ni},D_n\right) = \frac{(D_n\varphi_{ni})^{y_{ni}}\exp\left(-D_n\varphi_{ni}\right)}{y_{ni}!}.$$

The log-likelihood for the complete sample then:

$$\tilde{L}^P\left(\{\varphi_{ni},D_n\}_{n,i}\,;\{y_{ni}\}_{n,i}\right) = \sum_{n=1}^{N}\sum_{i=1}^{N} y_{ni}\log D_n + \sum_{n=1}^{N}\sum_{i=1}^{N} y_{ni}\log\varphi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N} D_n\varphi_{ni} + k_2,$$

---

[3]As will become apparent later, all we actually need is that the conditional expectation of the model be correctly specified. This is the insight in Gourieroux, Monfort, and Trognon (1984). The appendix includes all derivations.

where k_2 is also a constant. The first order condition of the log-likelihood with respect to the destination fixed effect yields:

$$\hat{D}_n \quad = \quad \frac{1}{\sum_{i=1}^{N} \varphi_{ni}}.$$

Substituting $\hat{D}_n$ back into the log-likelihood, we obtain the concentrated likelihood function:[4]

$$L^P\left(\{\varphi_{ni}\}_{n,i}\,; \{y_{ni}\}_{n,i}\right) \quad = \quad \sum_{n=1}^{N}\sum_{i=1}^{N} y_{ni} \log\left[\frac{\varphi_{ni}}{\sum_{j=1}^{N}\varphi_{nj}}\right] + k_2 \tag{4}$$

A comparison of (3) and (4) makes clear that L^P differs from L^M only by an additive constant, so the likelihoods are proportional. They will therefore yield the same first order conditions for the parameters contained in φ_{ni}. This implies a first important result: the Multinomial PML and the Poisson PML (when the latter includes a full set of fixed effects) produce exactly the same estimator.

2.4   The First Order Conditions

Let $\varphi_{ni} = \exp\left(\beta' z_{ni}\right)$ contain all the parameters β_k and the corresponding explanatory variables $z_{ni,k}$, for $k = 1, \ldots, K$. The first order conditions obtained from either log-likelihood (3) or (4) define the estimates of β:

$$\sum_{n=1}^{N} I_n \left\{ \sum_{i=1}^{N} \left[\frac{y_{ni}}{I_n} - \frac{\hat{\varphi}_{ni}}{\sum_{k=1}^{N}\hat{\varphi}_{nk}}\right] z_{ni} \right\} = 0, \tag{5}$$

where, recall, $I_n \equiv \sum_{i=1}^{N} y_{ni}$, and $\hat{\varphi}_{ni} = \exp\left(\hat{\beta}' z_{ni}\right)$. Again, the fact that both approaches produce the same first order conditions reveals that the estimators coincide.

---

[4]This use of the concentrated likelihood to relate the Poisson and the Multinomial likelihood is the same as in Cameron and Trivedi (1998), Ch. 9.

2.5   Implementation in EKS

Let us now relate these first order conditions to the EKS model. Evaluating the first order conditions (5) at the trade shares, i.e. setting $y_{ni} = s_{ni}$ and $I_n = 1$, we obtain:

$$\sum_{n=1}^{N} \sum_{i=1}^{N} \left[s_{ni} - \frac{\hat{\varphi}_{ni}}{\sum_{k=1}^{N} \hat{\varphi}_{nk}} \right] z_{ni,k} \;\; = \;\; 0. \tag{6}$$

Inspection of equation (6) reminds us that, if the expectation given by equation (1) is correctly specified, one can use these first order conditions – evaluated at s_{ni} – to obtain consistent estimates for the parameters in φ.

In practice, EKS estimate β using a Poisson estimator, with a full set of destination dummies, using trade *shares* as dependent variables, because Poisson PML routines are readily available. But it should be clear at this point that doing so is equivalent to applying the MPML estimator to trade shares data. Further, since the data satisfies the normalization $\sum_i s_{ni} = 1$, one can think of the destination dummies D_n as ensuring that the import shares of country n add up to one. The adding-up property of the Poisson estimator with fixed effects has been studied by Fally (2015).

3   MPML and PPML as Different Weighting Schemes

In this section, we take a step back and study a general PPML estimation problem in the presence of destination fixed effects. Doing so shows that the difference between modeling shares and levels can be understood as a difference in weighting schemes.

Suppose now that we are interested in a random variable variable $X_{ni} \geq 0$, which we model as

$$\mathbb{E}\left[X_{ni} | z_{ni}; \gamma\right] = D_n \phi_{ni}, \tag{7}$$

where $\phi_{ni} = \exp\left(\gamma' z_{ni}\right)$. For concreteness, suppose that X_{ni} is the imports of destination n from source i. Note that we have included destination fixed effects, D_n. Whether they

should be included depends on the model being estimated, but destination fixed effects arise naturally in many trade models that deliver a structural gravity equation, capturing the multilateral resistance terms (in the case of the destination, interpretable as a price index. See, e.g. Head and Mayer, 2014).

There are is a tight link between this model and that of EKS. First, note that equation (7) specializes to equation (1), with $D_n = 1/\sum_k \phi_{nk}$ and s_{ni} as a dependent variable. Second, in the EKS model, by conditioning on total spending in n (as well as all other observables), one could also compute an expectation for trade *flows* that has the form of equation (7), with $D_n = X_n/\sum_k \phi_{nk}$.[5]

To better understand the difference between specifying a model in shares and specifying it in levels, suppose we estimate the model in equation (7) via PPML. The resulting first order conditions, obtained by evaluating (5) at $y_{ni} = X_{ni}$, define $\hat{\gamma}$ and are given by[6]

$$\sum_n X_n \left\{ \sum_i \left[s_{ni} - \frac{\hat{\phi}_{ni}}{\sum_k \hat{\phi}_{nk}} \right] z_{ni,k} \right\} = 0, \tag{8}$$

where $\hat{\phi}_{ni} = \exp(\gamma' z_{ni})$.where

By comparing (6) and (8), the difference between the two estimators becomes clear. The contribution of each importer's observations to the score is computed in the same way in both estimators (the expression in curly brackets in equation 8). The difference is in how those contributions are weighted in the final calculation of the score. When applying PPML to trade flows, X_{ni}, the weight is the sum of the dependent variable for that importer, X_n, whereas it is equal to one in the MPML approach. In other words, specifying the equation on trade flows induces the estimator to try harder to fit the data for countries with larger total spending, X_n, whereas specifying it on trade shares treats all countries equally.

---

[5]This statement clarifies footnote 24 in EKS, by showing that, in fact, their model also produces a gravity relation for trade levels.

[6]To obtain these conditions we follow the concentrated log-likelihood approach as before.

4   Application to Trade Data

In this section we draw the implications from our previous results for common empirical specifications in trade, and provide an example with trade data. In common empirical gravity specifications, an additional implication follows. If ϕ_{ni} includes a destination-specific trade impediment – i.e., a shifter that uniformly affects destination n's costs of importing from other countries – the MPML model will fit the home trade share perfectly.[7] This form of trade impediment is implied in Eaton and Kortum's (2002) empirical gravity formulation. The reason is that, as shown in the Appendix, such an impediment will also make the model fit perfectly the partial sum of purchases, excluding home purchases, $\sum_{i \neq n} X_{ni}$. In the presence of destination dummies D_n, which ensure that the model fits $\sum_i X_{ni}$ perfectly, the implication is that the model fits the home purchases, X_{nn}, perfectly. In this case, as gleaned from equation (6), the home observation does not contribute to the identification of bilateral trade barriers.

Next, consider a model that includes both destination and source dummies, D_n and S_i, as in

$$\mathbb{E}\left[X_{ni} | z_{ni}; \gamma\right] = D_n S_i \phi_{ni}.$$

In this setup, including a destination-specific import cost or including a source-specific export cost[8] will yield numerically identical estimates for the coefficients on bilateral covariates. Both specifications will also yield the exact same predicted values for all observations. The reason is that, in both specifications, the model will fit exactly: (i) total sales *and* purchases of each country, and (ii) the home sales observation. Hence, both formulations produce equivalent first order conditions, for the identification of coefficients on bilateral impediments. Appendix B provides details on this equivalence. However, as shown by Waugh (2010), the formulation that includes source-specific exporting costs is preferable because it better accounts for differences in the costs of living between rich and poor countries.

---

[7]In the formulation above, if $\phi_{ni} = m_n \exp\left(\gamma' z_{ni}\right)$ if $n \neq i$ and $\phi_{ni} = \exp\left(\gamma' z_{ni}\right)$ if $n = i$.

[8]The exporter-cost formulation is given by: $\phi_{ni} = k_i \exp\left(\gamma' z_{ni}\right)$ if $n \neq i$ and $\phi_{ni} = k_i \exp\left(\gamma' z_{ni}\right)$ if $n = i$.

Table 1: Comparison of Estimators

	(1) Levels	(2) Shares	(3) Shares w/ source barrier
Distance	-0.562***	-0.735***	-1.072***
	(-9.45)	(-16.74)	(-20.99)
Lack of contiguity	-1.042***	-0.736***	-0.370**
	(-4.44)	(-4.55)	(-2.72)
lack of common language	-0.114	-0.300**	-0.511***
	(-0.87)	(-3.13)	(-4.82)
Lack of common legal origin	-0.0626	-0.222**	-0.133
	(-0.56)	(-3.16)	(-1.85)
Lack of common colonizer	1.389***	1.688***	-0.306
	(3.52)	(5.65)	(-1.50)
Lack of colonial ties	-0.331*	-0.861***	-0.953***
	(-2.57)	(-6.06)	(-6.84)
N	8464	8464	8464

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1 illustrates these findings, using the same data as EKS. All specifications contain source and destination fixed effects, and variables are defined to ensure that trade of a country with itself is unimpeded.[9] Column (1) presents the coefficients for some usual gravity variables, estimated on trade flow levels. Next, Column (2) estimates the model on trade shares. Focusing on the first row, we see the effect of scaling down large countries is to increase the distance elasticity, by close to 0.2. Column (3) repeats the shares specification, this time including source-specific trade barriers, which further increases this elasticity to close to 1.[10]

5   Conclusion

Provided that the model specification of the conditional expectation is correct, specifying a gravity equation in levels or in shares will both yield consistent estimates. But the MPML estimator in EKS focuses on trade shares, not levels. Relative to a levels specification, it down-weights the observations belonging to destinations with large overall purchases, reducing their influence in the estimation. This fact helps explain why, when applied on the same sample in EKS, levels and shares specifications yield quite different results, for example, for the elasticity of trade with respect to distance.

References

CAMERON, A. C., AND P. K. TRIVEDI (1998): *Regression analysis of count data.* Cambridge University Press.

---

[9]Specifically, in the regression the log of distance is normalized to zero. Other indicator variables take a value of one when two countries lack a common characteristic expected to facilitate trade, e.g. when they lack a common language. This way, the own trade barrier is always equal to one.

[10]As explained before, a specification that includes destination-specific trade barriers will produce the same coefficients.

CORREIA, S., P. GUIMARAES, AND T. ZYLKIN (2019): "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," .

EATON, J., AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica*, 70(5), 1741–1779.

EATON, J., S. S. KORTUM, AND S. SOTELO (2012): "International Trade: Linking Micro and Macro," NBER Working Papers 17864, National Bureau of Economic Research, Inc.

FALLY, T. (2015): "Structural gravity and fixed effects," *Journal of International Economics*, 97(1), 76–85.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 52(3), pp. 681–700.

GUIMARAES, P. (2016): "POI2HDFE: Stata module to estimate a Poisson regression with two high-dimensional fixed effects," .

GUIMARAES, P., O. FIGUEIREDO, AND D. WOODWARD (2003): "A Tractable Approach to the Firm Location Decision Problem," *The Review of Economics and Statistics*, 85(1), 201–204.

HAUSMAN, J., B. H. HALL, AND Z. GRILICHES (1984): "Econometric Models for Count Data with an Application to the Patents-R & D Relationship," *Econometrica*, 52(4), 909–938.

HEAD, K., AND T. MAYER (2014): *Gravity Equations: Workhorse, Toolkit, and Cookbook*vol. 4 of *Handbook of International Economics*, chap. 3, pp. 131–195. Elsevier.

SANTOS SILVA, J., AND S. TENREYRO (2011): "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters*, 112(2), 220–222.

SANTOS SILVA, J. M. C., AND S. TENREYRO (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88(4), 641–658.

SILVA, J. M. C. S., AND S. TENREYRO (2011): "poisson: Some convergence issues," *Stata Journal*, 11(2), 207–212(6).

WAUGH, M. E. (2010): "International Trade and Income Differences," *American Economic Review*, 100(5), 2093–2124.

A Appendix

A.1 Obtaining the Multinomial Log-Likelihood

In general, a multinomial r.v. counts the number of successes in each of N categories, after I independent trials. If p_n is the probability of success and y_n is the number of successes in category n, the probability of observing $\{y_n\}_{n=1}^N$ is $\frac{I!}{y_1!\ldots y_N!}p_1^{y_1}\ldots p_N^{y_N}$, where $\sum_n p_n = 1$ and $\sum_n y_n = I_n$.

If we treated all import shares for destination n as one observation, distributed like a Multinomial (M) r.v., the probability of observing the counts $\{y_{ni}\}_i$ is:

$$f_n^M\left(\{y_{ni}\}_i ; p_n\right) = \frac{I_n!}{y_{n1}!\ldots y_{nN}!}\prod_{i=1}^N p_{ni}^{y_{ni}}.$$

The likelihood for our complete sample, $\{y_{ni}\}_{n,i}$, is

$$f^M\left(\{y_{ni}\}_{n,i} ; p\right) = \prod_{n=1}^N f_n^M\left(\{y_{ni}\}_i ; p_n\right) = \prod_{n=1}^N \frac{I_n!}{y_{n1}!\ldots y_{nN}!}\prod_{i=1}^N p_{ni}^{y_{ni}}.$$

Finally, the log-likelihood is

$$L^M\left(\{y_{ni}\}_{n,i} ; \pi\right) = \sum_{n=1}^N \sum_{i=1}^N y_{ni}\log p_{ni} + k_1,$$

where k_1 is a constant that does not depend on p. Letting $p_{ni} = \varphi_{ni}/\sum_j \varphi_{nj}$ we obtain the log-likelihood in the main text.

A.2 Obtaining the Poisson Log-Likelihood

The likelihood of one count y_{ni} will be

$$f_{ni}^P = \frac{(D_n\varphi_{ni})^{y_{ni}}\exp(-D_n\varphi_{ni})}{y_{ni}!}.$$

The likelihood for the complete sample is

$$f^P = \prod_{n=1}^{N} \prod_{i=1}^{N} \frac{(D_n \varphi_{ni})^{y_{ni}} \exp(-D_n \varphi_{ni})}{y_{ni}!}$$

and the corresponding log-likelihood is

$$L^P = \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log D_n + \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log \varphi_{ni} - \sum_{n=1}^{N} \sum_{i=1}^{N} D_n \varphi_{ni} + \tilde{k}_2,$$

where $\tilde{k}_2$ does not depend on D or φ.

The first order condition of the log-likelihood with respect to the destination fixed effect is

$$\frac{\partial L^P}{\partial D_n} = \sum_{i=1}^{N} \frac{y_{ni}}{D_n} - \sum_{i=1}^{N} \varphi_{ni} = 0 \Rightarrow \sum_{i=1}^{N} y_{ni} = D_n \sum_{i=1}^{N} \varphi_{ni} \Rightarrow \hat{D}_n = \frac{\sum_{i=1}^{N} y_{ni}}{\sum_{i=1}^{N} \varphi_{ni}}.$$

Letting $Y_n \equiv \sum_i y_{ni}$ and substituting $\hat{D}_n$ back into the log-likelihood, we get the concentrated likelihood function:

$$
\begin{aligned}
L^P &= \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log \left(\frac{Y_n}{\sum_{j=1}^{N} \varphi_{nj}} \right) + \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log \varphi_{ni} - \sum_{n=1}^{N} \sum_{i=1}^{N} \left(\frac{Y_n}{\sum_{j=1}^{N} \varphi_{nj}} \right) \varphi_{ni} + \tilde{k}_2 \\
&= \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log \left(\frac{\varphi_{ni}}{\sum_{j=1}^{N} \varphi_{nj}} \right) + - \sum_{n=1}^{N} Y_n \left(\frac{\sum_{i=1}^{N} \varphi_{ni}}{\sum_{j=1}^{N} \varphi_{nj}} \right) + \tilde{k}_2 \\
&= \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log \left(\frac{\varphi_{ni}}{\sum_{j=1}^{N} \varphi_{nj}} \right) + k_2,
\end{aligned}
$$

where k_2 is a constant again. This is the expression in the main text.

A.3 Derivation of the First Order Conditions

Letting $\varphi_{ni} = \exp(\beta' z_{ni})$, β contain all the parameters we estimate and z the corresponding explanatory variables, the first order conditions that define the estimates of β are

$$
\begin{aligned}
\frac{\partial L^M}{\partial \beta} &= \frac{\partial}{\partial \beta} \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} \log \frac{\varphi_{ni}}{\sum_{k=1}^{N} \varphi_{nk}} = 0 \\
&\Leftrightarrow \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} z_{ni} - \sum_{n=1}^{N} \sum_{i=1}^{N} \frac{y_{ni}}{\sum_{k=1}^{N} \varphi_{nk}} \sum_{l=1}^{N} \varphi_{nl} z_{nl} = 0 \\
&\Leftrightarrow \sum_{n=1}^{N} \sum_{i=1}^{N} y_{ni} z_{ni} - \sum_{n=1}^{N} \frac{\sum_{l=1}^{N} \varphi_{nl} z_{nl}}{\sum_{k=1}^{N} \varphi_{nk}} Y_n = 0 \\
&\Leftrightarrow \sum_{n=1}^{N} \left\{ \sum_{i=1}^{N} y_{ni} z_{ni} - Y_n \frac{\sum_{i=1}^{N} \varphi_{ni} z_{ni}}{\sum_{k=1}^{N} \varphi_{nk}} \right\} = 0 \\
&\Leftrightarrow \sum_{n=1}^{N} Y_n \left\{ \sum_{i=1}^{N} \left[\frac{y_{ni}}{Y_n} - \frac{\varphi_{ni}}{\sum_{k=1}^{N} \varphi_{nk}} \right] z_{ni} \right\} = 0.
\end{aligned}
$$

If we evaluate these first order conditions at $y_{ni} = s_{ni}$, we obtain the first order conditions in the main text

$$
\sum_{n=1}^{N} \sum_{i=1}^{N} \left[s_{ni} - \frac{\varphi_{ni}}{\sum_{k=1}^{N} \varphi_{nk}} \right] z_{ni} = 0,
$$

since $\sum_i s_{ni} = 1$.[11]

A.4 Proof that MPML fits the home trade share exactly

Without loss of generality, consider the following simple model, without bilateral covariates

$$
\mathbb{E}[s_{ni}] = \frac{\varphi_{ni}}{\sum_k \varphi_{nk}}
$$

where $\varphi_{ni} = S_i k_n$ if $i \neq n$ and $\varphi_{ni} = S_i$ if $i = n$.

---

[11] Note that, because $\sum_i s_{ni} = 1$, we can rewrite the last set of equations as

$$
\sum_{n=1}^{N} \sum_{i=1}^{N} \left[s_{ni} - \frac{\bar{s}_n}{\bar{\varphi}_n} \varphi_{ni} \right] z_{ni} = 0.
$$

This way of writing the equations that define the estimator coincide exactly with those derived by Cameron and Trivedi (1998).

The first order conditions from the Poisson likelihood $\tilde{L}^P$ are

1. With respect to D_n

$$\frac{\partial \tilde{L}^P}{\partial D_n} = \sum_i s_{ni} \frac{1}{D_n} - \sum_i \varphi_{ni} = 0$$

$$\Rightarrow D_n = \frac{1}{\sum_k \varphi_{nk}}.$$

2. With respect to k_n

$$\frac{\partial \tilde{L}^P}{\partial k_n} = \sum_{i \neq n} s_{ni} \frac{1}{k_n} - \sum_{i \neq n} D_n S_i = 0$$

$$\Rightarrow \sum_{i \neq n} s_{ni} \frac{1}{k_n} = \sum_{i \neq n} D_n S_i$$

$$\frac{\sum_{i \neq n} s_{ni}}{\sum_{i \neq n} S_i} = k_n D_n.$$

Using the previous first order condition

$$\sum_{i \neq n} s_{ni} = \frac{\sum_{i \neq n} \varphi_{ni}}{\sum_k \varphi_{nk}},$$

and in our case with $\sum_i s_{ni} = 1$, this implies

$$s_{nn} = \frac{\varphi_{nn}}{\sum_k \varphi_{nk}}.$$

A.5   Derivations for the general Poisson approach

Suppose now that we are interested in a general model for variable $X_{ni} \geq 0$, whose conditional expectation is given by $D_n \phi_{ni}$, with $\phi_{ni} = \exp(\beta' z_{ni})$.

The likelihood for the complete sample is

$$f^P\left(\{X_{ni}\}_{n,i}; D, \phi\right) = \prod_{n=1}^N \prod_{i=1}^N \frac{(D_n \phi_{ni})^{X_{ni}} \exp(-D_n \phi_{ni})}{X_{ni}!}.$$

We start by eliminating D_n from the log-likelihood, which requires setting $\frac{\partial L^P}{\partial D_n} = 0$ for each n.

$$
\begin{aligned}
L^P &= \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\log D_n + \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\log\phi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N} D_n\phi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right) \Rightarrow \\
\frac{\partial L^P}{\partial D_n} &= \sum_{i=1}^{N}\frac{X_{ni}}{D_n} - \sum_{i=1}^{N}\phi_{ni} = 0 \Rightarrow \sum_{i=1}^{N} X_{ni} = D_n\sum_{i=1}^{N}\phi_{ni} \Rightarrow \hat{D}_n = \frac{X_n}{\sum_{i=1}^{N}\phi_{ni}},
\end{aligned}
$$

where we define $X_n = \sum_i X_{ni}$. Substituting $\hat{D}_n$ back into the likelihood, we get the concentrated likelihood function:

$$
\begin{aligned}
L^P &= \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\log D_n + \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\log\phi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N} D_n\phi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right) \\
&= \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\log\frac{X_n}{\sum_{i'=1}^{N}\phi_{ni'}} + \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\log\phi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N}\frac{X_n}{\sum_{i'=1}^{N}\phi_{ni'}}\phi_{ni} - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right) \\
&= \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\left[\log\frac{X_n}{\sum_{i'=1}^{N}\phi_{ni'}} + \log\varphi_{ni}\right] - \sum_{n=1}^{N} X_n\frac{\sum_{i=1}^{N}\varphi_{ni}}{\sum_{i'=1}^{N}\varphi_{ni'}} - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right) \\
&= \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\left[\log\frac{X_n\phi_{ni}}{\sum_{i'=1}^{N}\phi_{ni'}}\right] - \sum_{n=1}^{N} X_n - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right) \\
&= \sum_{n=1}^{N}\sum_{i=1}^{N} X_{ni}\left[\log\frac{\phi_{ni}}{\sum_{i'=1}^{N}\phi_{ni'}}\right] + \sum_{n=1}^{N} X_n\log X_n - \sum_{n=1}^{N} X_n - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right)
\end{aligned}
$$

We can simplify this expression by using $s_{ni}X_n = X_{ni}$:

$$
\begin{aligned}
L^P &= \sum_{n=1}^{N}\sum_{i=1}^{N} s_{ni}X_n\left[\log\frac{\phi_{ni}}{\sum_{i'=1}^{N}\phi_{ni'}}\right] + \sum_{n=1}^{N} X_n\log X_n - \sum_{n=1}^{N} X_n - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right) \\
&= \sum_{n=1}^{N} X_n\sum_{i=1}^{N} s_{ni}\left[\log\frac{\phi_{ni}}{\sum_{i'=1}^{N}\phi_{ni'}}\right] + \sum_{n=1}^{N} X_n\log X_n - \sum_{n=1}^{N} X_n - \sum_{n=1}^{N}\sum_{i=1}^{N}\log\left(X_{ni}!\right).
\end{aligned}
$$

Taking the derivative with respect to parameters, we obtain the equations that define the

estimator

$$
\begin{aligned}
\frac{\partial L^P}{\partial \beta} &= \sum_n \sum_i s_{ni} X_n z_{ni} - \sum_n \sum_i \frac{X_n s_{ni}}{\sum_k \phi_{nk}} \sum_l \phi_{nl} z_{nl} \\
&= \sum_n X_n \sum_i s_{ni} z_{ni} - \sum_n X_n \sum_l \phi_{nl} z_{nl} \sum_i \frac{s_{ni}}{\sum_k \phi_{nk}} \\
&= \sum_n X_n \sum_i s_{ni} z_{ni} - \sum_n X_n \frac{\sum_i \phi_{ni} z_{ni}}{\sum_k \phi_{nk}} \\
&= \sum_n X_n \left\{ \sum_i \left[s_{ni} - \frac{\phi_{ni}}{\sum_k \phi_{nk}} \right] z_{ni} \right\} = 0
\end{aligned}
$$

B Equivalence between the exporter- and importer-barrier formulations

Suppose now that we are interested in a model for $X_{ni} \geq 0$, whose conditional expectation is given by $D_n S_i \phi_{ni}$. The likelihood for the complete sample is

$$
f^P \left(\{X_{ni}\}_{n,i} ; D, S, \phi \right) = \prod_{n=1}^N \prod_{i=1}^N \frac{(D_n S_i \phi_{ni})^{X_{ni}} \exp\left(-D_n S_i \phi_{ni}\right)}{X_{ni}!}.
$$

The log-likelihood is

$$
\begin{aligned}
\tilde{L}^P &= \sum_{n=1}^N \sum_{i=1}^N X_{ni} \log D_n + \sum_{n=1}^N \sum_{i=1}^N X_{ni} \log S_i + \sum_{n=1}^N \sum_{i=1}^N X_{ni} \log \phi_{ni} \\
&- \sum_{n=1}^N \sum_{i=1}^N D_n S_i \phi_{ni} - \sum_{n=1}^N \sum_{i=1}^N \log\left(X_{ni}!\right).
\end{aligned}
$$

We start by finding the first order conditions associated with D_n and S_i $\forall i, n$:

$$
\begin{aligned}
\frac{\partial \tilde{L}^P}{\partial D_n} &= \sum_{i=1}^N \frac{X_{ni}}{D_n} - \sum_{i=1}^N S_i \phi_{ni} = 0 \\
&\Rightarrow \hat{D}_n = \frac{X_n}{\sum_{i=1}^N S_i \phi_{ni}}
\end{aligned}
$$

and

$$\frac{\partial \tilde{L}^P}{\partial S_i} = \sum_{n=1}^{N} \frac{X_{ni}}{S_i} - \sum_{n=1}^{N} D_n \phi_{ni} = 0$$
$$\Rightarrow \hat{S}_i = \frac{Y_i}{\sum_{n=1}^{N} D_n \phi_{ni}}.$$

where we define $Y_i \equiv \sum_{n=1}^{N} X_{ni}$. Alternatively, these two first order conditions mean that

$$\sum_{i=1}^{N} D_n S_i \phi_{ni} = X_n \tag{9}$$

and

$$\sum_{n=1}^{N} D_n S_i \phi_{ni} = Y_i \tag{10}$$

so the model fits total purchases and total sales of country i perfectly.

B.1 Symmetry between k_n and m_i formulations

Case 1 Suppose that $\phi_{ni} = m_i exp\left(\beta' z_{ni}\right)$ if $n \neq i$ and $\phi_{ni} = \exp\left(\beta' z_{ni}\right)$ if $n = i$. In that case, the first order conditions are

$$\frac{\partial \tilde{L}^P}{\partial m_j} = \sum_{n \neq j}^{N} X_{nj}/m_j - \sum_{n \neq j}^{N} \left(D_n S_j\right) \frac{\phi_{nj}}{m_j} = 0$$
$$\Rightarrow \sum_{n \neq j}^{N} X_{nj} = \sum_{n \neq j}^{N} D_n S_j \phi_{nj} \tag{11}$$

Using the fact that the total sales are fit perfectly with equation (11) yields

$$X_{ii} = \phi_{ii} D_i S_i \tag{12}$$

so the home observation is fit perfectly. However, having included destination fixed effects, total purchases of each country are fit exactly, and so we obtain, for each n

$$\sum_{i \neq n} X_{ni} = \sum_{i \neq n} S_i D_n \phi_{ni}, \tag{13}$$

which is the set of equations we would obtain if we included destination specific import costs (see Case 2 below).

Two further implications are that the home trade share in purchases and in sales are fit perfectly.

Case 2 Suppose that $\phi_{ni} = k_n exp\left(\beta' z_{ni}\right)$ if $n \neq i$ and $\phi_{ni} = \exp\left(\beta' z_{ni}\right)$ if $n = i$. By analogy, we the first order conditions with respect to k_n imply that

$$\sum_{i \neq n} X_{ni} = \sum_{i \neq n} D_n S_i \phi_{nj}.$$

Using the fact that the model fits purchases perfectly, we obtain again that

$$X_{nn} = \phi_{nn} D_n S_n. \tag{14}$$

Furthermore, since the model also fits sales perfectly, using this last equation we obtain for each i

$$\sum_{n \neq i} X_{ni} = \sum_{n \neq i} D_n S_i \phi_{ni},$$

which is the set of equations that we would obtain with source specific export costs (see equation 11, above)

Cases 1 and 2 yield the same β Now consider again the first order condition for β for Case 1:

$$\sum_{n=1}^{N}\sum_{i=1}^{N}\left[X_{ni} - D_n S_i \phi_{ni}\right] z_{ni,k} = 0,$$

for each k, which after using the fact that the home sales are fit perfectly, becomes:

$$\sum_{i=1}^{N}\sum_{n\neq i}\left[X_{ni} - D_n Q_i \exp\left(\beta' z_{ni}\right)\right] z_{ni,k} = 0.$$

where $Q_i \equiv S_i m_i$, because after eliminating the observations for which $n \neq i$, the previous equation cannot separately identify S_i and m_i, which work as an exporter shifter. The rest of the equations in the system are obtained by rewriting (11) and (13)

$$\sum_{n\neq i}^{N} X_{ni} = \sum_{n\neq i}^{N} D_n Q_i \exp\left(\beta' z_{ni}\right)$$

$$\sum_{i\neq n} X_{ni} = \sum_{i\neq n} D_n Q_i \exp\left(\beta' z_{ni}\right)$$

together with (12),

$$X_{nn} = \exp\left(\beta' z_{nn}\right) D_n Q_n / m_n,$$

which provides a system for β, D, Q, m.

The first order conditions for β in Case 2 are

$$\sum_{n=1}^{N}\sum_{i=1}^{N}\left[X_{ni} - D_n S_i \phi_{ni}\right] z_{ni,k} = 0,$$

which after eliminating the home observation can be written as

$$\sum_{i=1}^{N}\sum_{n\neq i}\left[X_{ni} - W_n S_i \exp\left(\beta' z_{ni}\right)\right] z_{ni,k} = 0$$

with $W_n \equiv D_n m_n$. Next, we rewrite (11) and (13) as

$$\sum_{n \neq i}^{N} X_{ni} = \sum_{n \neq i}^{N} W_n S_i \exp\left(\beta' z_{ni}\right)$$

$$\sum_{i \neq n} X_{ni} = \sum_{i \neq n} W_n S_i \exp\left(\beta' z_{ni}\right)$$

with (12)

$$X_{nn} = \exp\left(\beta' z_{nn}\right) W_n S_n / m_n$$

which is identical to the system of equations above, for β, W, S, m.

This does not mean, however, that the underlying S_n and D_n will be the same in the two specifications.