Trade, Gravity and Aggregation[1]

Holger Breinlich[2] Dennis Novy[3] J.M.C. Santos Silva[4]
University of Surrey University of Warwick University of Surrey

11 April 2022

Abstract

Gravity equations are an important tool in empirical international trade research. We study to what extent sector-level parameters can be recovered from aggregate gravity equations estimated via Poisson pseudo maximum likelihood. We show that in the leading case where trade cost regressors do not vary at the sector level, estimates obtained with aggregate data have a clear interpretation as a weighted average of sectoral elasticities. Otherwise the estimates are biased but researchers may possibly infer the direction of the bias. We illustrate our results by revisiting Baier and Bergstrand's (2007) influential study of the effects of free trade agreements.

JEL classification: C23, C43, F14, F15, F17

Keywords: Free trade agreements, gravity equation, heterogeneity, PPML, trade costs

[1]We thank the editor Daniel Xu and four anonymous referees for many helpful and interesting comments. We gratefully acknowledge research support from the Economic and Social Research Council (ESRC grant ES/P00766X/1). We also thank Thomas Zylkin and conference and seminar participants at the European Trade Study Group conference 2019 and the London School of Economics for helpful comments. The usual disclaimer applies.

[2]Holger Breinlich, School of Economics, University of Surrey, Guildford GU2 7XH, UK, CEP/LSE and CEPR. Email: h.breinlich@surrey.ac.uk.

[3]Dennis Novy, Department of Economics, University of Warwick, Coventry CV4 7AL, UK, CEP/LSE, CEPR and CESifo. Email: d.novy@warwick.ac.uk.

[4]João Santos Silva, School of Economics, University of Surrey, Guildford GU2 7XH, UK. Email: jmcss@surrey.ac.uk.

1 Introduction

The gravity equation is the workhorse model in international trade for estimating trade cost parameters and evaluating the effects of policy changes. Gravity equations have been used to estimate the trade effects of free trade agreements (FTAs), currency unions, WTO membership and colonial history, amongst other institutional features (see Anderson, 2011, and Head and Mayer, 2014). When trade costs change, the impact typically materializes at the level of firms and consumers in a particular sector. That is, the impact is governed by parameters at the sector level, for example sector-level demand elasticities. However, due to data constraints, gravity equations are routinely estimated at the aggregate level using country-level data. The resulting estimates are often assumed, explicitly or implicitly, to be informative about the more fundamental sector-level parameters.

In this paper, we investigate to what extent this practice is justified. Specifically, we ask whether we can identify sector-level elasticities from aggregate gravity regressions and, if yes, under what conditions?[5]

We show that if trade elasticities and regressors (i.e., trade cost variables and fixed effects) are the same at the sector level, we can recover the sector-level elasticity from estimation based on aggregate data when the model is estimated by Poisson pseudo maximum likelihood (PPML), as recommended by Santos Silva and Tenreyro (2006). For instance, this scenario applies when the trade cost variables are bilateral distance and a common language dummy whose elasticities do not differ at the sector level, and the fixed effects included in the regressions do not vary at the sector level either.

When trade elasticities and fixed effects vary at the sectoral level but trade cost variables do not, naturally it is not possible to recover the sectoral elasticities from aggregate

[5] Although we only explicitly consider aggregation across sectors, aggregation over time raises similar issues and our results can easily be applied to that problem.

data. However, we show that in this case PPML estimation using aggregate data will approximately recover a trade-weighted average of sector-level elasticities.

If the trade cost variables (e.g., tariffs) vary at the sector level, estimates obtained with aggregate data yield biased estimates of the parameters of interest, but we show that it may be possible to determine the direction of the bias when we have information about the distribution of the sector-level regressors.

Our work contributes to several strands of the literature. First, we contribute to the econometrics literature on cross-sectional aggregation of constant-elasticity models. However, rather than considering the consequences of aggregation in the context of log-linearized models estimated by OLS, as Lewbel (1992) and van Garderen, Lee and Pesaran (2000), we study the effects of aggregation when constant-elasticity models are estimated in their exponential form by PPML.

Second, we contribute to the literature that studies the consequences of aggregation in the particular context of gravity equations. Specifically, our work is related to that of Imbs and Mejean (2015) who exploit tariff variation and show that trade elasticity estimates based on aggregate data are smaller in absolute value than the average of sectoral elasticities. This finding is driven by the fact that tariffs tend to be less dispersed in elastic sectors. The main difference to Imbs and Mejean (2015) is that we focus on aggregation with PPML, not OLS, and we also consider cases where regressors are common across sectors and only the parameters vary, and vice versa.[6] Our results imply that aggregation can either reduce or increase estimates depending on the sectoral distribution of trade cost variables.

[6] In particular, we show that even if trade cost parameters are common across sectors, aggregation can shift coefficient estimates (due to fixed effects with a sector dimension). In contrast, their argument only applies if sectoral elasticities are heterogeneous (see Imbs and Mejean, 2015, p. 47).

In a related contribution, Redding and Weinstein (2019a and 2019b) show that the theoretical aggregation of gravity equations is not straightforward. They estimate the distance elasticity at the aggregate level and devise a decomposition that breaks down the estimated aggregate distance elasticity into various components. One of these components is a Jensen's inequality term that accounts for the fact that the typical gravity equation is log-linear while aggregation implies summation in levels. As Redding and Weinstein rely on OLS estimation for their decomposition, they do not account for the bias resulting from heteroskedasticity as highlighted by Santos Silva and Tenreyro (2006). In addition, as the Jensen's inequality term has to be computed from estimates of sectoral gravity equations, a researcher who only has access to aggregate data cannot resort to the decomposition method as a way of adjusting for the effects of aggregation. This differs from our intention of understanding the effects of aggregation in situations where sectoral data may not be available.

In independent work, French (2019) also considers the effects of aggregation on gravity equations estimated by PPML. We use a different analytical framework that provides an interpretation of aggregation effects that is arguably more practical and intuitive. In particular, we show that in a leading case, aggregate PPML estimates of gravity coefficients can be seen as weighted averages of sector-specific parameters, thereby providing a clear link between the estimated parameters and the parameters of interest.

The paper is structured as follows. In Section 2 we present a simple international trade model that delivers gravity equations at two levels of aggregation. This framework provides theoretical guidance for our approach and helps to clarify the link between parameter estimates and the underlying theoretical parameters common in international trade models. In Section 3 we present initial motivating evidence regarding the effects of aggregation in gravity estimation. In Section 4 we explain these findings by deriving theoretical results on the aggregation of constant-elasticity models under different assumptions. Section 5 concludes. We present additional results in an online appendix.

2 Gravity at different levels of aggregation

We sketch a theoretical framework that yields gravity equations at different levels of aggregation. It is based on a simple model of international trade with a two-tier nested constant elasticity of substitution (CES) demand system, following Redding and Weinstein (2019a and 2019b). The upper tier represents the aggregate level of the economy, and the lower tier the disaggregated (sector) level. Varieties in each sector are differentiated by origin according to the Armington assumption.

Aggregate consumption at the upper tier is given by

$$C_j = \left(\sum_s (c_{js})^{\frac{\nu-1}{\nu}} \right)^{\frac{\nu}{\nu-1}},$$

where c_{js} is real consumption by country j of sector s aggregates, and ν is the elasticity of substitution between sectors. The lower-tier aggregator is given by

$$c_{js} = \left(\sum_i (\theta_{ijs} c_{ijs})^{\frac{\sigma_s-1}{\sigma_s}} \right)^{\frac{\sigma_s}{\sigma_s-1}},$$

where c_{ijs} is real consumption by country j of sector s varieties originating from country i, σ_s is the elasticity of substitution across sector s varieties, and $\theta_{ijs} \geq 0$ is a taste parameter that implies zero trade flows between countries i and j in sector s if $\theta_{ijs} = 0$.

The CES demand relationship at the lower tier follows as

$$x_{ijs} = \left(\frac{p_{ijs}}{\theta_{ijs} P_{js}} \right)^{1-\sigma_s} E_{js}, \tag{1}$$

where x_{ijs} denotes nominal trade flows from country i to country j in sector s, p_{ijs} denotes their unit price, P_{js} is the sectoral CES price index in country j, and E_{js} is the corresponding sectoral expenditure. We assume that trade costs are of the iceberg type such that

$$p_{ijs} = \tau_{ijs} p_{is}, \tag{2}$$

where p_{is} denotes the price (or unit cost) at origin i. We assume a standard log-linear specification of the trade cost function with

$$\ln \tau_{ijs} = \rho_s \ln dist_{ij}, \tag{3}$$

where for simplicity we use bilateral distance $dist_{ij}$ as the sole trade cost component with an elasticity ρ_s that can vary by sector.[7]

Combining equations (1), (2) and (3), we can write the sector-level gravity equation in log-linearized form as

$$\ln x_{ijs} = \phi_{is} + \xi_{js} - (\sigma_s - 1)\, \rho_s \ln dist_{ij} + (1 - \sigma_s) \ln \theta_{ijs}, \tag{4}$$

where the sector-origin fixed effect ϕ_{is} captures the origin price p_{is}, the sector-destination fixed effect ξ_{js} captures the price index P_{js} and expenditure E_{js}, and $(1 - \sigma_s) \ln \theta_{ijs}$ is the error term, traditionally assumed to be independent of trade costs τ_{ijs}. The sector-level trade cost elasticity is thus a function of the elasticity of substitution.

[7] Feenstra, Luck, Obstfeld and Russ (2018) specify a monopolistic competition model with a separate 'macro' elasticity between home and foreign varieties and a 'micro' elasticity between different foreign varieties. We do not make such a distinction but rather focus on the variation of elasticities across sectors.

Aggregate bilateral trade is defined as the sum of bilateral trade flows across sectors $s = 1, \ldots, l$

$$x_{ij} \equiv \sum_{s=1}^{l} x_{ijs} \tag{5}$$

with $i \neq j$. We proceed to show that based on the framework in equations (1)-(3), an aggregate gravity equation can be constructed but only with non-standard properties. For this purpose, we substitute the demand function (1) into the definition of aggregate bilateral trade (5) using equation (2):

$$\begin{aligned} x_{ij} &= \sum_{s} \left(\frac{\tau_{ijs} p_{is}}{\theta_{ijs} P_{js}} \right)^{1-\sigma_s} E_{js} \\ &= \left(\frac{\tau_{ij} p_i}{P_j} \right)^{1-\sigma} E_j \exp(\varepsilon_{ij}), \end{aligned} \tag{6}$$

where σ denotes the aggregate demand elasticity, $\ln \tau_{ij} = \rho \ln dist_{ij}$, p_i is the unit price in country i, P_j is the CES price index in country j, E_j is the corresponding expenditure, and

$$\exp(\varepsilon_{ij}) = \sum_{s} \left(\frac{\tau_{ij} p_i}{P_j} \right)^{\sigma-1} \left(\frac{\tau_{ijs} p_{is}}{\theta_{ijs} P_{js}} \right)^{1-\sigma_s} \frac{E_{js}}{E_j}. \tag{7}$$

Taking logarithms of equation (6) implies

$$\ln x_{ij} = \Phi_i + \Xi_j - (\sigma - 1)\, \rho \ln dist_{ij} + \varepsilon_{ij}, \tag{8}$$

with $\Phi_i = (1 - \sigma) \ln p_i$ and $\Xi_j = (\sigma - 1) \ln P_j + \ln E_j$.

Superficially, equation (8) has the same structure as a conventional log-linearized gravity equation, but the key point is that ε_{ij} should not be considered a standard error term because it is by construction a function of bilateral trade costs τ_{ij}.[8] The exception is the case where θ_{ijs} is the only source of sectoral heterogeneity, and therefore $\sigma_s = \sigma$, $p_{is} = p_i$, $P_{js} = P_j$, $E_{js} = E_j$, and $\tau_{ijs} = \tau_{ij} = dist_{ij}^{\rho}$. In this special case we have

$$\exp(\varepsilon_{ij}) = \sum_s \theta_{ijs}^{\sigma-1},$$

and therefore (8) would be a proper log-linearized gravity equation. We will use this result later.

3 Motivating evidence

The theoretical framework in Section 2 delivers a gravity equation (4) at the disaggregate level as well as an equation (8) at the aggregate level that can be construed as a non-standard gravity equation. Building upon Baier and Bergstrand's (2007) seminal work on the effects of free trade agreements, we now explore empirically how estimated coefficients on gravity variables behave at different levels of aggregation, and in the next section we discuss the observed empirical patterns from a theoretical perspective.

Baier and Bergstrand's estimation framework is based on an OLS regression of logarithmic trade flows on multiple categories of fixed effects and dummies for whether two countries have a trade agreement in place. Specifically, they consider models of the form

$$\ln x_{ijt} = \alpha_{it} + \alpha_{jt} + \alpha_{ij} + \beta_1 FTA_{ijt} + \beta_2 FTA_{ijt-1} + \beta_3 FTA_{ijt-2} + \varepsilon_{ijt}, \tag{9}$$

[8] The result in equation (8) resonates with Redding and Weinstein (2019a and 2019b) who also demonstrate that in a nested CES demand system as above, a log-linear gravity equation can be derived at the aggregate level but only with an error term that is not orthogonal to bilateral trade costs.

where x_{ijt} are imports of country j from country i in year t, the FTA dummies (which include a contemporaneous term as well as two lags to allow for phasing-in effects) are the regressors of interest, α_{it} and α_{jt} denote exporter-year and importer-year fixed effects that control for price index and expenditure terms, α_{ij} are bilateral fixed effects introduced by Baier and Bergstrand to help address the potential endogeneity of free trade agreements, and ε_{ijt} is the error term.

Following Santos Silva and Tenreyro (2006), we also use the PPML estimator of Gourieroux, Monfort, and Trognon (1984) to estimate models of the form

$$x_{ijt} = \exp\left(\alpha_{it} + \alpha_{jt} + \alpha_{ij} + \beta_1 FTA_{ijt} + \beta_2 FTA_{ijt-1} + \beta_3 FTA_{ijt-2}\right)\eta_{ijt}, \qquad (10)$$

where η_{ijt} is a multiplicative error term.

Baier and Bergstrand estimate (9) using country-level bilateral trade data from the IMF's Direction of Trade Statistics (DOTS). In our context, the key question is how the coefficient estimates on the FTA terms change as we vary the level of aggregation. For this purpose we replicate Baier and Bergstrand's key results but using data from the UN Comtrade database which provides trade flows at different levels of aggregation.[9] This allows us to show results from estimating models (9) and (10) at three different levels

[9] Baier and Bergstrand use data from the IMF's DOTS for the years 1960-2000 at five-year intervals for 96 countries, excluding zero trade flows. To achieve a similar timespan, we rely on data from Comtrade, based on the SITC classification, for the same countries and for the years 1962, 1965, 1970, ..., 2000 (no data are available prior to 1962 so we use 1962 data for 1960). Specifically, we use the value of bilateral imports in current US dollars on a c.i.f. basis. These data are available at five different levels of aggregation, from SITC 4-digit to the country-level bilateral trade flows used by Baier and Bergstrand (SITC 0-digit). Our data on FTAs are the same as in Baier and Bergstrand (2007), based on their Table 3.

of aggregation: aggregate bilateral imports, and imports at the 2-digit and 4-digit SITC levels.[10]

The results obtained when estimating models (9) and (10) at different levels of aggregation are presented in Table 1. Specifically, for each estimator we present results at three levels of aggregation and, when using disaggregated data, we present results for models imposing that the fixed effects are the same across sectors and for models where the fixed effects are allowed to vary by sector.[11] For now the coefficients on the FTA dummies are constrained to be the same for all sectors. Although we present estimates for each of the three FTA dummies, we will focus our discussion on the total FTA effect, which is reported in the last column of the table and is computed as the sum of the estimated coefficients on the three FTA dummies.

Reassuringly, our results for the specification most directly comparable to Baier and Bergstrand's (the one using OLS estimation with aggregate trade) are similar to theirs. We obtain a total FTA effect of 0.714 log points (see the last column of the first line of Table 1) compared to 0.76 log points in the key specification by Baier and Bergstrand (see their Table 5, column 4). This demonstrates that changing the data source from the IMF

[10]For each country pair in the data, we observe trade flows for 61 2-digit SITC sectors and 625 4-digit SITC sectors. However, we drop all 4-digit sectors with fewer than $2,000$ observations of positive trade flows. This is because in the sector-level regressions discussed below, it is not always possible to identify all the parameters of interest when the number of positive observations is small. To keep our samples comparable across different sections of this paper, we also exclude such observations for the pooled regressions presented here. Dropping these observations reduces the number of 4-digit sectors to 576 and that of 2-digit sectors to 60.

[11]Models with sector-level fixed effects include importer-year-sector, exporter-year-sector and exporter-importer-sector fixed effects, whereas models without sector-level fixed effects include only importer-year, exporter-year and importer-exporter fixed effects.

DOTS database to UN Comtrade does not in itself change the basic findings in Baier and Bergstrand (2007).[12]

We now turn to our question of how results change as we change the degree of aggregation in our data. The results in Table 1 show that OLS estimates are sensitive to the level of aggregation, irrespective of the fixed effects we use. In contrast, with PPML the estimated coefficients and standard errors are invariant to aggregation if we do not allow the fixed effects to vary by sector, being exactly the same for the first three rows of the PPML panel in Table 1.

Moreover, when we use PPML, estimation with aggregate data is equivalent to using disaggregated data to estimate models where the fixed effects are assumed to be the same for all sectors. To see this, note that the change in the estimates and standard errors resulting from aggregation (e.g., going from the bottom row to the top row in the PPML panel) is the same as the change resulting from removing the sector dimension from the fixed effects (e.g., going from the bottom row to the middle row in the PPML panel).

Naturally, when we include the sector dimension into the fixed effects, the PPML estimates do depend on the level of aggregation because changing the level of aggrega-

[12] Note that the two samples are not fully comparable because Baier and Bergstrand use log exports as their dependent variable and thus have to exclude observations with zero flows from their sample. By contrast, the results in Table 1 are based on a fully rectangularized set of bilateral trade flows following current best practice in applied international trade research (see, e.g., Yotov, Piermartini, Monteiro and Larch, 2016). That is, we fill in all missing country pair-sector-year combinations and assign a trade flow value of zero for all such "filled in" observations. While the additional zero observations get dropped when taking logs (as we do for our OLS specifications), rectangularization also changes the structure of the lags of the FTA regressors, making Baier and Bergstrand's and our rectangularized data incompatible. As an additional comparability check, we have also re-estimated Baier and Bergstrand's key specification on our sample without rectangularizing the data, obtaining a total FTA effect of 0.77, which is very similar to the total FTA effect of 0.76 estimated by Baier and Bergstrand (note, however, that some differences in the individual coefficient estimates remain; results available on request).

tion changes the model specification. However, even in this case the PPML estimates are somewhat robust to changes in the level of aggregation, dropping from 0.591 at the aggregate level to 0.500 at the 4-digit level.

Until now we have constrained the coefficients on the FTA dummies to be the same across sectors. We now relax this restriction and estimate (9) and (10) separately for each SITC sector at both the 2-digit and the 4-digit levels. That is, we now allow both the fixed effects and the coefficients on the FTA dummies to vary by sector. This yields 60 sets of estimates for each equation at the 2-digit level, and 576 sets at the 4-digit level.

Figure 1 presents kernel density estimates of the total estimated FTA effects obtained with OLS and PPML for each of the sectors at the 2-digit and 4-digit levels. In each panel, we add two elements to help interpret the results. The vertical solid line represents the estimated effect obtained by estimating the models with the aggregate data (this corresponds to the results in the first line of each panel in Table 1). We also include a vertical dashed line representing the weighted average of the estimated coefficients using shares of 2-digit or 4-digit sectors in total trade as weights.

Figure 1 shows that in all cases the aggregate estimates lie reasonably close to the mode of the distribution of the sector-level estimates. Moreover, with PPML, the aggregate estimates are very close to the weighted averages of the disaggregate estimates (i.e., the solid and dashed lines are very close to each other); the same does not necessarily happen with OLS.

4 Aggregation of constant-elasticity models

We now examine the effects of aggregation from an econometric point of view and explain the patterns found in Section 3. We draw a distinction between disaggregate-level *parameters* on the one hand (where coefficients vary at the sector level, for instance by letting the coefficients on the FTA regressors from Section 3 vary across SITC sectors), and disaggregate-level *regressors* on the other hand (where regressors themselves differ across sectors, for example the tariffs considered by Imbs and Mejean, 2015).

4.1 Set-up and aggregation with OLS

Consistent with our earlier theoretical framework (see equation 1), we assume that sector-level trade flows (x_{ijs}) are described by the following constant-elasticity model

$$x_{ijs} = \exp\left(z'_{ijs}\beta_s\right)\eta_{ijs}, \tag{11}$$

where z_{ijs} is a vector of regressors that includes trade cost variables and fixed effects, η_{ijs} is a non-negative error term such that $\mathbb{E}\left(\eta_{ijs}|z_{ijs}\right) = 1$, and β_s is a vector of parameters that are potentially allowed to vary with s and in which the slope parameters have the usual interpretation as (semi-) elasticities. The model for the aggregate data is given by (see equation 5)

$$x_{ij} = \sum_{s=1}^{l} x_{ijs} = \sum_{s} \exp\left(z'_{ijs}\beta_s\right)\eta_{ijs}, \tag{12}$$

which in general is not a constant-elasticity model (see the discussion around equation 8).

The traditional approach to estimating models such as (11) is to take logarithms of both sides and estimate

$$\ln x_{ijs} = z'_{ijs}\beta_s + \ln \eta_{ijs} \tag{13}$$

by OLS, under the assumption that $\mathbb{E}\left(\ln \eta_{ijs}|z_{ijs}\right)$ is constant. Lewbel (1992) and van Garderen, Lee and Pesaran (2000), among others, have studied the consequences of estimating the aggregate counterpart of (13) and concluded that the parameters of interest can only be identified under very restrictive assumptions.

In the remainder of this section, we study the consequences of using aggregated data when the multiplicative model is estimated using PPML. We focus on PPML because it is the only pseudo maximum likelihood estimator that is valid in models with high-dimensional fixed effects (Weidner and Zylkin, 2021), that is not adversely affected by the possible non-existence of the estimates (Correia, Guimarães and Zylkin, 2021), and whose results are compatible with structural gravity models (Fally, 2015). PPML is therefore singularly suited to the estimation of gravity equations. However, in Appendix A, we present results for other pseudo maximum likelihood estimators based on the linear exponential family (see Gourieroux, Monfort, and Trognon, 1984).

4.2 Aggregation with PPML

Building on our earlier distinction between parameters and/or regressors varying at the sector level, we consider four particular cases of this problem. Case 1 is the simplest scenario where neither regressors nor parameters vary with s. In Case 2 the parameters vary with s but regressors do not, and the reverse holds in Case 3. Finally, in Case 4 both parameters and regressors vary with s. In Appendix B we report the results of a simulation experiment that illustrates the results of this section.

4.2.1 Case 1: Parameters and regressors are constant

We have seen in Section 2 that in this particular case we have proper gravity equations at the disaggregate and aggregate levels. In this case equation (11) can be written as

$$x_{ijs} = \exp\left(z_{ij}'\beta\right)\eta_{ijs}, \tag{14}$$

and expression (12) becomes

$$x_{ij} = \sum_{s} \exp\left(z'_{ij}\beta\right)\eta_{ijs} = \exp\left(\ln l + z'_{ij}\beta\right)\eta^{*}_{ij}, \tag{15}$$

where $\eta^{*}_{ij} = l^{-1}\sum_{s=1}^{l}\eta_{ijs}$ is an error term such that $\mathbb{E}\left(\eta^{*}_{ij}|z_{ij}\right) = 1$. Therefore, as discussed in Section 2, in this particular case both x_{ijs} and x_{ij} are described by stochastic constant-elasticity models.

It is easy to show that the PPML estimates of the slopes in (14) and (15) are identical, a result first noted in the simulation evidence reported by Amrhein and Flowerdew (1992). To see this, notice that the first-order condition of the PPML estimator of β in (14) is (Gourieroux, Monfort, and Trognon, 1984)

$$S\left(\hat{\beta}\right) = \sum_{ijs}\left(x_{ijs} - \exp\left(z'_{ij}\hat{\beta}\right)\right) z_{ij} = 0,$$

where a "hat" is used to denote parameter estimates and $\sum_{ijs}$ is shorthand for $\sum_{i}\sum_{j}\sum_{s}$. This condition can be written as

$$S\left(\hat{\beta}\right) = \sum_{ij}\left(x_{ij} - \exp\left(\ln l + z'_{ij}\hat{\beta}\right)\right) z_{ij} = 0,$$

which is the first-order condition of the PPML estimator of β in the aggregate model defined by (15). Hence, the estimation results are invariant to the level of aggregation of the data (with the exception of the intercept which is adjusted to reflect the number of sectors being aggregated). Moreover, if the dependent variable in the aggregate equation is the mean of x_{ijs} rather than its sum, the estimates are exactly the same at both levels, and the invariance result continues to apply. Additionally, it is possible to show that the cluster-robust estimate of the covariance matrix for the estimates from (14) is identical to the estimate of the robust covariance matrix for the estimates in the aggregate equation when the dependent variable is the average of x_{ijs} over s. Therefore, the level of aggregation does not matter for the significance of the estimates either.

It is important to note that the results above are obtained under the assumption that the number of sectors l is the same for every ij pair. When that is not the case, the same result holds in a panel where the models include pair fixed effects that will absorb the differences in the number of sectors by pair. If the disaggregate model does not include pair fixed effects, the aggregate and disaggregate elasticity estimates will not be numerically identical in finite samples. However, the two estimates converge to the same limit if the aggregate model includes pair fixed effects to account for the differences in the number of sectors across pairs. To simplify the exposition, in what follows we continue to assume that the number of sectors l is the same for every pair.

In summary, when both the parameters and the regressors are constant across s, both x_{ijs} and x_{ij} are given by constant-elasticity models with the same parameters, and the PPML estimates and standard errors are invariant to the level of aggregation of the data. This invariance result plays a central role in the analysis of Cases 2 to 4 because it implies that since aggregation in itself will not cause a bias, in those cases we only have to consider the effect of ignoring heterogeneity when using disaggregated data.[13]

Looking back at our results from Section 3, we note that the models underlying the estimates in the first three lines of the PPML panel of Table 1 fall into our Case 1 (neither parameters nor regressors vary with s). Thus, the invariance result just outlined explains why the estimates and standard errors obtained with these models are exactly the same.

[13]This result may be interesting in its own right. For example, it implies that wage equations estimated by PPML using monthly or annual data will deliver the same estimates as long as the regressors do not have within-year variation.

4.2.2 Case 2: Parameters vary with s but regressors do not

In this case the relevant model at the disaggregate level is

$$x_{ijs} = \exp\left(z'_{ij}\beta_s\right)\eta_{ijs}.$$

Clearly, now it is not possible to recover the sectoral parameters from aggregate data, but it is interesting to study what we estimate when using aggregate data.

To see the effect of ignoring parameter heterogeneity, write the first-order conditions for the estimates of β_s with sectoral data as

$$S_s\left(\hat{\beta}_s\right) = \sum_{ij}\left(x_{ijs} - \exp\left(z'_{ij}\hat{\beta}_s\right)\right)z_{ij} = 0, \qquad s = 1, \ldots, l.$$

Since we have $S_s\left(\hat{\beta}_s\right) = 0$ for each s, for the full sample we have $\sum_s S_s\left(\hat{\beta}_s\right) = 0$. Imposing homogeneity we estimate a single parameter for all s, say $\hat{\beta}^r$, which by definition will satisfy $S\left(\hat{\beta}^r\right) = \sum_s S_s\left(\hat{\beta}^r\right) = 0$.[14]

To study the relation between $\hat{\beta}^r$ and $\hat{\beta}_s$, $s = 1, \ldots, l$, we can use the mean value theorem to write

$$\sum_s S_s\left(\hat{\beta}_s\right) = \sum_s S_s\left(\hat{\beta}^r\right) - \sum_s H_s\left(\beta_s^*\right)\left(\hat{\beta}_s - \hat{\beta}^r\right)$$

with $H_s\left(\beta_s^*\right) = -\left.\partial S_s(b)/\partial b\right|_{b=\beta_s^*}$, where β_s^* is a point between $\hat{\beta}_s$ and $\hat{\beta}^r$.

[14] Note, however, that $S_s\left(\hat{\beta}^r\right) \neq 0$; that is, the estimates obtained imposing homogeneity do not satisfy the first-order conditions for each s.

Because $\sum_s S_s\left(\hat{\beta}_s\right) = \sum_s S_s\left(\hat{\beta}^r\right) = 0$, we can write

$$\hat{\beta}^r = \left[\sum_s H_s\left(\beta_s^*\right)\right]^{-1} \sum_s H_s\left(\beta_s^*\right)\hat{\beta}_s, \tag{16}$$

and therefore $\hat{\beta}^r$ can be interpreted as an average of the estimates of β_s weighted by the matrices $H_s\left(\beta_s^*\right)$.

Noting that $H_s\left(\beta_s^*\right) = \sum_{ij}\left(\exp\left(z_{ij}'\beta_s^*\right) z_{ij} z_{ij}'\right)$, we can see that $H_s\left(\beta_s^*\right)$ is itself a weighted sum of $\exp\left(z_{ij}'\beta_s^*\right)$, where the weights do not depend on s. Because $\exp\left(z_{ij}'\beta_s^*\right)$ is closely related to the expectation of x_{ijs}, we can interpret $\hat{\beta}^r$ as a weighted average of the estimates of β_s, giving more weight to the estimates from the subsamples where x_{ijs} tends to be larger. That is, the estimate obtained with aggregate data is approximately given by $\sum_s q_s\hat{\beta}_s$, where $q_s = \sum_{ij} x_{ijs}/\sum_{ijs} x_{ijs}$ denotes the share of sector s trade in total trade.

It follows from the invariance result in Case 1 that the parameters estimated with aggregated data can also be interpreted as weighted averages of the estimates of the individual parameters with weights given by $H_s\left(\beta_s^*\right)$, which are approximately equal to the average of the individual estimates weighted by q_s.

Looking again at our findings from Section 3, the results for Case 2 explain why in Figure 1 the trade-weighted average of the sectoral estimates (indicated by the dashed vertical line) is always close to the PPML estimate obtained with aggregate data (represented by the solid vertical line). This finding is also confirmed by the simulation results reported in Appendix B.

4.2.3 Case 3: Regressors vary with s but parameters do not

Now the relevant model is

$$x_{ijs} = \exp\left(z'_{ijs}\beta\right)\eta_{ijs}, \tag{17}$$

and we start by considering the effect of estimating

$$x_{ijs} = \exp\left(z'_{ij}\beta^a\right)\eta^a_{ijs}, \tag{18}$$

where z_{ij} is obtained by aggregating z_{ijs}, β^a denotes the parameters of the aggregate equation, and η^a_{ijs} is a non-negative error term whose properties are determined by how β^a is defined.

Letting $z_{ijs} = z_{ij} + \varepsilon_{ijs}$, we can write equation (17) as

$$x_{ijs} = \exp\left(z'_{ij}\beta + \varepsilon'_{ijs}\beta\right)\eta_{ijs}, \tag{19}$$

and we can then interpret (18) as resulting from omitting $\varepsilon'_{ijs}\beta$ from (19).

General results on the effects of omitted variables are difficult to obtain for non-linear models (see, Kiefer and Skoog, 1984, Neuhaus and Jewell, 1993, and Drake and McQuarrie, 1995). Therefore, to gain some insight into the effect of omitting $\varepsilon'_{ijs}\beta$, it is useful to start by considering as an illustrative example the case where, conditional on z_{ij}, $\varepsilon'_{ijs}\beta$ has a normal distribution with mean $z'_{ij}\mu$ and variance $z'_{ij}\omega$, and

$$\mathbb{E}\left[x_{ijs}|z_{ijs}\right] = \mathbb{E}\left[x_{ijs}|z_{ij},\varepsilon_{ijs}\right] = \exp\left(\left(z'_{ij} + \varepsilon'_{ijs}\right)\beta\right).$$

In this example $\exp\left(\varepsilon'_{ijs}\beta\right)$ is log-normal with

$$\mathbb{E}_{\varepsilon_{ijs}}\left[\exp\left(\varepsilon'_{ijs}\beta\right)|z_{ij}\right] = \exp\left(z'_{ij}\mu + 0.5z'_{ij}\omega\right),$$

and therefore

$$\begin{aligned}\mathbb{E}\left[x_{ijs}|z_{ij}\right] &= \mathbb{E}_{\varepsilon_{ijs}}\left[\mathbb{E}\left[x_{ijs}|z_{ij},\varepsilon_{ijs}\right]|z_{ij}\right] = \exp\left(z_{ij}'\beta\right)\mathbb{E}_{\varepsilon_{ijs}}\left[\exp\left(\varepsilon_{ijs}'\beta\right)|z_{ij}\right] \\ &= \exp\left(z_{ij}'\beta + z_{ij}'\mu + 0.5z_{ij}'\omega\right) = \exp\left(z_{ij}'\beta^{a}\right),\end{aligned}$$

with $\beta^a = \beta + \mu + 0.5\omega$, which simplifies to $\beta^a = \beta + 0.5\omega$ when z_{ij} is defined as the mean of z_{ijs}.

If z_{ijs} consists only of tariffs and z_{ij} is the mean of z_{ijs}, ignoring sectoral heterogeneity will bias the estimates upward (towards zero) because β is negative and ω is positive (as $z_{ij}\omega$ is a variance). If z_{ij} is defined in a different way, the bias will also depend on μ, and therefore it is more difficult to predict its direction.

More generally, the difference between the parameters at the two levels of aggregation depends on how z_{ij} is defined and on how the conditional expectation of $\exp\left(\varepsilon_{ijs}'\beta\right)\eta_{ijs}$ is related to z_{ij}. Since in most applications $\varepsilon_{ijs}'\beta$ will have little relation with η_{ijs}, the bias of β^a will be largely determined by the relation between the conditional moments of z_{ijs} and z_{ij}, as in the example above. In any case, $\hat{\beta}^a$ is such that the fitted values of the aggregate model approximate some of the characteristics of the fitted values of the regression with disaggregated data, and in that sense $\hat{\beta}^a$ provides an approximation to $\hat{\beta}$. Indeed, assuming that the models include intercepts, the residuals of both models will have zero mean, with the residuals of the disaggregate model being orthogonal to z_{ijs}, while the residuals of the aggregate model are orthogonal to z_{ij} but only approximately orthogonal to the disaggregate regressors.

Combining these results with those for Case 1, we can conclude that the aggregate model will estimate β^a rather than β (except for the intercept), and it may be possible to predict the magnitude and sign of the differences between the elements of the two vectors when, as in Imbs and Mejean (2015), we have information on how the conditional moments of the omitted variable $\varepsilon'_{ijs}\beta$ vary with z_{ij}. For example, if tariffs are the only regressor with disaggregate-level variation and the aggregate regressor is the average tariff, the bias is likely upward because higher average collected tariffs tend to be associated with a higher variance across sectors (see, e.g., Pritchett and Sethi, 1994). Further information on the performance of the PPML estimator in this case is provided by the simulation study reported in Appendix B. It suggests that the bias can be small even when the regressors have a reasonable amount of variation at the disaggregate level.

4.2.4 Case 4: Regressors and parameters vary with s

The case where both the regressors and the parameters vary with s can be addressed by combining earlier results. As we know from Case 3, the effect of replacing z_{ijs} with z_{ij} in the regressions for each s is that in each case we estimate a vector β^a that is an approximation to β_s. From Case 2 we know that imposing the same coefficients for all s leads to a weighted average of these individual estimates. Therefore, it follows from the invariance result for Case 1 that in Case 4 we estimate a weighted average of the approximations to β_s.

As in Case 3, in this case we need additional information, including on how the conditional moments of z_{ijs} are related to z_{ij}, to be able to interpret the estimates obtained with aggregate data. However, the simulation experiment presented in Appendix B suggests that in Case 4 the parameter identified in Case 2 is estimated with a bias similar to that observed in Case 3, which is not particularly severe in some of the scenarios we consider. Of course, if there is strong disaggregate-level variation of the regressors, meaningful estimates can only be obtained by using appropriately disaggregated data.

4.2.5 Summary of the main results and practical implications

Which of the four cases considered above is the more relevant one will depend on the particular application. However, as illustrated in Section 3, Case 2 is arguably the case of interest when estimating conventional gravity equations. Indeed, in many applications, gravity equations do not include trade cost variables such as tariffs that vary by sector, and therefore only the fixed effects depend on s.

These sectoral fixed effects can be interpreted as a set of dummies that depend on s but with constant coefficients, and hence can be seen as examples of Case 3. This way of approaching the problem is similar to that of French (2019) in that he also establishes that the consequences of aggregation in this context are equivalent to the omission of a variable.

Alternatively, the sectoral fixed effects can be interpreted as a set of dummies whose coefficients vary with s. Therefore, these models can also be seen as an example of Case 2 where at least the coefficients on the fixed effects vary with s. This alternative interpretation is interesting because it follows from our results that in this leading case, the PPML estimates of the aggregate model have a meaningful interpretation as a weighted average of the sector-specific parameters, something that is illustrated in Figure 1 and confirmed by the simulation results in Appendix B.

In applications where trade cost variables have disaggregate-level variation, PPML estimates will be biased but, as illustrated in Case 3, it may be possible to determine the direction of the bias when there is information on the distribution of the disaggregate regressors. Moreover, the simulation results reported in Appendix B reveal that in the settings we consider the PPML biases can be relatively small.

Our results also have important implications for the use of gravity equation parameter estimates, such as for forecasting purposes. We illustrate this point in Appendix C where we use the OLS and PPML parameter estimates to predict the trade effects of free trade agreements.

5 Conclusion

We investigate the consequences of aggregation for the estimation of gravity equations using PPML. We show that the estimation results are invariant to changes in the level of aggregation in the most favorable case where neither the regressors nor the parameters vary at the disaggregate level. In less favorable cases where the parameters, the regressors, or both vary across units, the PPML estimates depend on the level of aggregation, but it may still be possible to provide a meaningful interpretation of the estimates obtained with aggregate data.

Specifically, we argue that for the empirically most relevant case where parameters vary across sectors but trade cost variables do not (such as distance and common language), PPML estimates are approximately trade-weighted averages of the underlying sector-level parameters, and hence they still provide economically meaningful information. When trade cost variables vary at the sector level, PPML produces biased estimates, but in some cases it may be possible to determine the direction of the bias when, as in Imbs and Mejean (2015), we have information about the distribution of the regressors at the sector level. Encouragingly, our simulation results suggest that this bias can be small.

Overall, our findings provide guidance on the interpretation of PPML estimates obtained using aggregate data and should be helpful to applied researchers who may not have disaggregated data at their disposal.

References

Amrhein, C.G, and Flowerdew, R. (1992). "The Effect of Data Aggregation on a Poisson Regression Model of Canadian Migration," *Environment and Planning A: Economy and Space* 24, 1381-1391.

Anderson, J.E. (2011). "The Gravity Model," *Annual Review of Economics* 3, 133-160.

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

Correia, S., Guimarães, P., and Zylkin, T. (2021). "Verifying the Existence of Maximum Likelihood Estimates for Generalized Linear Models," arXiv:1903.01633v6.

Drake, C., and McQuarrie, A. (1995). "A Note on the Bias due to Omitted Confounders," *Biometrika* 82, 633-638.

Fally, T. (2015). "Structural Gravity and Fixed Effects," *Journal of International Economics* 97, 76-85.

Feenstra, R., Luck, P., Obstfeld, M., and Russ, K. (2018). "In Search of the Armington Elasticity," *Review of Economics and Statistics* 100, 135-150.

French, S. (2019). "Comparative Advantage and Biased Gravity," mimeo, available from http://research.economics.unsw.edu.au/sfrench/documents/French_Biased_Gravity_4_3.pdf.

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

Head, K., and Mayer, T. (2014). "Gravity Equations: Workhorse, Toolkit, and Cookbook," in G. Gopinath, E. Helpman, and K. Rogoff, eds., *Handbook of International Economics*, Volume 4, Chapter 3, Amsterdam: Elsevier, 131-195.

Imbs, J., and Mejean, I. (2015). “Elasticity Optimism,” *American Economic Journal: Macroeconomics* 7, 43-83.

Kiefer, N.M., and Skoog, G.R. (1984). “Local Asymptotic Specification Error Analysis,” *Econometrica* 52, 873-885.

Lewbel, A. (1992). “Aggregation with Log-Linear Models,” *Review Economic Studies* 59, 635-642.

Neuhaus, J.M., and Jewell, N.P. (1993). “A Geometric Approach to Assess Bias Due to Omitted Covariates in Generalized Linear Models,” *Biometrika* 80, 807-815.

Pritchett, L., and Sethi, G. (1994). “Tariff Rates, Tariff Revenue, and Tariff Reform: Some New Facts,” *World Bank Economic Review* 8, 1-16.

Redding, S., and Weinstein, D. (2019a). “Aggregation and the Gravity Equation,” *AEA Papers and Proceedings* 109, 450-455.

Redding, S., and Weinstein, D. (2019b). “Aggregation and the Gravity Equation,” NBER Working Paper 25464.

Santos Silva, J.M.C., and Tenreyro, S. (2006). “The Log of Gravity,” *Review of Economics and Statistics* 88, 641-658.

van Garderen, K.J., Lee, K., and Pesaran, M.H. (2000). “Cross-Sectional Aggregation of Non-Linear Models,” *Journal of Econometrics* 95, 285-331.

Weidner, M., and Zylkin, T. (2021). “Bias and Consistency in Three-way Gravity Models,” *Journal of International Economics* 132, 103513.

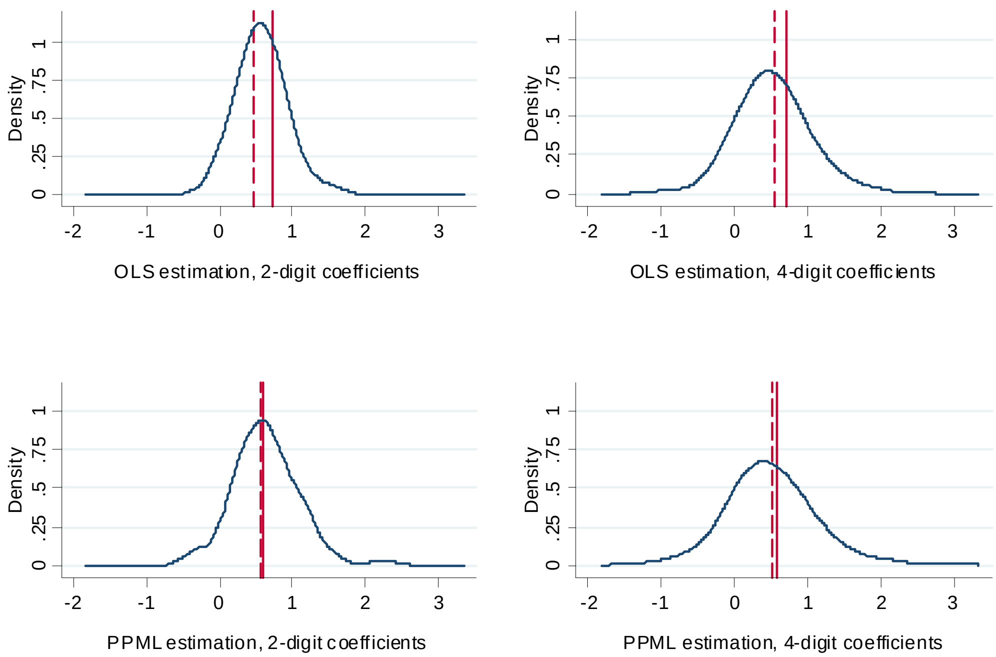
Yotov, Y.V., Piermartini, R., Monteiro, J.-A., and Larch, M. (2016). *An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model*, Geneva: World Trade Organization.

Table 1: Regression results at different aggregation levels

Aggregation level	Sample size (# clusters)	Sector-level fixed effects	Regressor coefficients (Standard errors clustered by pair) FTA_t	FTA_{t-1}	FTA_{t-2}	**Total**
			Estimator: OLS			
Aggr. trade	43,480 (7,896)	No	0.174*** (0.0453)	0.379*** (0.0455)	0.161*** (0.0510)	**0.714*** (0.0608)**
SITC 2-digit	1,142,000 (8,286)	No	0.355*** (0.0212)	0.191*** (0.0206)	−0.004 (0.0246)	**0.542*** (0.0307)**
SITC 4-digit	5,562,720 (8,307)	No	0.285*** (0.0165)	0.161*** (0.0140)	0.036** (0.0180)	**0.481*** (0.0268)**
SITC 2-digit	1,079,880 (7,487)	Yes	0.334*** (0.0225)	0.168*** (0.0209)	0.064*** (0.0255)	**0.566*** (0.0326)**
SITC 4-digit	5,039,915 (7,274)	Yes	0.326*** (0.0195)	0.116*** (0.0159)	0.087*** (0.0199)	**0.529*** (0.0314)**
			Estimator: PPML			
Aggr. trade	50,013 (8,409)	No	0.278*** (0.0324)	0.224*** (0.0242)	0.089*** (0.0267)	**0.591*** (0.0455)**
SITC 2-digit	3,009,960 (8,562)	No	0.278*** (0.0324)	0.224*** (0.0242)	0.089*** (0.0267)	**0.591*** (0.0455)**
SITC 4-digit	28,895,616 (8,562)	No	0.278*** (0.0324)	0.224*** (0.0242)	0.089*** (0.0267)	**0.591*** (0.0455)**
SITC 2-digit	1,710,888 (8,409)	Yes	0.235*** (0.0252)	0.179*** (0.0171)	0.118*** (0.0217)	**0.533*** (0.0338)**
SITC 4-digit	9,718,394 (8,400)	Yes	0.210*** (0.0270)	0.172*** (0.0160)	0.117*** (0.0201)	**0.500*** (0.0353)**

Notes: The table presents the results of estimating three-way gravity equations. The dependent variable is logarithmic trade in the OLS panel and the level of trade in the PPML panel. The reported sample sizes exclude singletons and observations separated by fixed effects; the full sample sizes at the 0-, 2-, and 4-digit SITC levels are, respectively, 63,840, 3,830,400, and 36,771,840. Models with sector-level fixed effects ("Yes") include importer-year-sector, exporter-year-sector and exporter-importer-sector fixed effects, whereas models without sector-level fixed effects ("No") include only importer-year, exporter-year and importer-exporter fixed effects; the coefficients on the FTA dummies are constrained to be the same for all sectors. Statistically significant at *** 0.01, ** 0.05, * 0.1.

Figure 1: Kernel density plot of the estimated FTA effects at sectoral level



Notes: The dashed line is the trade-weighted average of the estimated sectoral effects, and the solid line is the effect estimated with aggregate data. The left panels show estimates at the 2-digit level, and the right panels show estimates at the 4-digit level. The top panels are estimated with OLS and the bottom panels with PPML.