Trade, Gravity and Aggregation

Holger Breinlich[1] Dennis Novy[2] J.M.C. Santos Silva[3]

University of Surrey University of Warwick University of Surrey

16 June 2022

ONLINE APPENDIX

Appendix A: Additional results on the effects of aggregation

Appendix B: Simulation evidence

Appendix C: Implications for gravity-based predictions: An application to free trade agreements

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

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

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

Appendix A:
Additional results on the effects of aggregation

Introduction and setup

Besides the advantages highlighted by Santos Silva and Tenreyro (2006), the Poisson pseudo maximum likelihood (PPML) estimator is the only pseudo maximum likelihood (PML) 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). Therefore, PPML is particularly well suited to estimate gravity equations. However, in other contexts, other PML estimators can be interesting. In this appendix, we sketch the extension of our results on aggregation effects in models estimated by PPML to the case where other PML estimators based on the linear exponential family are used; we focus on the popular non-linear least squares (NLS) and gamma PML (GPML) estimators (see, e.g., Manning and Mullahy, 2001) but the results apply more widely.

We assume that a sector-level dependent variable y_{ijs} is described by a stochastic constant-elasticity model of the form

$$y_{ijs} = \exp\left(z'_{ijs}\beta_s\right)\eta_{ijs}, \tag{1}$$

where z_{ijs} is a vector of explanatory variables, η_{ijs} is a non-negative error term such that $\mathbb{E}\left(\eta_{ijs}|z_{ijs}\right) = 1$, β_s is a vector of parameters that are allowed to vary with s and in which the slope parameters have the usual interpretation as (semi-) elasticities. As

before, aggregation leads to

$$y_{ij} = \sum_{s=1}^{l} y_{ijs} = \sum_{s} \exp\left(z'_{ijs}\beta_s\right) \eta_{ijs}, \tag{2}$$

and we consider four particular cases of the aggregation problem. In Case 1 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.

Case 1: Parameters and regressors are constant

The first-order condition of the estimation of β in (1) using a PML estimator based on the linear exponential family has the form (see Gourieroux, Monfort, and Trognon, 1984)

$$S\left(\hat{\beta}\right) = \sum_{ijs} \left(y_{ijs} - \exp\left(z'_{ij}\hat{\beta}\right)\right) f\left(z'_{ij}\hat{\beta}\right) z_{ij} = 0, \tag{3}$$

where a "hat" is used to denote parameter estimates and $\sum_{ijs}$ is shorthand for $\sum_i \sum_j \sum_s$. In (3), $f\left(z'_{ij}\hat{\beta}\right)$ is a function of $z'_{ij}\hat{\beta}$ that defines the estimator. In particular, we have $f\left(z'_{ij}\hat{\beta}\right) = \exp\left(z'_{ij}\hat{\beta}\right)$ for NLS, $f\left(z'_{ij}\hat{\beta}\right) = 1$ for PPML, and $f\left(z'_{ij}\hat{\beta}\right) = \exp\left(-z'_{ij}\hat{\beta}\right)$ for GPML. That is, NLS gives extra weight to the observations with large expected values of the dependent variable, GPML down-weights observations with large expected values of the dependent variable, whereas PPML gives the same weight to all observations.

Because $f\left(z'_{ij}\hat{\beta}\right)$ does not depend on s, (3) can be written as

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

which is the first-order condition of the PML estimator of β in the aggregate model defined by (2) when $f\left(\ln l + z_{ij}'\hat{\beta}\right)$ is proportional to $f\left(z_{ij}'\hat{\beta}\right)$, as is the case for the NLS, PPML, and GPML estimators. Therefore, these estimators are invariant to aggregation (except for the intercept) when neither the parameters nor the regressors vary with s.

Case 2: Parameters vary with s but regressors do not

As in Section 4, it is possible to show that estimating the model with disaggregated data but ignoring the parameter heterogeneity, we estimate a parameter defined by

$$\hat{\beta}^r = \left[\sum_s H_s\left(\beta_s^*\right)\right]^{-1} \sum_s H_s\left(\beta_s^*\right)\hat{\beta}_s, \tag{4}$$

and thus the estimates obtained in this case can be interpreted as an average of the estimates of β_s weighted by the matrices $H_s\left(\beta_s^*\right) = -\,\partial S_s\left(b\right)/\partial b|_{b=\beta_s^*}$, where β_s^* is a point between $\hat{\beta}_s$ and $\hat{\beta}^r$.

Therefore, the interpretation of $\hat{\beta}^r$ depends only on the form of $H_s\left(\beta_s^*\right)$, which in turn depends on the form of $f\left(z_{ij}'\hat{\beta}\right)$. In particular, it follows from (3) that

$$H_s\left(\beta_s^*\right) = \sum_{ij} z_{ij}z_{ij}'\exp\left(z_{ij}'\beta_s^*\right) f\left(z_{ij}'\beta_s^*\right) - \sum_{ij} z_{ij}\left(y_{ijs} - \exp\left(z_{ij}'\beta_s^*\right)\right)\frac{\partial f\left(z_{ij}'\beta_s^*\right)}{\partial \beta_s^*}. \tag{5}$$

Because the second term on the right-hand side of (5) tends to be small,[4] $H_s\left(\beta_s^*\right)$ is approximately a weighted sum of $\exp\left(z_{ij}'\beta_s^*\right) f\left(z_{ij}'\beta_s^*\right)$, with weights given by $z_{ij}z_{ij}'$. We have seen that when PPML is used, $\hat{\beta}^r$ can be interpreted as being approximately equal to the weighted average of the estimates of β_s with weights given by $\sum_{ij} y_{ijs}/\sum_{ijs} y_{ijs}$. Likewise, when NLS is used, $\hat{\beta}^r$ can be interpreted as being approximately the average

[4] Note that this term is zero in the case of PPML.

of the estimates of β_s weighted by the square of expectation of y_{ijs}, and for GPML $\hat{\beta}^r$ is approximately the average of $\hat{\beta}_s$.

Invoking the invariance result from Case 1, we conclude that the parameters estimated with aggregated data can also be interpreted as weighted averages of the estimates of the individual parameters, and that the particular weighted average that is estimated depends on the PML estimator that is used.

Case 3: Regressors vary with s but parameters do not

Our earlier results for this case are not specific to the PPML estimator, and therefore readily apply to the other estimators based on generalized linear models considered by Gourieroux, Monfort, and Trognon (1984). Therefore, for the NLS, PPML, and GPML estimators, ignoring that the regressors vary with s leads to an estimate whose relation to β depends on how the conditional moments of $\exp\left(z'_{ijs}\beta - z'_{ij}\beta\right)$ are related to z_{ij}. Again, when the ommited variable $\left(z'_{ijs} - z'_{ij}\right)\beta$ contributes little to the overall variance of y_{ijs}, it may be possible to predict the magnitude and sign of the differences between the elements of the two vectors when we have information on how the conditional moments of the omitted variable $\left(z'_{ijs}\beta - z'_{ij}\beta\right)$ vary with z_{ij}. Furthermore, the estimated parameters are 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 the estimates can be seen as an approximation to $\hat{\beta}$.

Case 4: Regressors and parameters vary with s

The case where both the regressors and the parameters vary with s can again 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 of parameters 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, with the weights depending on the particular PML estimator that is used.

References

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

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

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

Manning, Willard G. and John Mullahy (2001). "Estimating Log Models: To Transform or Not to Transform?," *Journal of Health Economics* 20, 461-494.

Santos Silva, João M.C. and Silvana Tenreyro (2006). "The Log of Gravity," *Review of Economics and Statistics* 88, 641-658.

Weidner, Martin and Thomas Zylkin (2021). "Bias and Consistency in Three-way Gravity Models," *Journal of International Economics* 132, 103513.

Appendix B:
Simulation evidence

Design

In this appendix we present the results of a simulation study illustrating the effects of aggregation in models estimated by PML. To simplify the design of the experiments, we consider a design without fixed effects where each observation only has two indices.

Following Santos Silva and Tenreyro (2011), for $i = 1, \ldots, n$ and $s = 1, \ldots, l$, we generate y_{is} as draws from a $\chi^2_{(m_{is})}$ distribution where, conditionally on the regressors z_{is}, m_{is} has a negative binomial distribution with $\mathrm{E}(m_{is}|z_{is}) = \exp(z'_{is}\beta_s)$ and $\mathrm{Var}(m_{is}|z_{is}) = a\mathrm{E}(m_{is}|z_{is}) + b\mathrm{E}(m_{is}|z_{is})^2$.

Estimation is performed with aggregate data defined as $y_i = \sum_{s=1}^{l} y_{is}$ and $z_i = \frac{1}{l}\sum_{s=1}^{l} z_{is}$. To keep the design close to that used in Santos Silva and Tenreyro (2011), we specify $\mathrm{E}(m_{is}|z_{is}) = \exp(0.4 + \beta_s z_{is})/l$ and set $a = 1$ and $b = 2$, which implies that neither PPML nor GPML are optimal estimators, and that in Case 1 the aggregate data have about 50 percent of zeros. Note that the distribution of the aggregate data does not depend on l, but the distribution of the disaggregated data does.

The way β_s and z_{is} are generated depends on the case we consider. For Case 1, we have $\beta_s = -1$ and $z_{is} = z_i \sim \mathcal{N}(0, 1)$. For Case 2, $\beta_s \sim \mathcal{N}(-1, \sigma_\beta)$ and z_{is} is obtained as in Case 1. For Case 3, $\beta_s = -1$ as in Case 1, but now $z_{is} = \mathcal{N}(z_i, \sigma(z_i))$, with $z_i \sim \mathcal{N}(0, 1)$ and $\sigma(z_i) = 1 - \sigma_z$ if $z_i < 0$ and $\sigma(z_i) = 1 + \sigma_z$ if $z_i > 0$. Therefore, in Case 3 the difference between z_{is} and z_i has mean zero and a variance that has a positive correlation with z_i for $\sigma_z > 0$. Finally, in Case 4 we have β_s as in Case 2 and z_{is} as in Case 3.

To complete the description of the data generating process, we need to set the values of n, l, σ_β, and σ_z. As in Santos Silva and Tenreyro (2011), we perform simulations

with $n \in \{1,000, 10,000\}$. The choice of l is not particularly important because the aggregation results do not depend on it. We, therefore, set $l = 60$, which is close to the number of 2-digit SITC sectors in our illustrative example; using a larger value of l should not substantially change the results but has a high computational cost. The choice of σ_z is more important because this parameter affects the magnitude of the bias in Case 3 and also affects the results in Case 4. We therefore perform experiments with $\sigma_z \in \{0.20, 0.40\}$. Finally, for the more interesting parameter σ_β, we choose values based on the dispersion of the estimates we obtained in our illustrative example (see Figure 1), and perform experiments with $\sigma_\beta \in \{0.40, 0.56\}$.

Results

The results of this simulation exercise are summarized in Table B1, which displays the average and standard deviation (in parentheses) of the PPML and GPML estimates of the coefficient on z_i obtained with $10,000$ replicas of the simulation procedure described above.[5] Additionally, the table also reports the weighted average $\sum_i y_{is}\beta_s / \sum_{is} y_{is}$, which provides a benchmark for the performance of the PPML estimator in Cases 2 and 4.

Overall, the results in Table B1 are in line with our theoretical results and in that sense they are not surprising. Nevertheless, the results are useful in that they provide information on the magnitude of the effects of aggregation.

The results for Case 1 confirm that both estimators identify the parameter of interest and that, in this particular design, they have comparable precision. For Case 2, as expected, we find that the GPML estimator identifies the average of sectoral parameters, whereas PPML identifies a parameter that is close to the weighted average of the sectoral

[5] We do not report results for NLS and for the least squares estimates based on the log-linearized model because these are too poor to be interesting.

parameters, with weights given by the sectoral shares of the dependent variable (trade, in the case of the gravity equation). Even with $\sigma_\beta = 0.56$, the difference between the two quantities is relatively small, confirming the findings in our illustrative application (see Figure 1).

The results for Case 3 show that, as in the example considered in Subsection 4.2.3, the estimates are biased upwards (towards zero) when the variance of the sectoral regressors

Table B1: Simulation results

	$n = 1,000$			$n = 10,000$		
	Gamma PML	Poisson PML	Weighted average	Gamma PML	Poisson PML	Weighted average
Case 1	−1.010 (0.090)	−0.994 (0.101)	−1	−1.001 (0.029)	−1.000 (0.035)	−1
Case 2						
σ_β= 0.40	−1.011 (0.101)	−1.143 (0.222)	−1.173 (0.088)	−1.002 (0.034)	−1.175 (0.113)	−1.185 (0.043)
σ_β= 0.56	−1.001 (0.119)	−1.275 (0.318)	−1.356 (0.167)	−0.997 (0.046)	−1.338 (0.163)	−1.395 (0.090)
Case 3						
σ_z= 0.20	−0.930 (0.092)	−0.940 (0.134)	−1	−0.924 (0.029)	−0.948 (0.048)	−1
σ_z= 0.40	−0.853 (0.090)	−0.883 (0.128)	−1	−0.847 (0.029)	−0.892 (0.045)	−1
Case 4						
σ_β= 0.40,σ_z= 0.20	−1.039 (0.134)	−1.127 (0.278)	−1.300 (0.112)	−1.041 (0.049)	−1.173 (0.124)	−1.323 (0.050)
σ_β= 0.56,σ_z= 0.20	−1.099 (0.166)	−1.202 (0.315)	−1.550 (0.161)	−1.102 (0.066)	−1.252 (0.127)	−1.595 (0.072)
σ_β= 0.40,σ_z= 0.40	−0.933 (0.130)	−1.055 (0.282)	−1.228 (0.105)	−0.935 (0.050)	−1.102 (0.131)	−1.301 (0.050)
σ_β= 0.56,σ_z= 0.40	−0.984 (0.165)	−1.151 (0.336)	−1.533 (0.166)	−0.986 (0.069)	−1.201 (0.143)	−1.575 (0.075)

increases with their average.[6] Interestingly, in the scenarios we considered, PPML is always less sensitive to this problem than GPML.

As in the theoretical results, the findings for Case 4 follow from the ones for the previous cases. In particular, in Case 4 both PPML and GPML estimate the quantity identified in Case 2 with a bias that is generally similar to the one observed in Case 3. Therefore, for GPML the results in Case 4 tend to be similar to those in Case 3, whereas for PPML the results for Case 4 are similar to those for Case 2.

Overall, these results are encouraging and can be summarized as follows. For both estimators, the effect of aggregation results only from the impossibility to account for sectoral variation in parameters and regressors. When there is sectoral variation in the parameters, both estimators identify different potentially interesting averages of the sectoral parameters. Sectoral variation in the regressors biases the estimates, but in some applications researchers may have information to determine at least the direction of the bias. Perhaps the more surprising result of these simulations is that in the scenarios we considered, both estimators are relatively robust to this problem. Finally, when both kinds of heterogeneity are present, the parameter identified in Case 2 is estimated with a bias similar to that observed in Case 3. Therefore, in this case we obtain a biased estimate of an average of the sectoral parameters. These findings provide guidance on the interpretation of estimates obtained using aggregate data and should be helpful to applied researchers who do not have disaggregated data at their disposal.

[6] We also performed some exploratory experiments where the variance of the sectoral regressors decreases with their average and found that in that case the estimates are biased downwards (away from zero).

References

Santos Silva, João M.C. and Silvana Tenreyro (2011). "Further Simulation Evidence on the Performance of the Poisson Pseudo-Maximum Likelihood Estimator," *Economics Letters* 112, 220-222.

Appendix C:

Implications for gravity-based predictions: An application to free trade agreements

The gravity equation has been used extensively to evaluate the impact of free trade agreements (FTAs) on trade flows, which is the application that we used in Section 3 to illustrate how aggregation affects parameter estimates. A natural question to ask is whether aggregation also matters for predictions of the trade flow increases expected after the implementation of FTAs. That is, would a researcher who has access to trade data at the sector level reach the same conclusion as another researcher who only has country-level trade data available?

In trying to answer this question, we use the same data as in Section 3 and consider again two estimation methods (OLS and PPML), three levels of aggregation, and models that impose coefficient homogeneity or allow the estimates to vary at the sector level.[7] Note that there are three types of counterfactuals we can perform. First, we could ask by how much trade flows between existing FTA partners are higher because of the FTAs in place. Second, we might be interested in finding out by how much trade would be larger if countries without FTAs put such agreements in place. Third, we could consider the change in trade moving from a situation without FTAs to a situation with FTAs in place between all countries. Conceptually, the first counterfactual corresponds to the average treatment effect on the treated (ATT), the second captures the average treatment effect on the untreated (ATU), and the third captures the average treatment effect (ATE).[8]

[7] See the description of Figure 1 and Table 1 for details.

[8] Note, however, that we are interested in changes in total trade flows rather than the average change in bilateral flows. That is, if we allow for sectoral coefficient heterogeneity, estimates for sectors with more trade get more weight. We also note that we are only concerned with the direct trade cost effects, not the indirect general equilibrium effects that operate through price indices, income and expenditure.

Denoting by $x_{ijst,1}$ the value of trade for country pair ij in sector s at time t in the presence of an FTA, and by $x_{ijst,0}$ the same flow in the absence of an FTA, the relevant counterfactuals are easily computed from the estimates of β_1, β_2 and β_3 obtained either by estimating (9) by OLS or (10) by PPML. For example, trade among FTA partners is simply the observed trade flow, $x_{ijst,1,FTA_{ijt}=1} = \exp\left(\hat{\alpha}_{ist} + \hat{\alpha}_{jst} + \hat{\alpha}_{ijs} + \hat{\beta}_{1s} + \hat{\beta}_{2s} + \hat{\beta}_{3s}\right)\hat{\eta}_{ijst}$, where we let the estimated parameters vary with s and we have assumed that the agreement is fully phased in (for the purpose of this illustration, we drop all pairs for which there is an FTA that is not fully phased in).[9] The (counterfactual) trade flow between the partners in the absence of an agreement is then given by $x_{ijst,0,FTA_{ijt}=1} = \exp\left(\hat{\alpha}_{ist} + \hat{\alpha}_{jst} + \hat{\alpha}_{ijs}\right)\hat{\eta}_{ijst} = x_{ijst,1,FTA_{ijt}=1} \times \exp\left(-\hat{\beta}_{1s} - \hat{\beta}_{2s} - \hat{\beta}_{3s}\right)$, which can be computed using data on actual trade flows and the OLS or PPML estimates obtained at the disaggregated level. Likewise, current trade among non-FTA partners can be expressed as $x_{ijst,0,FTA_{ijt}=0} = \exp\left(\hat{\alpha}_{ist} + \hat{\alpha}_{jst} + \hat{\alpha}_{ijs}\right)\hat{\eta}_{ijst}$, and the (counterfactual) trade in the presence of an FTA would be $x_{ijst,1,FTA_{ijt}=0} = x_{ijst,0,FTA_{ijt}=0} \times \exp\left(\hat{\beta}_{1s} + \hat{\beta}_{2s} + \hat{\beta}_{3s}\right)$. Once we have computed these counterfactuals, we can calculate the implied percentage changes in trade flows as

$$\begin{aligned} ATT &= \frac{\sum_{ijst} x_{ijst,1,FTA_{ijt}=1}}{\sum_{ijst} x_{ijst,0,FTA_{ijt}=1}} - 1 \\ &= \frac{\sum_{ijst} x_{ijst,1,FTA_{ijt}=1}}{\sum_{ijst} x_{ijst,1,FTA_{ijt}=1} \times \exp\left(-\hat{\beta}_{1s} - \hat{\beta}_{2s} - \hat{\beta}_{3s}\right)} - 1, \end{aligned}$$

[9] Using standard Neyman–Rubin notation, the subscript FTA_{ijt} indicates whether or not countries i and j have an FTA in place at time t. Thus, $x_{ijst,1,FTA_{ijt}=1}$ is the trade flow with an FTA for country pair ij in sector s at time t, given that country pair ij has an FTA in place. Note that this is of course simply the observed trade flow. By contrast, $x_{ijst,0,FTA_{ijt}=1}$ is the trade flow for country pair ij in sector s at time t without an FTA, which is a counterfactual trade flow given there is currently an FTA in place.

$$
\begin{aligned}
ATU &= \frac{\sum_{ijst} x_{ijst,1,FTA_{ijt}=0}}{\sum_{ijst} x_{ijst,0,FTA_{ijt}=0}} - 1 \\
&= \frac{\sum_{ijst} x_{ijst,0,FTA_{ijt}=0} \times \exp\left(\hat{\beta}_{1s} + \hat{\beta}_{2s} + \hat{\beta}_{3s}\right)}{\sum_{ijst} x_{ijst,0,FTA_{ijt}=0}} - 1,
\end{aligned}
$$

and

$$
\begin{aligned}
ATE &= \frac{\sum_{ijst} x_{ijst,1}}{\sum_{ijst} x_{ijst,0}} - 1 = \frac{\sum_{ijst}\left[x_{ijst,1,FTA_{ijt}=1} + x_{ijst,1,FTA_{ijt}=0}\right]}{\sum_{ijst}\left[x_{ijst,0,FTA_{ijt}=1} + x_{ijst,0,FTA_{ijt}=0}\right]} - 1 \\
&= \frac{\sum_{ijst}\left[x_{ijst,1,FTA_{ijt}=1} + x_{ijst,0,FTA_{ijt}=0} \times \exp\left(\hat{\beta}_{1s} + \hat{\beta}_{2s} + \hat{\beta}_{3s}\right)\right]}{\sum_{ijst}\left[x_{ijst,1,FTA_{ijt}=1} \times \exp\left(-\hat{\beta}_{1s} - \hat{\beta}_{2s} - \hat{\beta}_{3s}\right) + x_{ijst,0,FTA_{ijt}=0}\right]} - 1,
\end{aligned}
$$

where the summations are over all country pairs ij, sectors s and time periods t in our data.[10]

Table C1 presents the results of this exercise. The first row of the table shows the predicted increases in trade flows when we estimate our FTA coefficients using country-level data (i.e., 0-digit). As there is no sector-level dimension, we have that ATT=ATU=ATE. We also have that ATT=ATU=ATE whenever we impose coefficient homogeneity. The reason is obvious on inspection of the relevant expressions above. Indeed, if the coefficient estimates do not vary by sector s, the trade flow terms in the numerator and denominator of the ATT and ATU cancel so that the estimated treatment effect is simply the

[10] Since we compute treatment effects as percentage changes, the above definition of the ATE yields the same results as the more traditional ATE definition in terms of the average effect of a treatment (here: the presence of an FTA) across the units in a population (here: all country pairs, sectors and time periods) when the effect is expressed relative to the average baseline trade flows without FTAs. To see this write

$$
ATE = \frac{\sum_{ijst}\left(x_{ijst,1} - x_{ijst,0}\right)}{\sum_{ijst} x_{ijst,0}} = \frac{\sum_{ijst} x_{ijst,1} - \sum_{ijst} x_{ijst,0}}{\sum_{ijst} x_{ijst,0}} = \frac{\sum_{ijst} x_{ijst,1}}{\sum_{ijst} x_{ijst,0}} - 1.
$$

exponential of the sum of the coefficients for both the ATT and ATU. Since the ATE is simply a weighted mean of the ATT and the ATU, it will also be equal to whatever value the ATT and ATU take.

Table C1: Estimated treatment effects at different aggregation levels

Aggregation level	Heterogeneous coefficients	Treatment effect type	Estimator	
			OLS	PPML
SITC 0-digit	No	ATT=ATU=ATE	104.2%	80.7%
SITC 2-digit	Yes	ATT	56.3%	66.2%
SITC 2-digit	Yes	ATU	61.7%	83.4%
SITC 2-digit	Yes	ATE	60.5%	79.8%
SITC 2-digit	No	ATT=ATU=ATE	71.9%	80.7%
SITC 4-digit	Yes	ATT	61.0%	57.6%
SITC 4-digit	Yes	ATU	92.0%	105.4%
SITC 4-digit	Yes	ATE	85.3%	95.0%
SITC 4-digit	No	ATT=ATU=ATE	61.8%	80.7%

Notes: The table shows the predicted effect of FTAs at the 0-digit, 2-digit and 4-digit levels of aggregation. ATT is average treatment effect on the treated, ATU is average treatment effect on the untreated, ATE is average treatment effect. See text for details.

After these preliminary observations, we now move on to the more interesting comparison of how predicted trade flow increases vary with the level of aggregation and the underlying estimation method. Consistent with our results from Case 1, which demonstrated the invariance of PPML estimates when coefficient estimates do not vary at the sector level, Table C1 shows that the predicted trade flow increase under PPML with

homogeneous estimates is the same regardless of whether we use aggregate, 2-digit or 4-digit data (it is always 80.7%). However, the same is not true for predictions based on OLS estimates, even if we impose coefficient homogeneity. Thus, using country-level data instead of 4-digit sector-level data can lead to substantially different predictions regarding the trade effects of FTAs when based on traditional OLS estimation.

As expected from our results for Case 2, however, aggregation matters even with PPML when the underlying sector-level elasticities are heterogeneous. Looking at the results in Table C1, when we use 4-digit data we estimate an ATE of 95.0%. When we instead use 2-digit data (allowing coefficient estimates to vary at that level), the estimated ATE is 79.8%. When we aggregate up further to bilateral trade at the country level, we obtain an ATE of 80.7% as mentioned previously.[11] The corresponding results for OLS estimation are considerably more heterogeneous. This variability reflects the fact that OLS combines the aggregation bias and the bias resulting from log-linearization, and these biases can partially offset or compound each other.

[11] Note that with aggregate bilateral trade, there is of course no sector dimension and so we cannot allow for sector heterogeneity in our FTA estimates. Accordingly, Table C1 only reports results without coefficient heterogeneity at the aggregate (0-digit) level.