UNSW Business School Research Paper No. 2017 ECON 03

Comparative Advantage and Biased Gravity

Scott French

Comparative Advantage and Biased Gravity

Scott French*

January, 2017

Abstract

Gravity estimation based on sector-level trade data is generally misspecified because it ignores the role of product-level comparative advantage in shaping the effects of trade barriers on sector-level trade flows. Using a model that allows for arbitrary patterns of product-level comparative advantage, I show that sector-level trade flows follow a generalized gravity equation that contains an unobservable, bilateral component that is correlated with trade costs and omitted by standard sector-level gravity models. I propose and implement an estimator that uses product-level data to account for patterns of comparative advantage and find the bias in sector-level estimates to be significant. I also find that, when controlling for product-level comparative advantage, estimates are much more robust to distributional assumptions, suggesting that remaining biases due to heteroskedasticity and sample selection are less severe than previously thought.

JEL Classification: F10, F14, C13, C21, C50

Keywords: international trade, product-level, misspecification, heteroskedasticity, multi-sector

*School of Economics, University of New South Wales. scott.french@unsw.edu.au. scott.french@unsw.edu.au. Some of the content of this paper has been adapted from an unpublished paper previously circulated under the title "The Composition of Exports and Gravity". I am thankful for helpful comments from James Morley and Seojeong Lee.

1 Introduction

The gravity model has long been celebrated as a parsimonious yet empirically successful way to characterize bilateral trade flows. It is also useful for estimating the effect of factors that determine bilateral trade flows. Since Anderson (1979) showed that the gravity equation is theoretically founded, it has also been used to parameterize trade models, allowing general equilibrium analysis of the effects of these factors on economic outcomes and welfare.

In this paper, I show that standard gravity estimation using highly aggregated bilateral trade flows is misspecified in the presence of comparative advantage across products because it ignores how this product-level comparative advantage shapes the effects of trade barriers on sector-level trade flows. I demonstrate this using a simple model of international trade that allows for arbitrary patterns of product-level comparative advantage. The model is consistent with generalizations of a wide class of quantitative trade models and imposes minimal structure on the form of demand across products, details of factor markets, or the sources of comparative advantage. The key insight of this model is that trade barriers have weaker effects on trade flows between county pairs with relatively strong patterns of comparative advantage. This effect is embodied in an additional term that appears in a generalized sector-level gravity equation. In general, this term varies across country pairs and is a function of the trade barriers faced by all countries interacted with countries' patterns of comparative advantage. As a result, failing to control for the effects of product-level comparative advantage will cause gravity estimation using sector-level trade flow data to suffer from omitted variable bias. Unfortunately, this effect is not observable using sector-level data and, because it is country-pair-specific, it cannot be absorbed into country-specific fixed effects as is commonly done with other endogenous and unobservable variables in gravity estimation.

Because bilateral trade barriers cannot be inferred from sector-level trade data, I propose a method to estimate trade barriers using pooled product-level trade data. The estimation specification based on my model uses its product-level gravity structure to overcome practical issues related to the lack of available product-level data on domestic trade flows. I implement this estimator using data on bilateral product-level trade flows for 130 countries and 4,608 manufactured products. The results indicate that sector-level trade flows depend significantly on product-level comparative advantage and that coefficient estimates based on sector-level data are significantly biased. This bias is also economically significant. For example, the estimated distance elasticity differs by between 8% and 69% across estimations that do and do not control for product-level comparative advantage. For estimation based on log-linear OLS with importer and exporter fixed effects, controlling for product-level comparative advantage reduces the median estimated ad valorem tariff equivalent trade cost from 2,250% to 741%.[1]

I also find that estimates based on product-level data are remarkably consistent, and statistically indistinguishable, across estimators that make different assumptions on the properties of the error term and use different methods to control for unobserved country-specific effects. This is in contrast to sector-level estimates, which differ markedly across estimators, both in the application in this paper and in the literature as a whole. Based on the theoretical insights and Monte Carlo experiments of Santos Silva and Tenreyro (2006), these differences are typically attributed to bias in log-linear OLS estimates, due to sample selection in the presence of zero-valued trade flows and heteroskedasticity interacting with the log transformation, and finite-sample bias in pseudo-maximum likelihood (PML) estimators. However, my results indicate that differences in parameter estimates based on sector-level gravity estimation are, in fact, largely due to a failure to control for product-level comparative advantage.

Another finding is that border costs estimated using product-level data are more highly (negatively) correlated with income per worker than are sector-level estimates. This is important because it indicates that reductions in border-related trade barriers, which are likely to be policy-related, will disproportionately benefit low-income countries to a greater extent than sector-level estimates would indicate. In addition, I use the model and coefficient estimates to calculate the trade impact of eliminating these border costs. The product-level model predicts substantially smaller changes in trade flows than a standard sector-level model, and the product-level model predictions vary much less with the estimator used to estimate the trade cost coefficients. I show that the trade impact measure employed – labeled the Modular Trade Impact (MTI) by Head and Mayer (2014) – can be calculated without adding any more structure to the model than is needed to derive the product-level estimating equation. This makes the MTI a very useful tool for evaluating the effects of changes in trade costs estimated within such a framework.

For simplicity, my baseline estimation treats manufacturing as a single sector and estimates a parsimonious empirical specification. However, I extend the empirical analysis along each dimension by allowing for parameter heterogeneity across multiple manufacturing sectors, defined as 2-digit ISIC industries, and by including variables related to trade policy and historical country ties that are common in gravity estimation. I find evidence of significant heterogeneity across sectors, even after controlling for product-level comparative advantage. Yet, the main results are unaffected by allowing for sectoral heterogeneity. Consideration of additional covariates also does not significantly change the main findings, although there

[1] This assumes a value of the elasticity of trade flows with respect to trade costs equal to 6, a value between the mean and median estimate reported in the meta-analysis of Head and Mayer (2014).

is substantially more heterogeneity in the product-level estimates for some of the additional variables, in particular colonial ties and regional trade agreements. This may indicate that there is actual heterogeneity across country pairs in the effects of these variables.[2] Interestingly, sector- and product-level Poisson PML estimates for variables other than the primary determinants of trade costs – distance and border effects – are remarkably similar, which may indicate that sector-level Poisson PML is a reasonably robust estimator for the effects of variables of secondary importance, such as these.[3]

Taken together, the model and empirical results illustrate that it is crucial to control for the effects of product-level comparative advantage when analyzing the determinants of sector-level trade flows. The pooled product-level estimator that I propose makes use of data which are widely available and contain a wealth of information that is lost when the data are aggregated to the sector level. While the product-level estimator is more computationally demanding than a typical sector-level estimator, it is far from prohibitively so.[4] Thus, pooled product-level estimation using data at the lowest level of aggregation available is generally preferable in any setting where sector-level estimation would typically be employed.

This paper contributes to two strands of the international trade literature. The first is devoted to developing efficient and unbiased estimators of gravity models. Among many others, this literature includes Anderson and van Wincoop (2003), who demonstrated that estimation must control for endogenous factors embodied in the "multilateral resistance" terms of structural gravity models, and Santos Silva and Tenreyro (2006), who highlighted the pitfalls of estimation based on log-linear OLS and popularized the estimation of gravity equations in their multiplicative form using PML estimators. Head and Mayer (2014) provide a detailed review of this literature. My paper contributes to this literature by demonstrating that unbiased estimation of sector-level gravity equations requires controlling for product-level comparative advantage and by developing and applying an estimator that uses product-level data to do so.

My paper is also related to a fast-growing literature developing and exploring the properties of gravity models with sectoral heterogeneity. Costinot and Rodríguez-Clare (2014) and Kehoe et al. (2016) review recent advances in this area. These models have been used to study, for example, the effect of sectoral specialization on the incidence of trade costs (Anderson and Yotov, 2010), the degree heterogeneity in trade costs across sectors (Chen

[2] Baier et al. (2016) find empirical support for this conclusion in the case of trade agreements. In addition, these variables are likely endogenous.

[3] Distance and border costs account for more than 70% of the variation in estimated trade costs, based on product-level Poisson PML.

[4] For example, the baseline product-level Poisson PML estimation took about 15 minutes on a relatively standard desktop computer.

and Novy, 2011), the effect of sectoral productivity differences on inter-sectoral trade flows (Costinot et al., 2012), and the welfare implications of sectoral heterogeneity and changes in sectoral comparative advantage over time (Ossa, 2015, and Levchenko and Zhang, 2016). These models allow for heterogeneity across broadly-defined sectors. My paper is distinct in that it studies the effects of comparative advantage across narrowly-defined products within sectors – i.e., thousands of products versus at most a few dozen sectors. In another paper, French (2016), I show that accounting for product-level comparative advantage is important for predicting the welfare gains from trade. In this paper, by contrast, I show that patterns of comparative advantage at such a low level of aggregation are important for determining sector-level trade flows and that controlling for them is necessary for unbiased gravity estimation.

In the next section, I specify the model and demonstrate how product-level comparative advantage influences sector-level trade flows and causes sector-level gravity estimates to be biased. Section 3 develops my proposed product-level estimation procedure and presents the baseline empirical results. Section 4 presents extensions to multiple sectors and the inclusion of additional covariates. Section 5 concludes.

2 Theoretical Framework

Bilateral international trade flows obey gravity equations in a very large range of models.[5] In this section, I outline a general theoretical framework in which gravity equations for bilateral product-level trade flows arise and in which countries can have arbitrary patterns of comparative advantage across a potentially large number of products.

2.1 A Product-Level Gravity Model

The model economy consists of N countries, each of which contains buyers who demand goods from $j = 1, ..., J$ sectors. Each sector is made up of a finite number of product categories, $k = 1, ..., K^j$, and each category contains a continuum of product varieties $\omega \in \Omega^k$, of which a weak subset $\Omega_n^k \subseteq \Omega^k$ are available in n.[6] Thus, a particular product is identified by the pair (j, k), and a product variety is identified by the triple (j, k, ω).

Anderson and van Wincoop (2004) make clear that gravity-like structure arises in models in which the allocation of expenditure across product varieties can be analyzed separately

[5] See Anderson (2011) and Head and Mayer (2014) for a recent discussions of the set of theoretical models that imply gravity equations.

[6] The assumption of a continuum of varieties is purely for analytical convenience. If the number of varieties per product category were finite, the results below would hold in expectation.

from the allocation of production and consumption across sectors within countries, which they term trade separability. In the present framework, I introduce this simplification by assuming that buyers take as given sector-level output and expenditure, $\{Y_i^j, X_i^j\}$, when allocating expenditure across product categories and varieties.

I further assume that buyers in each country maximize an objective function that is separable across products, which implies that the allocation of expenditure across product varieties can be analyzed separately from the allocation of sector-level expenditure across products. Formally, this is identical to assuming that expenditure on variety (j, k, ω) by buyers in n is given by

$$x_n^{jk}(\omega) = f_n^{jk}(\omega, \mathbf{p}_n^{jk}, X_n^{jk}),$$

where $\mathbf{p}_n^{jk}$ is the set of prices of varieties of product (j, k), and X_n^{jk} represents total expenditure on all varieties of product (j, k) by consumers in n.

Given this basic setup, there are a wide variety of assumptions regarding technologies and market structure that imply that product-level trade flows follow a gravity equation. Without relying on a particular set of micro-foundations, I make the following assumption, analogous to one of the macro-level restrictions of Arkolakis et al. (2012), regarding the product-level import demand system:

Assumption 1 (Product-Level Gravity)**.** *Expenditure by buyers in n on all varieties of product (j, k) from i is given by*

$$X_{ni}^{jk} = \frac{T_i^{jk}(d_{ni}^{jk})^{-\theta^j}}{\Phi_n^{jk}} X_n^{jk}, \tag{1}$$

where $\Phi_n^{jk} = \sum_i T_i^{jk}(d_{ni}^{jk})^{-\theta^j}$.

The key condition imposed by equation (1) is that the share of product-level expenditure by destination n on varieties from i can be decomposed multiplicatively into an exporter-product-specific component, an importer-product-specific component, and a trade cost effect. The term T_i^{jk} includes all factors that affect country i's overall ability to supply product (j, k), including both exogenous and endogenous variables – for example, product-level productivity, factor prices, and the mass of firms that produce varieties of (j, k). Trade barriers, $d_{ni}^{jk} > 1$, take the standard form of "iceberg" trade costs, as in Samuelson (1954), and θ^j is the elasticity of product-level trade flows to trade costs in sector j. The term Φ_n^{jk} is an index of all exporters' abilities to deliver product (j, k) to destination n, which also serves to ensure that product-level trade flows sum across exporters to product-level expenditure.

Summing over n to obtain product-level output, denoted Y_i^{jk}, and substituting back into

(1) yields a product-level version of the structural gravity equation in its most common form:

$$X_{ni}^{k} = \frac{Y_i^{jk}}{\Psi_i^{jk}} \frac{X_n^{jk}}{\Phi_n^{jk}} (d_{ni}^{jk})^{-\theta}, \tag{2}$$

where

$$\Psi_i^{jk} = \sum_n \frac{X_n^{jk}}{\Phi_n^{jk}} (d_{ni}^{jk})^{-\theta} \quad \text{and} \quad \Phi_n^{jk} = \sum_i \frac{Y_i^{k}}{\Psi_i^{jk}} (d_{ni}^{jk})^{-\theta}, \tag{3}$$

and Ψ_i^{jk} and Φ_n^{jk} are the "multilateral resistance" terms defined by Anderson and van Wincoop (2003).

Assumption 1 does not necessarily require further restrictions on the form of buyers' objective functions.[7] However, such restrictions will be useful in obtaining certain analytical results. For this purpose, where necessary, I follow most of the gravity literature and assume that buyers maximize an identical nested constant elasticity of substitution (CES) utility function, which implies the following demand system:

Assumption 1′ (CES Demand)**.** *Expenditure on variety (j, k, ω) by buyers in n is given by*

$$x_n^{jk}(\omega) = \left(\frac{p_n^{jk}(\omega)}{P_n^{jk}} \right)^{1-\eta^{jk}} X_n^{jk},$$

and

$$X_n^{jk} = \left(\frac{P_n^{jk}}{P_n^{j}} \right)^{1-\sigma^{j}} X_n^{j},$$

where $P_n^{jk} = \left(\int_{\omega \in \Omega^k} p_n^{jk}(\omega)^{1-\eta^{jk}} \right)^{\frac{1}{1-\eta^{jk}}}$, $P_n^{j} = \left(\sum_{k=1}^{K^j} (P_n^{jk})^{1-\sigma^j} \right)^{\frac{1}{1-\sigma^j}}$, *and* $\eta^{jk} > \sigma^j > 1$.

Finally, I make the following assumption regarding product-level bilateral trade costs:

Assumption 2 (Trade Costs)**.** *Iceberg trade costs take the following form:*

$$d_{ni}^{jk} = \begin{cases} d_{ni}^{j} d_n^{jk} d_i^{*jk} & \text{if } n \neq i \\ d_{nn}^{j} & \text{if } n = i \end{cases}$$

This restriction on the form of d_{ni}^{jk} implies that product-level trade costs can be decomposed into a bilateral sector-specific component and importer- and exporter-product-specific

[7] Different micro structures require different functional form restrictions. For example, the Armington model of Anderson and van Wincoop (2003) requires consumers to have CES preferences, while French (2015) and Arkolakis et al. (2015) provide examples of gravity models with perfect competition and monopolistic competition, respectively, that relax the assumption of CES preferences.

border costs.[8] This assumption greatly simplifies the analysis that follows, while still allowing a large degree of flexibility in d_{ni}^{jk}, and is much less the restrictive than the form of trade barriers implicitly assumed by most sector-level gravity models.[9] Assumption 2 is also consistent with import tariffs and non-tariff barriers that obey the Most Favored Nation principle of the WTO.

2.2 Sector-Level Trade Flows

In what follows, every variable is sector-specific. To avoid excessive notation, in the remainder of the paper, wherever there is no ambiguity, I omit the sector superscript, j and refer to product k in place of references to product (j, k).

Aggregating product-level trade flows and imposing sector-level market clearing conditions, it is possible to derive an expression relating total sector-level trade flows to countries' total output and expenditure and bilateral trade costs.

Proposition 1 (Sector-Level Gravity)**.** *Given Assumptions 1 and 2, sector-level trade flows are given by*

$$X_{ni} = \frac{X_n}{\Phi_n}\frac{Y_i}{\Psi_i}d_{ni}^{-\theta}\tilde{T}_{ni} \tag{4}$$

where Y_i is sector-level output in i, and Φ_n and Ψ_i are defined by the system of equations

$$\Psi_i = \sum_n \frac{X_n}{\Phi_n}d_{ni}^{-\theta}\tilde{T}_{ni} \quad \text{and} \quad \Phi_n = \sum_i \frac{Y_i}{\Psi_i}d_{ni}^{-\theta}\tilde{T}_{ni}, \tag{5}$$

where $\tilde{T}_{ni}$ is given by

$$\tilde{T}_{ni} = \sum_k \left(d_n^k d_i^{*k}\right)^{-\theta}\frac{X_n^k}{X_n}\frac{\Phi_n}{\Phi_n^k}\frac{T_i^k}{\bar{T}_i}, \quad \forall n \neq i, \tag{6}$$

and where $\bar{T}_i$ is a sector-level index of T_i^k that is homogeneous of degree 1, and $\tilde{T}_{ii} = \sum_k \frac{X_n^k}{X_n}\frac{\Phi_n}{\Phi_n^k}\frac{T_i^k}{\bar{T}_i}$, for all i.

Proofs of all propositions are given in the Appendix. Equation (4) is very nearly a standard sector-level gravity equation, except for the presence of the term $\tilde{T}_{ni}$. This term summarizes the effect of countries' patterns of product-level comparative advantage (and border costs) on sector-level trade flows. To understand how $\tilde{T}_{ni}$ embodies this effect, note that its value is closely related to the covariance of T_i^k and $(\Phi_n^k)^{-1}$. Recall that Φ_n^k is an index of all producers' ability to provide product k to n. The ratio $(T_i^k/\bar{T}_i)/(\Phi_n^k/\Phi_n)$ can

[8] Exporter-specific border costs are distinguished by an asterisk.

[9] This specification also does not rule out the possibility that trade barriers are asymmetric (i.e., $d_{ni}^j \neq d_{in}^j$) and does not require that domestic trade be frictionless (i.e., $d_{nn}^j \geq 1$).

be seen a measure of i's comparative advantage for product k in destination n vis-à-vis all possible sources. In this context, $\tilde{T}_{ni}$ is an index of the overall strength of i's comparative in n, across all products, in that $\tilde{T}_{ni}$ is relatively large if i tends to be relative proficient at producing the products which n has limited opportunities to purchase from other sources.[10]

2.3 Biased Gravity

As is standard in the gravity literature, I specify d_{ni} as a parametric function of observable variables:

$$d_{ni}^{-\theta} = g(\boldsymbol{Z}_{ni}, \boldsymbol{\beta}).$$

Typically, $\boldsymbol{\beta}$ is estimated using data on sector level trade flows, controlling for the country-specific endogenous variables using fixed effects or by imposing the structure of the model. However, the presence of $\tilde{T}_{ni}$ in (4) presents a fundamental problem for such sector-level gravity estimation. Failing to control for $\tilde{T}_{ni}$ causes sector-level gravity estimators to be misspecified in the presence of product-level comparative advantage, which implies that estimates of $\boldsymbol{\beta}$ will suffer from a form of omitted variable bias. Further, $\tilde{T}_{ni}$ is unobservable, so it cannot be controlled for directly, and it varies by country pair, so, unlike Φ_n and Ψ_i, it cannot be controlled for with fixed effects.

The severity and direction of this bias depends on the correlation in the data between $\boldsymbol{Z}_{ni}$ and $\tilde{T}_{ni}$. Because $\tilde{T}_{ni}$ is a function of bilateral trade costs, through the Φ_n^k terms, $\tilde{T}_{ni}$ and d_{ni} are likely to be strongly related and the bias severe. To gain some intuition for how the omission of $\tilde{T}_{ni}$ is likely to bias sector-level gravity estimates, it is helpful to consider a few implications of the model. First, consider the special case in which there is no product-level comparative advantage.

Proposition 2 (No Comparative Advantage)**.** *Given assumptions 1 and 2, if* $T_i^k = T_i T^k$ *and* $d_i^{*k} = \bar{d}_i^*$*, for all* i *and* k*, and* $d_n^k = \bar{d}_n$*, for all* n *and* k*, then* $\tilde{T}_{ni} = 1$*, for all* n *and* i*, and sectoral trade flows are given by*

$$X_{ni} = \frac{X_n}{\Phi_n} \frac{Y_i}{\Psi_i} d_{ni}^{-\theta}.$$

Proposition 2 demonstrates that, in the special case in which there is no product-level comparative advantage for any country, (4) reduces to a standard structural gravity equation, and $\boldsymbol{\beta}$ can be estimated consistently using sector-level data and standard techniques. In

[10]Because $\tilde{T}_{ni}$ weights the ratio $(T_i^k/\bar{T}_i)/(\Phi_n^k/\Phi_n)$ by endogenous expenditure shares, this statement implicitly assumes that the elasticity of X_n^k with respect to Φ_n^k is not greater than one. Under Assumption 1′, this is equivalent to assuming that $\theta > \sigma - 1$ – i.e., that the effective elasticity of substitution across sources of a product is greater than the elasticity of substitution across products, a natural assumption.

general, however, the forces of comparative advantage interact with trade barriers to shape sector-level trade flows.

To gain more insight into how estimation of $\boldsymbol{\beta}$ will be biased by omitting $\tilde{T}_{ni}$, it is helpful to place additional structure on buyers' objective function. Therefore, I impose Assumption $1'$ to derive the following two results.

Proposition 3 (Free Trade)**.** *Given assumptions 1 and $1'$, if $d_{ni}^k = 1$, for all n, i, and k, then $\tilde{T}_{ni} = \tilde{T}_i$, for all n and i, and sectoral trade flows are given by*

$$X_{ni} = \frac{X_n}{\Phi_n}\frac{Y_i}{\Psi_i'},$$

where $\Psi_i' = \Psi_i/\tilde{T}_i$.[11]

Proposition 3 shows that, even in the presence of product-level comparative advantage, sector-level trade flows still obey a standard gravity equation if trade is frictionless. This serves to demonstrate that it is the interaction of trade barriers and patterns of comparative advantage that causes bilateral trade flows to deviate from a sector-level gravity equation. It is worth noting that patterns of comparative advantage still influence the volume of trade flows under free trade. Trade flows follow a sector-level gravity equation because, under free trade, prices are equalized in every destination, so each source country's patterns of comparative advantage shift its sales by the same proportion in every destination.

The final result demonstrates how trade barriers and product-level comparative advantage interact to shape sector-level trade flows outside of these special cases.

Proposition 4 (Trade Elasticity)**.** *Given assumptions 1 and $1'$, changes in $\tilde{T}_{ni}$ associated with changes in bilateral trade costs are given by*

$$\mathrm{d}\ln(\tilde{T}_{ni}) = [\theta - (\sigma - 1)] \sum_l \mathrm{d}\ln(d_{nl}) \sum_k \frac{X_n^k}{X_{ni}} \left(\frac{X_{ni}^k}{X_n^k} - \frac{X_{ni}}{X_n}\right)\left(\frac{X_{nl}^k}{X_n^k} - \frac{X_{nl}}{X_n}\right), \tag{7}$$

holding constant all values of X_n, T_i^k, and d_n^k.

To interpret this expression, note that the values in parentheses can be thought of as measures of revealed comparative advantage (RCA) for countries i and l, respectively, in market n. The summation over k is proportional to the weighted covariance of these product-level RCA measures. This means that an increase in d_{nl} will disproportionately shift n's sector-level expenditure toward i if l and i have highly correlated patterns of comparative advantage.

[11] Assumption $1'$ is sufficient but not strictly necessary for Proposition 3 to hold. All that is required is that product-level expenditure shares are equal in every destination if all relative prices are equal, which will be the case, for example, if preferences are identical and homothetic.

In addition, the partial elasticity of $\tilde{T}_{ni}$ with respect to d_{ni} is given by

$$\frac{\partial \ln(\tilde{T}_{ni})}{\partial \ln(d_{ni})} = [\theta - (\sigma - 1)] \sum_k \frac{X_n^k}{X_{ni}} \left(\frac{X_{ni}^k}{X_n^k} - \frac{X_{ni}}{X_n} \right)^2 . \tag{8}$$

The direct effect of an increase in d_{ni}, through (4), is to reduce X_{ni}. Equation (8) demonstrates that this effect is weakened in proportion to the strength of i's product-level comparative advantage in n, measured by the weighted variance of its RCA values.[12]

What does this imply for the bias in estimates of $\boldsymbol{\beta}$ based on a sector-level gravity equation? Clearly, the bias will be more severe the stronger are countries' patterns of comparative advantage. It is more difficult to anticipate the sign of the bias. If the bilateral effect of trade costs on $\tilde{T}_{ni}$, given by (8), dominates the multilateral effects, given by (7), then $\tilde{T}_{ni}$ must be positively correlated with d_{ni} after controlling for exporter fixed effects, because $\tilde{T}_{ni} = \tilde{T}_i$ under free trade. Then, because X_{ni} is mechanically increasing in $\tilde{T}_{ni}$, omission of $\tilde{T}_{ni}$ will tend to bias estimates of $\boldsymbol{\beta}$ toward zero.

3 Gravity Estimation with Product-Level Comparative Advantage

I have shown theoretically that sector-level gravity estimation that ignores the effects of product-level comparative advantage is likely to yield biased estimates. To assess severity of this bias in practice, I propose and implement a method for estimating trade costs which is not susceptible to this problem. For the purpose of estimation, I assume that $g(\boldsymbol{Z}_{ni}, \boldsymbol{\beta}) = e^{\boldsymbol{Z}'_{ni}\boldsymbol{\beta}}$, consistent with most gravity estimations in the literature. I also assume that $d_i^{*k} = 1$, for all i and k, meaning that all border costs are importer-specific, consistent with, for example, import tariffs.[13] Together with Assumption 2, this implies that

$$-\theta \ln(d_{ni}^k) = -\theta \ln(d_n^k) \times \mathbb{1}_{\text{bord}} + \boldsymbol{Z}'_{ni}\boldsymbol{\beta},$$

where $\mathbb{1}_{\text{bord}}$ is an indicator function, which is equal to 1 when $n \neq i$.

Sector-level gravity estimation implicity assumes that product-level patterns of compara-

[12] This elasticity is always positive under the natural assumption that $\theta > \sigma - 1$ – i.e., that the effective elasticity of substitution across sources of a product is greater than the elasticity of substitution across products.

[13] The estimation and analyses of this paper are isomorphic to the specification of border costs as importer-specific, exporter-specific, or some combination. While Waugh (2010) argues that exporter-specific border costs are more consistent with international data on prices of tradeable goods, Ramondo et al. (2016) argue that domestic trade frictions account for much of this phenomenon.

tive advantage take a very special form, such as that imposed by the conditions of Proposition 2. Instead, I propose an estimator that imposes only Assumptions 1 and 2. This approach takes advantage of the product-level gravity structure of (1) and uses data on product-level trade flows, which are now widely available for most countries.

3.1 Estimation Methods

There are several options for estimating $\boldsymbol{\beta}$ based on (1).[14] As with sector-level gravity estimation, there are two basic approaches to controlling for the unobserved components of (1). The first is to take advantage of the log-linear form of (1) and control for these variables with country-by-product fixed effects. The second is to impose the model's market clearing conditions and use a structural estimator. In the context of the product-level gravity model, both of these approaches rely on the following result:

Proposition 5. *Given assumptions 1 and 2, product-level trade flows are given by the following system of equations:*

$$X_{ni}^k = \frac{E_i^k}{\tilde{\Psi}_i^k} \frac{M_n^k}{\tilde{\Phi}_n^k} d_{ni}^{-\theta}, \tag{9}$$

where

$$\tilde{\Psi}_i^k = \sum_{n \neq i} \frac{M_n^k}{\tilde{\Phi}_n^k} d_{ni}^{-\theta} \quad \text{and} \quad \tilde{\Phi}_n^k = \sum_{i \neq n} \frac{E_i^k}{\tilde{\Psi}_i^k} d_{ni}^{-\theta}, \tag{10}$$

and where E_i^k is total exports of k by i, and M_n^k is total imports of k by n.

Equation (9) is a slight variation on (3), which is more closely analogous to a standard sector-level structural gravity equation. Equation (9) has two particularly important features that are useful for estimation. First, it is specified in terms of total product-level exports and imports, rather than total production and expenditure. This is very useful because data on production, expenditure, and domestic trade flows are typically not available at a level of disaggregation comparable to international trade flow data. This specification implies that highly disaggregated trade data can be used to estimate $\boldsymbol{\beta}$ without comparable domestic data. Second, it is not necessary to identify the values of d_n^k in order to identify $\boldsymbol{\beta}$, which is important because, without data on domestic trade flows, it is not possible to identify these border costs at the product level. However, I show below that it is possible to identify a sector-level index of these parameters if domestic data are available at higher levels of aggregation.

[14] All of these are based on an assumption that some form of (1) holds in expectation. This implicitly assumes that deviations from (1) result from measurement error or exogenous shocks to bilateral trade flows. Egger and Nigai (2015) show that it is problematic to relegate unobserved trade costs to the error term, in contrast to the argument of Anderson and van Wincoop (2003) that resulting biases were likely to be small.

Both the fixed-effects (FE) and structural approaches use pooled product-level data. It is also possible to apply both approaches product-by-product. In addition, there are several valid choices for the objective function of the estimation. The first generation of sector-level theoretically-consistent gravity estimations used a structural nonlinear least squares (NLS) or a FE OLS estimator based on the logged form of (1).[15] More recently, pseudo-maximum likelihood (PML) estimation based on the multiplicative form of (1) has become more common due to the finding of Santos Silva and Tenreyro (2006) that the log transformation can lead to inconsistent estimates if the error term is heteroskedastic. Taking logs also requires dropping zero-valued trade flows, which are extremely prevalent in product-level trade data, leading to potential sample selection bias.[16] The family of PML estimators, on the other hand, are consistent as long as the conditional expectation of X_{ni}^k is given by (1) and allow the estimation to include zeros. The only difference between these estimators is how they weight observations based on the assumed form of heteroskedasticity. To see this, note that the gamma PML, Poisson PML, and Gaussian PML (NLS in levels) respectively impose the following moment conditions:

$$\sum(X_{ni}^k - \hat{X}_{ni}^k)\boldsymbol{Z}_{ni}/\hat{X}_{ni}^k = 0, \quad \sum(X_{ni}^k - \hat{X}_{ni}^k)\boldsymbol{Z}_{ni} = 0, \quad \sum(X_{ni}^k - \hat{X}_{ni}^k)\hat{X}_{ni}^k\boldsymbol{Z}_{ni} = 0.$$

For comparison, the moment conditions for log-linear least squares (log LS) are given by $\sum[\ln(X_{ni}^k) - \ln(\hat{X}_{ni}^k)]\boldsymbol{Z}_{ni} = 0$. Poisson PML is preferred by Santos Silva and Tenreyro (2006) due to its performance in Monte Carlo experiments. However, I also employ these other PML estimators for comparison.[17]

The relative merits of the FE and structural approaches depend on several practical considerations. It is tempting to favor FE estimators based on the argument that, because they impose less structure, they are more robust to model misspecification. However, the structural approach actually imposes no more structure on the estimation. Other than the functional form of (1), the structural approach imposes only the adding-up constraints that $\sum_{i\neq n}\hat{X}_{ni}^k = M_n^k$ and $\sum_{n\neq i}\hat{X}_{ni}^k = E_i^k$, where hats indicate fitted values. Fally (2015) shows that the FE estimators implicitly impose similar adding-up constraints, meaning that the only difference between the two approaches is in the functional form of the adding-up

[15] Anderson and van Wincoop (2003) and Eaton and Kortum (2002) are seminal examples, respectively, of these methods.

[16] Eaton and Tamura (1994), Hallak (2006), and Helpman et al. (2008) propose estimators that deal with the issue of sample selection bias while maintaining the log-linear regression approach. However, in Monte Carlo experiments, Santos Silva and Tenreyro (2006) find that such methods perform comparatively poorly.

[17] Another popular PML estimator is the negative binomial, which nests Poisson PML as a special case. However, Head and Mayer (2014) summarize several compelling arguments against using this estimator, including that estimates depend on the unit of measurement of the dependent variable.

constraints.[18] In the case of Poisson PML, the structural and FE approaches coincide.[19] Product-by-product estimation is considerably more flexible than pooled estimation, as it does not impose Assumption 2 or that θ is constant across products within a sector. However, it will be quite inefficient if these conditions are reasonably close to holding in the data.

The primary practical issue is the shear size of pooled product-level trade datasets when using trade flows at the lowest levels of aggregation available, as the model implies is appropriate. This creates particular problems for the FE approach. Standard practice is to control for the fixed effects using dummy variables. In a pooled product-level estimation, this requires including $2K(N-1)$ importer-product and exporter-product dummy variables, which quickly becomes infeasible for large samples.[20] In the log-linear OLS case, the within estimator can be computed by numerically solving the least squares normal equations to remove the country-by-product fixed effects from all other variables and then running the regression using the residuals, in accordance with the Frisch-Waugh-Lovell theorem.[21] This method is computationally intensive but typically feasible. In the PML cases, creating the matrix of dummy variables may be avoidable by numerically solving the system of equations implied by the first-order conditions of the likelihood function. However, this can also be a very computationally intensive task. The structural approach can be computationally intensive, as well, but the system of equations (10) is known to have a unique solution and efficient algorithms to find it. Thus, in practice, the structural approach tends to be less computationally demanding than the FE approach for all specifications (other than Poisson PML).[22]

The FE approach has an additional deficiency in the non-Poisson PML specifications. The multitude of fixed effects creates an incidental parameters problem. The log-linear OLS and Poisson PML estimators are two special cases in which $\boldsymbol{\beta}$ is consistently identified even

[18] These constraints are $\sum_{i\neq n} \ln\left(\hat{X}_{ni}^k\right) = \sum_{i\neq n} \ln\left(X_{ni}^k\right)$ for log-linear OLS, $\sum_{i\neq n} X_{ni}^k \Big/ \hat{X}_{ni}^k = N$ for gamma PML, and $\sum_{i\neq n} \hat{X}_{ni}^k X_{ni}^k = \sum_{i\neq n} \left(X_{ni}^k\right)^2$ for NLS in levels, along with the analogous constraints for exports for each estimator.

[19] Fally (2015) and Arvis and Shepherd (2013) provide formal proofs of this equivalence.

[20] In the estimation below, the dataset contains trade flows among 130 countries in 4,208 product categories. Estimation using fixed effects would require 1,188,864 dummy variables, meaning that forming the matrix of independent variables (given $N(N-1)K$ trade flow observations) would require nearly 80 terabytes of computer memory.

[21] See, for example, the algorithm of Guimarães and Portugal (2010).

[22] Fally (2015) provides a proof of uniqueness of the solution to (10). Gourieroux et al. (1984) show that the Poisson PML likelihood function has a unique maximum and globally negative semidefinite Hessian. Alvarez and Lucas (2007) and Allen et al. (2014) provide general proofs of existence and uniqueness of equilibria of trade models that imply that trade flows take the form of (9)-(10).

when the fixed effects are not.[23] Due to the bilateral nature of trade data, the number of observations for each country-product pair increases with N, meaning that this is not a problem if N grows toward infinity. However, the identification of the fixed effects requires non-zero observations within each country-product group, so the prevalence of zeros in product-level trade data means that the incidental parameters problem could be a significant issue for these PML estimators in finite samples even for a relatively large N.[24]

For these reasons, pooled product-level Poisson PML is my preferred estimator. However, for comparison, I also estimate both structural and FE log LS and structural gamma PML and NLS specifications as well as a product-by-product specification and specifications that define a sector at different levels of aggregation.

3.2 Estimating Border Costs

While (9) shows that identification of $\boldsymbol{\beta}$ does not require information on the values d_n^k, estimates of border costs are of interest in many contexts. If data on X_{nn}^k are available, it is straightforward to recover values of d_n^k. Based on (9)-(10), the predicted value of X_{nn}^k is given by

$$\hat{X}_{nn}^k = (\hat{d}_n^k)^{\theta} \frac{M_n^k}{\hat{\tilde{\Phi}}_n^k} \frac{E_i^k}{\hat{\tilde{\Psi}}_i^k} \hat{d}_{nn}^{-\theta},$$

where $\hat{\tilde{\Phi}}_n^k$ and $\hat{\tilde{\Psi}}_i^k$ solve (10) given $\hat{\boldsymbol{\beta}}$ and values of M_n^k and E_i^k from the data. Thus, a valid estimate of $(d_n^k)^{-\theta}$ is

$$(\hat{d}_n^k)^{-\theta} = \frac{\hat{d}_{nn}^{-\theta}}{X_{nn}^k} \frac{M_n^k}{\hat{\tilde{\Phi}}_n^k} \frac{E_i^k}{\hat{\tilde{\Psi}}_i^k},$$

Data on X_{nn}^k are not typically available, meaning that identification of d_n^k is not possible in most cases. However, sector-level domestic trade flow data often are available, in which case it is possible to calculate a sector-level index of d_n^k. In particular, let $\bar{d}_n$ denote the uniform importer-specific border cost for which sector-level domestic trade flows equal their values in data, holding fixed product-level imports and exports.[25] Because the values of $\tilde{\Phi}_n^k$

[23] Fernández-Val and Weidner (2016) demonstrates this for the specific case of Poisson PML with two-way fixed effects under the assumption that the regressors are strictly exogenous. Also, due to its equivalence with the structural estimator, FE Poisson PML will be unbiased as long as the structural equations (10) are correctly specified.

[24] For the estimation below, $N = 130$, so there are 129 observations per country-by-product effect, but 95% of the observations are zero.

[25] This index is closely related to the ideal summary index developed by Anderson and Neary (2003). See also Anderson and van Wincoop (2004) for a review of the literature on aggregation issues regarding trade costs.

and $\tilde{\Psi}_i^k$ do not depend on d_n^k, this index can be easily calculated as

$$\hat{\bar{d}}_n^{-\theta} = \frac{\hat{d}_{nn}^{-\theta}}{X_{nn}} \sum_k \frac{M_n^k}{\hat{\tilde{\Phi}}_n^k} \frac{E_i^k}{\hat{\tilde{\Psi}}_i^k}.$$

3.3 Data

I use data on bilateral product-level trade flows from the U.N. Comtrade database for the year 2003, classified at the 6-digit level of the 1996 revision of the Harmonized System. Data on bilateral relationships are taken from CEPII's Gravity dataset. When manufacturing is treated as a single sector, total manufacturing output data are taken from the OECD STAN database, where available, or the UNIDO INDSTAT database. Where not available from either source, total manufacturing output is imputed based on manufacturing value added from the World Bank's WDI database. When a sector is defined as a 2-digit ISIC (Rev. 3) industry, disaggregated output data are taken from the UNIDO INDSTAT database.

The full sample consists of trade flows among 130 countries classified into 4,608 product categories. When a sector is defined as a 2-digit ISIC industry, $\bar{d}_n^j$ is not identified for all countries due to a lack of disaggregated output data. Such data are available for between 33 and 79 countries, depending on the industry, with a median value of 62 countries. Table A1 lists the countries in the sample and the source of output data for each, and Table A2 lists the set of industries. Further details are in the Appendix.

3.4 Estimation Results

For the baseline set of estimations, to keep the specification as parsimonious as possible, I treat manufacturing as a single sector and assume that $\boldsymbol{Z}_{ni}$ consists of (logged) bilateral distance and an indicator for whether a pair of countries shares a common border. Each of these simplifications is later relaxed. Table 1 presents the coefficient estimates from the baseline specification. The results from the sector-level estimations are roughly in line with the literature. Bilateral trade is generally decreasing in distance and higher if countries share a border.

There is a great deal of variation in the coefficient estimates across the sector-level specifications. The distance elasticity varies by a factor of more than 2.5 from -0.94 to -2.41, and the effect of sharing a border varies from a statistically insignificant 0.20 for structural gamma PML to 0.96 for FE log LS. Santos Silva and Tenreyro (2006) attribute such differences to bias due to heteroskedasticity and sample selection associated with log-linear least squares and to finite sample biases for the PML estimators.

Table 1: Trade Cost Coefficient Estimates

	Log LS		Gamma PML		Poisson PML		NLS	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
	Sector-Level Fixed Effects Estimations							
mean($\ln \hat{d}_n$)	−1.63		0.61		−3.70		−4.40	
Distance	−1.95	(0.07)	−2.41	(0.14)	−0.96	(0.07)	−0.94	(0.13)
Shared Border	0.96	(0.19)	0.66	(0.29)	0.54	(0.12)	0.49	(0.18)
	Sector-Level Structural Estimations							
mean($\ln \hat{d}_n$)	−0.78		−2.42		−3.70		−3.77	
Distance	−2.11	(0.10)	−1.50	(0.07)	−0.96	(0.07)	−0.95	(0.11)
Shared Border	0.54	(0.35)	0.20	(0.24)	0.54	(0.12)	0.50	(0.21)
	Produt-Level Fixed Effects Estimations							
mean($\ln \hat{d}_n$)	−2.74				−2.95			
Distance	−1.15	(0.05)			−1.16	(0.08)		
Shared Border	0.79	(0.10)			0.51	(0.12)		
	Product-Level Structural Estimations							
mean($\ln \hat{d}_n$)	−2.24		−2.96		−2.95		−3.32	
Distance	−1.35	(0.07)	−1.20	(0.15)	−1.16	(0.08)	−1.02	(0.11)
Shared Border	0.75	(0.11)	0.37	(0.32)	0.51	(0.12)	0.55	(0.21)

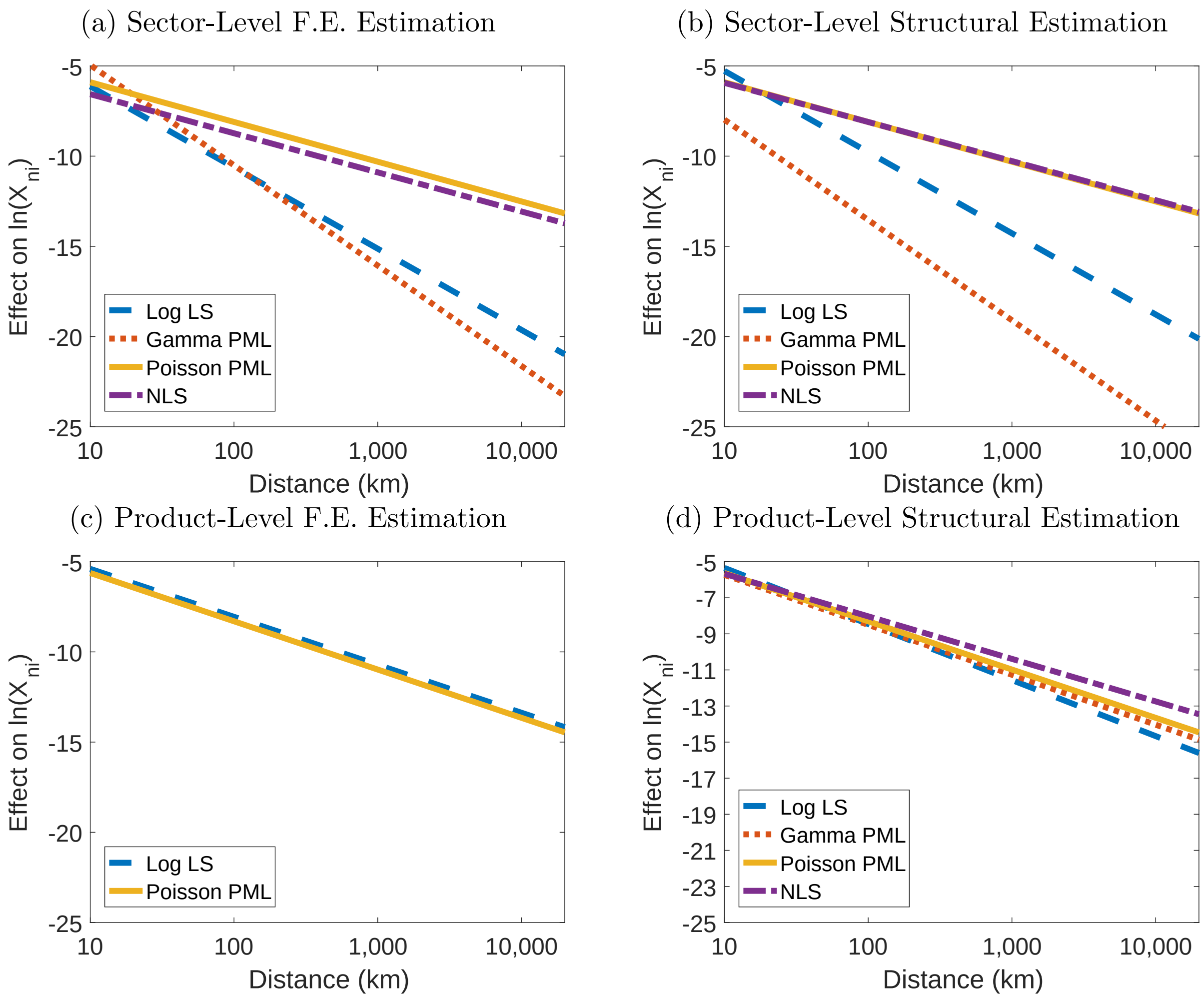
Notes: Standard errors (in parentheses) are robust to multi-way clustering by both importer and exporter. Parameters reported represent $\hat{b} = -\theta\hat{\beta}$. The implied percentage effect of each coefficient on the ad valorem tariff equivalent trade cost is $100 \times (e^{-\hat{b}/\theta} - 1)$. Number of observations: 11,193 for sector-level log LS, 16,770 for sector-level PML, 3,571,896 for product-level log LS, 77,276,160 for product-level PML.

By contrast, the estimates from the pooled product-level specifications are much more similar. To demonstrate the economic significance of these differences, Figure 1 plots the cumulative effect of distance on bilateral trade flows estimated by each specification.[26] It is clear that, over the relevant range of distances, the sector-level estimates diverge markedly, while the product-level estimates are nearly identical. To place these numbers in perspective, if we assume that $\theta = 6$, then the sector-level Poisson PML estimates imply a median bilateral trade cost that is equivalent to a 640% ad valorem tariff, versus 2,250% for the FE log LS estimates. The equivalent values for the product-level estimates are 782% and 741%.

These results indicate that estimates based on sector-level data suffer from significant bias from failing to control for product-level comparative advantage. The similarity between product-level estimates indicates that much of the discrepancy between sector-level estimates, which has been attributed to heteroskedasticity, sample-selection, and finite sample biases, is due to misspecification of the conditional expectation of X_{ni} by the omission of $\tilde{T}_{ni}$.

[26] Specifically, it plots the value of mean($\ln \hat{d}_n$) + $\hat{\beta}_{\text{Distance}} \ln(\text{Distance})$.

Figure 1: Estimated Distance-Related Trade Costs



The estimates based on Poisson PML and NLS confirm the intuition, based on equation (8), for the effect of product-level comparative advantage on sector-level estimates. To the extent that they differ significantly, the coefficient estimates move away from zero as we move from sector-level to product-level estimation. The opposite tends to be true for the log LS and gamma PML estimates. However, this is not surprising given the properties of these estimators, which place the most weight on observations for which the expected dependent variable is small. Because trade flows are bounded below by zero, the variance of the RCA measures in (8) will tend to be relatively small for country pairs that tend to trade relatively little. Thus, the value of $\tilde{T}_{ni}$ will be driven more by the multilateral effects of trade costs, given patterns of comparative advantage vis-à-vis third countries, as in (7). As a result, these estimators are the ones for which the bias is most likely to go in the opposite direction.

In addition, Santos Silva and Tenreyro (2006) demonstrate in a Monte Carlo experiment that the gamma PML estimator tends to be very sensitive to a particular form of measurement error in the data, whereby small trade flows are rounded to zero. Given that the strength of comparative advantage will tend to be small for small sector-level trade flows, these trade flows will tend to be more sensitive to trade costs. If such small expected trade flows show up as zeros in the data, this estimator will require that trade flows, overall, be overly sensitive to trade costs to make the predicted values of these small observations very close to zero. It is also possible that the biases in the log LS estimators pointed out by Santos Silva and Tenreyro (2006) are exacerbated by the bias due to comparative advantage. However, the fact that the log LS estimates and gamma PML estimates follow a similar pattern suggests that these issues are independent, and the similarity between the product-level log LS and PML estimates suggest that the former issues are relatively less important.

It is also interesting to note that the standard errors for the sector-level and product-level estimations are of similar magnitude despite the substantially larger sample size of the pooled product-level estimations. Standard errors are clustered by both importer and exporter, meaning that the number of clusters is the same in both sets of estimations and very small relative to the number of product-level observations. That the estimated standard errors are similar indicates that there is significant correlation among the errors of product-level trade flows to and from a common country.

3.5 Specification Tests

The estimation results summarized by Table 1 strongly indicate that sector-level gravity estimates are biased due to product-level comparative advantage. However, it is useful to formally test for this bias. The model makes clear that consistent estimation of $\boldsymbol{\beta}$ must be based on a product-level specification as long as sector-level trade flows depend on the interaction among country-product-specific effects, which cannot be controlled for in sector-level specifications. Directly testing for the presence of these effects is not straightforward for a couple reasons. First, the shear number of these effects makes it practically infeasible to compute the variance matrix required to test such a hypothesis. Second, if there is significant clustering in estimation errors, which appears to be the case with product-level trade data, the estimated variance matrix will be rank deficient, making such a test impossible even given sufficient computing power.[27] Therefore, I consider several indirect tests for the presence of these effects in the data.

The first test is an auxiliary estimation designed to test whether sector-level trade flows depend on patterns of product-level comparative advantage. Based on a pooled product-level estimation, it is straightforward to calculate fitted values of $\tilde{T}_{ni}$, up to an importer- and

[27]See Cameron and Miller (2015) for a discussion of this issue.

Table 2: Hausman Tests for Bias in Sector-Level Estimates

H_0	Test Statistic	FE Log LS	Structural Log LS	Structural GPML	Structural PPML	Structural NLS
$\gamma_{\tilde{T}_{ni}} = 0$	$t(128)$	20.69	23.72	34.87	24.05	16.82
	$P > \lvert t \rvert$	0.000	0.000	0.000	0.000	0.000
$\boldsymbol{\beta}^{Agg} = \boldsymbol{\beta}^{Pool}$	$F(2, 129)$	131.43	74.09	2.82	21.64	3.11
	$P > F$	0.000	0.000	0.064	0.000	0.048
$\boldsymbol{\beta}^{Agg} = \boldsymbol{\beta}^{Agg}_{PPML}$	$F(2, 129)$	112.94	142.11	33.18		0.19
	$P > F$	0.000	0.000	0.000		0.824
$\boldsymbol{\beta}^{Pool} = \boldsymbol{\beta}^{Pool}_{PPML}$	$F(2, 129)$	1.06	9.97	0.10		1.89
	$P > F$	0.350	0.000	0.907		0.155

Notes: All tests are based on estimated variance matrices that are robust to multi-way clustering by both importer and exporter. In the calculation of p-values, the test statistic as being distributed $t(N-L)$, for the first row, and $F(L, N-L)$, for the remaining rows, where $N = 130$ is the number of clusters, and $L = 2$ is the number of elements of $\boldsymbol{\beta}$.

exporter-specific scale factor. Using these fitted values, I perform a sector-level estimation that takes the following form:

$$E[X_{ni}] = \phi_n \psi_i e^{\boldsymbol{Z}'_{ni}\boldsymbol{\beta}} (\hat{\tilde{T}}_{ni})^{\gamma_{\tilde{T}_{ni}}}$$

where ϕ_n and ψ_i are importer and exporter fixed effects. I do this for each of the sector-level estimators with product-level analogues in Table 1, using the values of $\hat{\tilde{T}}_{ni}$ calculated from the analogous product-level estimation. The first row of Table 2 presents the test statistics and p-values of the null hypothesis $\gamma_{\tilde{T}_{ni}} = 0$. This hypothesis is resoundingly rejected for each estimator, clearly indicating that product-level comparative advantage influences sector-level trade flows, even after controlling for the appropriate sector-level effects.

The remaining rows of Table 2 present the results of three cluster-robust Hausman tests based on the comparison of estimates of $\boldsymbol{\beta}$ across estimations. Specifically, if the conditional expectation of X^k_{ni} is correctly specified and if sector-level trade flows are unaffected by product-level comparative advantage, then all of the estimators presented in Table 1 will have the same probability limit.[28] Therefore, I test for these issues by testing for differences between these estimators.

The second row of Table 2 presents tests for bias in sector-level estimates by testing the equality of sector-level estimates with their product-level counterparts. Equality is resoundingly rejected for both log LS estimators and Poisson PML. It is rejected at the 5% significance level for structural NLS and at the 10% level for structural gamma PML. This

[28] In the case of log LS estimators, this also requires that these estimators do not suffer from inconsistency due to heteroskedasticity and sample-selection bias.

result allows the formal conclusion that, to the extent that the product-level estimators are more robust to patterns of product-level comparative advantage, sector-level gravity estimates are biased. While it may be tempting to conclude, based on these results, that sector-level structural gamma PML and NLS are more robust than the log LS and Poisson PML estimators, the failure to reject at higher levels of significance stems largely from the lack of precision of their estimates, especially for gamma PML. While the NLS coefficient estimates are surprisingly consistent across specifications, this is a notoriously unreliable estimator (see, e.g., Santos Silva and Tenreyro, 2006) that is very sensitive to outliers, so I would strongly caution against the adoption of sector-level NLS as a workhorse gravity estimator based on these results alone.

The final two rows of Table 2 test the hypotheses that estimates based on the various sector- and product-level estimators are equal to those of the respective Poisson PML estimator. Equality is resoundingly rejected for all sector-level estimators other than NLS and cannot be rejected at any reasonable level of significant for any of the product-level estimators other than structural log LS. Even in the latter case, the value of the test statistic is dramatically smaller for the product-level estimator. These tests formally confirm the patterns evident on inspection of Table 1 and support the conclusion that differences in sector-level estimates across estimators are largely due to failure to control for product-level patterns of comparative advantage. In fact, it is not possible to reject the hypothesis that product-level FE OLS estimation is unbiased.

Finally, I evaluate the ability of sector-level estimators to predict variation in the product-level trade data. Though it is not straightforward to directly test for the patterns of comparative advantage that bias sector-level estimation, it is still informative to evaluate the performance of sector-level estimators in this regard. The product-level FE estimators are particularly useful for this task, as they nest sector-level FE estimators when the fixed effects are constrained to be equal across products. In fact, in the Poisson PML case, the sector-level FE estimator is isomorphic to the constrained product-level estimator. This is not the case with log LS due to the log transformation of X_{ni}^k. To give the constrained FE log LS estimator the best chance of success, I derive the specification under the null hypothesis that the assumptions of Proposition 2 hold, which allows for demand-side variation in product-level trade flows. In this case, product-level trade flows can be expressed as

$$X_{ni}^k = \frac{T_i d_{ni}^{-\theta}}{\tilde{\Phi}_n} M_n^k.$$

Thus, the constrained FE log LS estimator can be implemented by regressing $\ln(X_{ni}^k / M_n^k)$ on $\boldsymbol{Z}_{ni}$ and importer and exporter fixed effects.

Table 3: Relative Fit of Constrained Fixed-Effects Estimators

(a) Product-Level Fixed Effects Estimations

	Log LS		Poisson PML	
	Coeff.	S.E.	Coeff.	S.E.
Distance	−1.15	(0.04)	−1.16	(0.05)
Shared Border	0.79	(0.10)	0.51	(0.10)
	R^2: 0.6360		Pseudo-R^2: 0.9212	

(b) Constrained Product-Level Fixed Effects Estimations

	Log LS		Poisson PML	
	Coeff.	S.E.	Coeff.	S.E.
Distance	−0.93	(0.05)	−0.96	(0.04)
Shared Border	0.76	(0.13)	0.54	(0.11)
	R^2: 0.2797		Pseudo-R^2: 0.7700	

Notes: Standard errors (in parentheses) are robust to multi-way clustering by both importer and exporter. Parameters reported represent $\hat{b} = -\theta\hat{\beta}$.

Table 3 presents the coefficient estimates and (pseudo) R^2 values based on the constrained and unconstrained FE log LS and Poisson PML estimators.[29] As expected, the R^2 measures are significantly larger for the unconstrained estimators. It is also interesting to note that the constrained log LS estimates are much more in line with the product-level Poisson PML estimator than with the sector-level log LS estimator. The constrained estimator implicitly places more weight on country pairs with positive trade flows for a relatively large number of products, which apparently offsets the relative weighting of the log LS estimators toward small expected trade flows.

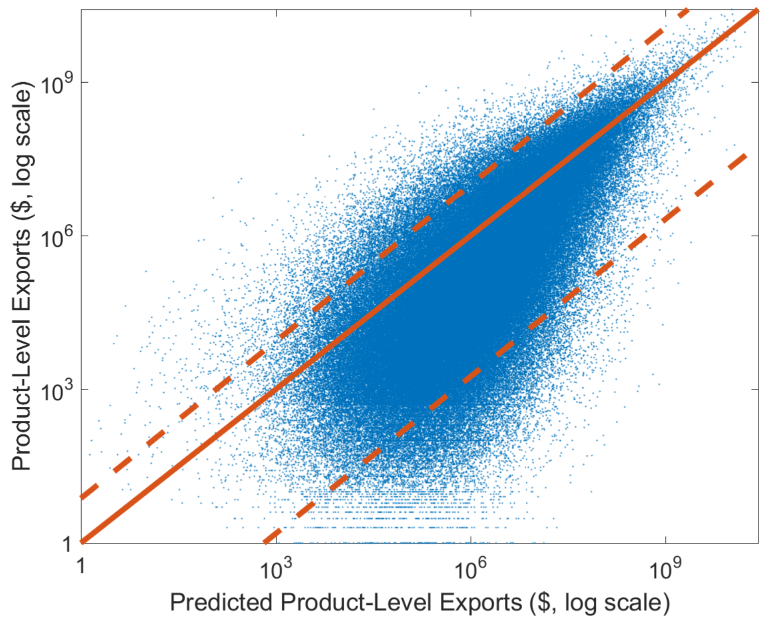
Sector-level estimators implicitly assume patterns of comparative advantage that imply particular patterns for product-level trade flows. Under the assumptions of Proposition 2, the following relationship holds:

$$X_{ni}^k = X_{ni} \frac{M_n^k}{M_n}.$$

Thus, variation in E_i^k is only due to the interaction between bilateral factors and product-level import demand. This provides another way to evaluate the predictive power of sector-level estimators. Figure 2 plots the actual against the predicted values of E_i^k based on the constrained (sector-level) Poisson PML estimation along with the 45-degree line and the 90% prediction interval of a log-linear OLS regression of E_i^k on $\hat{E}_i^k$. The product-level Poisson

[29] If estimated using $\ln(X_{ni}^k/M_n^k)$ as the dependent variable, the R^2 for constrained log LS must be adjusted to be consistent with the unconstrained estimator – i.e., $R^2 = \sum[\ln(X_{ni}^k) - \ln(\hat{X}_{ni}^k)]^2 / \sum[\ln(X_{ni}^k) - \ln(\bar{X})]^2$, where $\ln(\bar{X})$ is the mean over all finite values of $\ln(X_{ni}^k)$. For Poisson PML, the pseudo-R^2 is the measure proposed by Cameron and Windmeijer (1997) for exponential family regression models.

Figure 2: Predicted Values of E_i^k from Constrained Estimation



PML estimator fits these values exactly. It is clear that predicted and actual values of E_i^k are positively correlated, meaning that countries that export more overall, tend to export more of a given product. However, the prediction errors are enormous. A one-standard-deviation error is equivalent to shifting exports of a product by a factor 13.8.

3.6 Heteroskedasticity

The estimation results suggest that parameter estimates depend less on heteroskedasticity than previously thought, after controlling for product-level comparative advantage. However, to the extent that coefficient estimates differ across product-level estimators and for the sake of efficiency, it is still useful to use information on the properties of the estimation errors to help in selecting among these estimators. I follow Manning and Mullahy (2001) in estimating the relationship between the squared residual and a power function of the model predicted values – referred to by Head and Mayer (2014) as a "MaMu" test – which is given by

$$(X_{ni}^k - \hat{X}_{ni}^k)^2 = \lambda_0 (\hat{X}_{ni}^k)^{\lambda_1}.$$

Manning and Mullahy (2001) suggest estimating this relationship by OLS in its log-linear form, but Santos Silva and Tenreyro (2006) point out that this is only appropriate under the same conditions for which log LS is the appropriate estimator. Therefore, I estimate this relationship using the same estimator for which the residuals were obtained in each

Table 4: MaMu Tests

	FE Log LS	Structural Log LS	Structural GPML	Structural PPML	Structural NLS
H_0	$\lambda_1 = 2$	$\lambda_1 = 2$	$\lambda_1 = 2$	$\lambda_1 = 1$	$\lambda_1 = 0$
Aggregate	1.78	1.81	1.66	1.42	1.27
	(0.009)	(0.009)	(0.044)	(0.050)	(0.108)
Product-Level	2.69	1.30	1.13	1.43	0.92
	(0.003)	(0.001)	(0.010)	(0.062)	(0.054)

Notes: Estimated values of $\hat{\lambda}_1$ and standard errors (in parentheses) are reported. Standard errors for log least squares specifications are non-robust. All other standard errors are robust to multi-way clustering by both importer and exporter.

case.[30] Because the log LS estimators are valid when $\lambda_1 = 2$, the MaMu tests based on these estimators are valid tests of this hypothesis. The MaMu tests based on the PML estimators are asymptotically valid given that inference regarding λ_1 is based on heteroskedasticity-robust standard errors.

The results of the MaMu tests are presented in Table 4. For each of the estimators, the assumed value of λ_1 can be rejected. The hypothesis that $\lambda_1 = 0$ and $\lambda_1 = 2$ can be easily rejected in all cases. In most cases, the estimates of λ_1 based on product-level estimators are smaller. This is consistent with the model's implication that $\tilde{T}_{ni}$, which is relegated to the error term in sector-level estimation, is smaller for country pairs with smaller expected sector-level trade flows. The exceptions are product-level FE log LS, which appears to be an outlier, and Poisson PML, whose estimate of λ_1 is virtually the same across specifications. The preponderance of the evidence suggests that λ_1 likely lies between 1 and 2, and the MaMu tests based on product-level estimators suggest that λ_1 is likely closer to 1 than to 2, including the failure to reject $\lambda_1 = 1$ based on product-level structural NLS. This suggests that product-level Poisson PML should be the preferred gravity estimator, though the gamma PML and log LS estimators should be considered for robustness.

3.7 Estimated Border Costs

Thus far, primary attention has been placed on the determinants of the bilateral component of trade costs, but border costs also make up a large share of estimated trade costs. These costs are particularly important in welfare analysis because, as Waugh (2010) argues, the asymmetric component of estimated border costs likely reflects policy differences across countries. Table 1 shows that average estimated border effects, like the parameters that determine bilateral trade costs, are much more similar across product-level estimations than

[30] Specifically, I use log-linear OLS with non-robust standard errors in the FE and Structural log LS cases and the same PML estimator with multi-way cluster-robust standard errors for the PML cases.

Table 5: Correlation of Border Costs with Income Components

	Aggregate (constant d_{ii})		Aggregate		Product-Level	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Output	−0.65	(0.06)	−0.34	(0.06)	−0.21	(0.09)
GDP per Worker	−0.23	(0.12)	−0.66	(0.14)	−0.89	(0.15)
R^2	0.6568		0.5188		0.3817	
Mean $\ln(\hat{d}_n^{-\theta})$	−3.70		−3.70		−2.95	
Std. Dev. $\ln(\hat{d}_n^{-\theta})$	2.27		2.10		2.41	

Notes: Dependent variable is $\ln(\hat{d}_n^{-\theta})$. All variables specified in logs. Heteroskedasticity robust standard errors are in parentheses.

across sector-level estimations. In this section, I briefly explore the implications for the distribution of estimated border costs across countries.

Based on a sector-level estimation, Waugh (2010) finds that estimated border costs are systematically larger for low-income countries, which implies that reducing these costs would substantially decrease cross-country income inequality. Ramondo et al. (2016) argue that much of this variation in estimated border costs is actually caused by assuming that internal trade costs are zero instead of the more reasonable assumption that they are increasing in country size. Table 5 reports the coefficients of a regression of estimated border costs on country size (measured by total manufacturing output) and GDP per worker. In the first two columns, border costs are derived from the sector-level Poisson PML estimation.[31] In the first column, the estimation imposes the assumption that internal trade costs are identical in every country.[32] In the second and third columns, internal trade costs vary with internal distance as in the baseline estimations, where the third column uses border costs derived from the product-level Poisson PML estimation.

The first set of border cost estimates are only weakly correlated with GDP per capita after controlling for country size. Adding variation in internal trade costs reduces the overall variation of estimated border costs and their correlation with country size. Both of these findings are consistent with the conclusions of Ramondo et al. (2016). The third column indicates that estimating border costs using product-level data further reduces their correlation with country size to a level that is only marginally statistically significant. In addition, though the average level of border costs falls in the product-level Poisson PML estimates, both variation across countries and the correlation with GDP per worker increase, meaning

[31] For parsimony, I present only the results based on Poisson PML. Results based on the other estimators are similar.

[32] I do this by setting the log of internal distance in each country to its average value in the sample. This is similar to the specification of Waugh (2010), which divides bilateral distance into six discrete categories, omitting the range 0-375 miles.

that analysis based on sector-level estimates will understate the effect of border costs on the cross-country income distribution. In this sense, these results support the conclusions of Waugh (2010) without contradicting the insights of Ramondo et al. (2016).

3.8 The Trade Impact of Changes in Border Costs

Often, gravity estimation is intended not only to estimate the effect of variables of interest on trade flows but to parameterize a structural model to predict changes in trade flows due to a change in trade costs. In this spirit, I conduct a simple counterfactual experiment based on the baseline gravity estimations in order to demonstrate the effects on predictions of accounting for product-level comparative advantage: I consider the effect on bilateral trade flows of eliminating all border costs.

Based on the structural gravity framework that forms the basis of the estimation specifications, it is straightforward to predict the effect of changes in trade costs, holding constant output and expenditure. Head and Mayer (2014) refer to this as the Modular Trade Impact (MTI) because it takes advantage of the fact that structural gravity models allow the allocation of bilateral trade flows to be determined separately from the allocation of production and expenditure.[33] The MTI allows the multilateral resistance terms in (3) to adjust in response to changes in trade costs but is not a full general equilibrium impact (GETI) because factor prices are held constant.

The MTI is very useful tool for evaluating the effects of changes in trade costs estimated based on a multi-sector and/or product-level gravity model because, as my model makes clear, it is not necessary to specify the form of demand across products, details of factor markets, or the sources of comparative advantage to estimate the parameters of a trade cost function. Computing the GETI would involve specifying and parameterizing each of these, as well as the value of the trade cost elasticity (θ), but the MTI does not. In addition, as Head and Mayer (2014) and Anderson and van Wincoop (2003) report, the differences between an MTI and GETI tend to be relatively small in practice, especially for small changes in trade costs.

For the product-level gravity model, the MTI is defined as follows:

$$\text{MTI}_{ni} \equiv \frac{X'_{ni}}{X^0_{ni}} = \left(\frac{d'_{ni}}{d^0_{ni}}\right)^{-\theta} \sum_k \frac{\Psi_i^{0k}}{\Psi_i'^k}\frac{\Phi_n^{0k}}{\Phi_n'^k},$$

where naughts denote baseline values and primes denote values after a change in d_{ni}. For

[33] See Anderson (2011) for a detailed analysis of this property of structural gravity models.

Table 6: Median MTI of Elimination of Border Costs

	Sector-Level	Product-Level	Product-Level (Δd_{ni}^{SL})
FE Log LS	3.207	1.548	1.301
Structural Log LS	1.541	1.398	1.112
Structural GPML	2.185	1.585	1.471
Structural PPML	2.311	1.592	1.736
Structural NLS	2.305	1.687	1.767

sector-level specifications, this expression reduces to

$$\text{MTI}_{ni} \equiv \frac{X'_{ni}}{X^0_{ni}} = \left(\frac{d'_{ni}}{d^0_{ni}}\right)^{-\theta} \frac{\Psi^0_i}{\Psi'_i}\frac{\Phi^0_n}{\Phi'_n}.$$

The values of Φ'^k_n and Ψ'^k_i are calculated according to (3). Because data on X^k_{nn} are unavailable, computations based on the product-level model use the predicted values of Y^k_i and X^k_n, calculated using (9) and (10) and the baseline estimates of d_{ni} and $\bar{d}_n$.

Table 6 presents the median MTI of setting all border costs to zero, using estimates from the baseline specifications.[34] To aide comparison between the MTIs based on sector-level and product-level specifications, the third column presents MTIs calculated using the product-level model and baseline trade cost estimates but reduces border costs by the amount estimated in the corresponding sector-level specification.

As with the coefficient estimates, the MTIs based on sector-level specifications are much more heterogeneous than those based on product-level specifications. Also, the product-level MTIs tend to be much smaller, about half the size of sector-level MTIs in terms of percentage changes in trade flows. The former result is due to the heterogeneity in border costs estimated by the sector-level specifications. The latter depends mostly on computing the MTI using the product-level model. To see this, note that the partial trade impact (PTI), $(d'_{ni}/d^0_{ni})^{-\theta}$, is the same in the first and third columns of Table 2. Because the effects of trade costs are ameliorated by product-level comparative advantage, the product-level model, which accounts for this, predicts a smaller MTI for a given PTI than the sector-level model, which does not. This effect can be offset if border effects are estimated to be much smaller by the sector-level specifications than the product-level specifications. This is true of the first three estimators in Table 2, but the latter effect is never large enough to offset the former.

[34]For a very small number of countries, borders are estimated to enhance trade – i.e., $\hat{\bar{d}}_n < 1$. For these countries, I leave the value of $\hat{\bar{d}}_n$ unchanged in computing MTIs. Setting $d_n = 1$ for these countries, as well, does not significantly affect the results.

This simple exercise demonstrates that accounting for product-level comparative advantage is important not only for estimating the size of trade barriers but also for predicting the effects of changes in trade costs. It is worth noting that the welfare effects of change in trade flows depend on the scope of comparative advantage in the product-level model. For example, in models closely related to mine, Levchenko and Zhang (2014) and French (2016) show that the welfare gains from trade are much larger in the presence of comparative advantage across sectors or products. Thus, the fact that the product-level model predicts smaller changes in trade flows does not indicate that it would necessarily predict smaller gains from removing border costs. Of course, these welfare effects will also depend on assumptions regarding the form of demand across products, factors of production, etc.

4 Extensions

The baseline set of estimations show that product-level comparative advantage causes substantial bias in trade cost estimates based on sector-level data. To demonstrate this phenomenon as clearly as possible, the baseline specification was quite simple. In this section, I extend the analysis along two dimensions. First, I allow parameters of the trade cost function to vary across more narrowly defined sectors. Second, I consider the effects of a broader set of covariates, which reflect trade policy and historical country ties.

4.1 Multiple Sectors

Though gravity estimations have almost always treated countries as one-sector economies (typically focussing on manufacturing or merchandise trade), recent papers such as Anderson and Yotov (2010), Chen and Novy (2011), and Levchenko and Zhang (2016) have considered multiple, more narrowly-defined sectors, studying trade flows within manufacturing industries defined approximately at the 2-digit ISIC level. To evaluate the extent to which ignoring product-level comparative advantage is problematic for gravity estimations focused on somewhat more disaggregated sectors, I repeat the estimations from above industry-by-industry.

Tables A3 - A5, in the Appendix, present the results of multi-sector estimations comparable to those reported in Table 1, which were conducted separately for the 21 industries defined in Table A2, as well as p-values for Hausman tests analogous to those reported in Table 2. The overall pattern that emerges is that, while there is significant heterogeneity in coefficient estimates across sectors, the differences between multi-sector, sector-level and multi-sector, pooled product-level estimates are similar to those of the one-sector estimations. These results are summarized in Tables 7 and 8. Table 7 presents the share of

Table 7: Multi-Sector Hausman Tests for Bias: Share with H > F_α

H_0	α	FE Log LS	Structural Log LS	Structural GPML	Structural PPML	Structural NLS
$\boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	0.10	1.00	1.00	0.62	0.90	0.62
	0.01	1.00	1.00	0.33	0.81	0.52
$\boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Agg}_{PPML}$	0.10	1.00	1.00	0.86		0.14
	0.01	1.00	1.00	0.76		0.05
$\boldsymbol{\beta}^{j,Pool} = \boldsymbol{\beta}^{j,Pool}_{PPML}$	0.10	0.43	0.81	0.81		0.14
	0.01	0.24	0.57	0.38		0.14

Notes: H represents the test statistic for the stated null hypothesis. All tests are based on estimated variance matrices that are robust to multi-way clustering by both importer and exporter. Test statistics are treated as being distributed $t(N-L)$ where $N = 130$ is the number of clusters, and $L = 2$ is the number of elements of $\boldsymbol{\beta}$, and α is the level of significance.

sectors for which the hypotheses tested in Table 2 are rejected, at the 10% and 1% levels of significance. As with the one-sector estimations, in most cases we can reject the hypothesis that the sector-level and product-level estimates are equal, indicating that there is bias even in multi-sector estimates. Also, for all but NLS, we can reject equality between the estimates based on Poisson PML and the other estimators for almost all sector-level estimations, where again the failure to reject for NLS is largely due to the inefficiency of this estimator. Meanwhile, we cannot reject equality with Poisson PML for the product-level estimations for a large share of sectors. Together, as with the one-sector estimations, the findings support the conclusion that sector-level estimates are significantly biased and that differences among sector-level estimators can be largely attributed to failure to control for product-level patterns of comparative advantage.

This is not to say that heterogeneity in trade cost parameters across industries is not a concern, only that any heterogeneity does not affect the main conclusions regarding bias in sector-level gravity estimation due to product-level comparative advantage. The first two columns of Table 8 show that, though the median parameter values from the multi-sector estimations are not substantially different from the single-sector estimation, we can reject the hypothesis that these values are equal for a large share of industries, especially for the distance elasticity.[35] This result does not appear to differ significantly between sector-level and product-level estimations.

Finally, I allow for the most extreme form of parameter heterogeneity by performing the estimations product-by-product. There is evidence of heterogeneity beyond the industry level, and again this conclusion does not differ significantly between sector-level and

[35]The median value is reported to minimize the effect of outliers. Results based on trade-weighted averages are similar. Unlike the case with border costs, it is not possible to construct an ideal index of industry-level coefficients without assuming a functional form for expenditure across industries.

Table 8: Multi-Sector Hausman Tests for Heterogeneity (Poisson PML)

		$\boldsymbol{\beta}^j_{Agg} - \boldsymbol{\beta}_{Agg}$	$\boldsymbol{\beta}^j_{Pool} - \boldsymbol{\beta}_{Pool}$	$\boldsymbol{\beta}^{jk} - \boldsymbol{\beta}^j_{Agg}$	$\boldsymbol{\beta}^{jk} - \boldsymbol{\beta}^j_{Pool}$
Median Coefficient Differences					
	Distance	-0.15	-0.03	-0.18	-0.01
	Shared Border	0.08	0.08	0.00	-0.01
Share with H $> t_\alpha$					
$\alpha = 0.10$	Distance	0.714	0.667	0.471	0.433
	Shared Border	0.333	0.333	0.270	0.262
$\alpha = 0.01$	Distance	0.476	0.619	0.249	0.221
	Shared Border	0.143	0.095	0.090	0.086

Notes: H represents the test statistic for the null hypothesis that the stated function of parameters is equal to zero for each industry and product. All tests are based on estimated variance matrices that are robust to multi-way clustering by both importer and exporter. Test statistics are treated as being distributed $t(N-L)$ where $N = 130$ is the number of clusters, and $L = 2$ is the number of elements of $\boldsymbol{\beta}$, and α is the level of significance.

product-level estimation. This is summarized in the last two columns of Table 8.[36] While equality between product-by-product and pooled estimates cannot be rejected for a majority of products, it is rejected for a share much larger than the significance level of the test. However, whereas the pooled product-level estimators are clearly superior to the sector-level estimators, as they eliminate the bias due to product-level comparative advantage, there is a tradeoff to allowing for a large degree of parameter heterogeneity. While there may be some model misspecification due to pooling product-level trade flows, there are significant efficiency gains. For example, for Poisson PML, the standard errors for multi-sector, pooled product-level estimates are on average 35% larger than the one-sector estimates, and those for product-by-product estimates are on average more than 3 times larger. Especially for the product-by-product estimations, the huge number of parameters and imprecision with which they are identified makes interpretation and inference very difficult. Therefore, a useful rule of thumb would seem to be to pool the estimation to the level at which there is interest in the effect of a variable unless there is explicit interest in heterogeneity. When estimates are being used to parameterize a fully-specified model, then it may be possible to specify an ideal aggregation index based on disaggregated parameter estimates, and it may be possible to estimate the index itself more efficiently than the disaggregated parameters.[37] It may also be possible to improve the efficiency of product-by-product estimates by using seemingly-unrelated-regression-type techniques.[38]

[36]These results omit a small number (less that 1%) of products whose estimations did not converge.

[37]In fact, the pooled product-level coefficients constitute such indexes, though they may not be "ideal", given a particular model.

[38]Doing this is beyond the scope of this paper, but it is a potentially fruitful avenue of future research.

4.2 Policy-Related Gravity Variables

While the baseline estimation results clearly demonstrate that sector-level gravity estimation is biased due to product-level comparative advantage, they restricted attention to estimation of the effects of geographical barriers to trade. In this section, I repeat the baseline set of estimations including other common gravity variables that proxy for cultural and political ties between countries and trade policy. Specifically, I include indicators for whether country pairs share a common language, historical colonial ties, a regional trade agreement (RTA), or a common currency. Table 9 presents the estimation results.

The results from the sector-level estimations are roughly in line with the literature. As with the baseline estimation results, bilateral trade is generally decreasing in distance and higher if countries share a border. Trade is also generally increasing if countries share a common language, colonial ties, an RTA, or a common currency. A similar pattern to the baseline also emerges in the differences between the sector-level and product-level estimations. There is a great deal of heterogeneity among the sector-level specifications, which is generally reduced in the product-level estimations, especially for distance, shared border, and common language.

Interestingly, a great deal of heterogeneity remains in the product-level estimates of the effects of colonial ties and RTAs. The fact that the coefficients for these variables change monotonically, moving from gamma PML to Poisson PML to NLS, indicates that the true elasticity of trade flows with respect to these variables may be non-constant. In particular, it seems reasonable to speculate that colonial ties are more important for relatively small former colonies whose economies may have been greatly shaped by their colonizers and may still depend heavily on investment as well as political and military support from the former colonial power. In the case of RTAs, this could reflect the fact that trade agreements among blocs of large countries, such as NAFTA and the EU customs union, go much farther in scope than other regional agreements.[39]

Also interesting is that the product-level estimates are quite pessimistic regarding the trade-enhancing effects of sharing a common currency, with only the FE log LS estimator finding a significantly positive effect. Given the wide range of estimates of this effect in the literature, this result is not an outlier.[40] However, it does appear that, after controlling for product-level comparative advantage, the estimates are converging to a value at the low end of the range in the literature. It is important to note that the estimated effects of the policy

[39] In a meta-analysis of gravity estimations, Head and Mayer (2014) note that "the North-American agreement seems to be associated with larger amounts of trade creation." Baier et al. (2016) also find evidence of heterogeneous effects of trade agreements.

[40] Head and Mayer (2014) review the literature on gravity estimation of the effects of common currencies.

Table 9: Trade Cost Coefficient Estimates

	Log LS		Gamma PML		Poisson PML		NLS	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
	Sector-Level Fixed Effects Estimations							
mean($\ln \hat{d}_n$)	−0.07		3.39		−3.61		−4.71	
Distance	−1.71	(0.07)	−2.14	(0.14)	−0.72	(0.07)	−0.61	(0.08)
Shared Border	0.72	(0.20)	0.58	(0.34)	0.42	(0.11)	0.29	(0.09)
Common Language	0.96	(0.14)	1.09	(0.20)	0.36	(0.11)	0.53	(0.12)
Colonial Ties	0.95	(0.18)	1.44	(0.49)	0.02	(0.15)	−0.24	(0.18)
RTA	0.57	(0.16)	0.28	(0.31)	0.83	(0.13)	0.97	(0.18)
Common Currency	0.39	(0.42)	1.29	(0.65)	−0.05	(0.19)	0.07	(0.17)
	Sector-Level Structural Estimations							
mean($\ln \hat{d}_n$)	−0.27		−1.40		−3.61		−4.07	
Distance	−1.65	(0.16)	−1.30	(0.09)	−0.72	(0.07)	−0.55	(0.09)
Shared Border	−0.44	(0.52)	−0.23	(0.23)	0.42	(0.09)	0.32	(0.09)
Common Language	1.48	(0.45)	0.55	(0.17)	0.36	(0.09)	0.31	(0.19)
Colonial Ties	1.38	(0.41)	0.90	(0.26)	0.02	(0.12)	−0.11	(0.17)
RTA	1.30	(0.39)	0.72	(0.19)	0.83	(0.11)	1.30	(0.22)
Common Currency	−0.67	(0.49)	0.15	(0.27)	−0.05	(0.17)	−0.05	(0.15)
	Produt-Level Fixed Effects Estimations							
mean($\ln \hat{d}_n$)	−1.84				−2.83			
Distance	−1.07	(0.06)			−0.93	(0.07)		
Shared Border	0.63	(0.08)			0.40	(0.09)		
Common Language	0.55	(0.07)			0.44	(0.09)		
Colonial Ties	0.43	(0.10)			0.02	(0.12)		
RTA	0.21	(0.12)			0.80	(0.10)		
Common Currency	0.34	(0.14)			−0.04	(0.15)		
	Product-Level Structural Estimations							
mean($\ln \hat{d}_n$)	−1.41		−1.80		−2.83		−3.73	
Distance	−1.25	(0.08)	−1.21	(0.19)	−0.93	(0.07)	−0.75	(0.09)
Shared Border	0.55	(0.10)	0.29	(0.24)	0.40	(0.09)	0.30	(0.12)
Common Language	0.70	(0.09)	0.73	(0.16)	0.44	(0.09)	0.76	(0.16)
Colonial Ties	0.66	(0.13)	0.67	(0.16)	0.02	(0.12)	−0.63	(0.16)
RTA	0.31	(0.13)	−0.35	(0.30)	0.80	(0.10)	0.82	(0.23)
Common Currency	−0.05	(0.18)	0.36	(0.27)	−0.04	(0.15)	0.05	(0.26)

Notes: Standard errors are clustered by exporter. Parameters reported represent $\hat{b} = -\theta\hat{\beta}$. The implied percentage effect of each coefficient on the ad valorem tariff equivalent trade cost is $100 \times (e^{-\hat{b}/\theta} - 1)$. Number of observations: 11,193 for sector-level log LS; 16,770 for sector-level PML; 3,571,896 for product-level log LS; 77,276,160 for product-level PML.

variables – RTA and common currency – as well as, to a lesser extent, common language and colonial ties, likely suffer from endogeneity bias, as these relationships will be more likely to form between countries that tend to trade a great deal for other reasons. Controlling for product-level comparative advantage, which is likely one of these reasons, should lessen but likely not eliminate the endogeneity problem.

It is also worth noting that the Poisson PML estimates, other than border costs and the distance elasticity, are quite similar between sector-level and product-level estimation. There are a couple factors that may contribute to this result. First, the contribution to d_{ni} of all of these variables is quite small compared to border costs and distance. Therefore, the bilateral effect of these variable on $\tilde{T}_{ni}$, which operates through (8), will also be relatively small. If these variables are not strongly correlated with patterns of comparative advantage or other countries' trade costs, then the bias in sector-level estimates of the effects of these variables will also be small. Second, Poisson PML weights all observations equally. The fact that sector-level estimates of these coefficients appear to be biased for the other estimators suggests that this reasoning only applies on average for country pairs near the middle of the distribution of bilateral trade flows.

Consider, for example, colonial ties. Poisson PML estimates the same effect whether controlling for comparative advantage or not. However, it may be the case that small, relatively distant countries' colonial ties are highly correlated with their patterns of comparative advantage. For example Sub-Saharan African countries tend to have strong comparative advantage vis-à-vis their former European colonizers, meaning that ignoring comparative advantage will tend to overestimate the effect of colonial ties, as it does for log LS and gamma PML. This is less likely to be the case for country pairs who trade more in absolute terms. The NLS estimates suggest that a similar correlation exists between colonial ties and comparative advantage for countries with very large bilateral trade flows. However, as discussed previously, this estimator is quite sensitive to outliers. Since there are comparatively few large trading partners with colonial relationships, especially ones that do not share a common language, it is likely this particular result is driven by a small number of observations.

The consistency of the Poisson PML estimates does suggest that, if one is only interested in the parameter governing the effect of a variable that has a relatively small effect on overall trade costs and is unlikely to have a heterogeneous effect across country pairs, then sector-level Poisson PML may provide an estimate that is reasonably robust to the omission of product-level comparative advantage. However, it is far from guaranteed that this phenomenon will persist in other samples.

5 Conclusion

This paper has demonstrated both theoretically and empirically that sector-level gravity estimation is biased in the presence non-trivial patterns of product-level comparative advantage. I have characterized the effect of product-level comparative advantage on aggregated trade flows in a very general theoretical framework and developed an approach for estimating trade costs using pooled product-level trade data. Comparing coefficient estimates based on sector and product-level data indicate that this bias is both statistically and economically significant. Overall, my results indicate that pooled product-level Poisson PML using data at the lowest available level of aggregation should be the primary estimator of choice for most empirical analyses in which a standard sector-level estimator is typically employed.

The application in this paper was based on a cross-section of bilateral, product-level trade flows. However, the pooled product-level estimator can also be applied to panel data. The estimation techniques used in the application for both the structural and FE approaches are relatively efficient in terms of memory requirements, so adding a time dimension would not require extraordinary computing resources beyond what would be required to handle the raw data.

There are many avenues to pursue for increasing the efficiency and robustness of gravity estimation. Two briefly mentioned in this paper for estimators that allow for heterogeneous product-level effects are (a) direct estimation of an ideal index or (b) employing a generalized estimator that takes into account information on the covariance structure of the product-level errors. Another way to potentially improve efficiency is explicitly modelling countries' patterns of product-level comparative advantage or the determinants of product-level heterogeneity, which would substantially increase the degrees of freedom available for a product-level estimator.

References

Allen, Treb, Costas Arkolakis, and Yuta Takahashi (2014) "Universal Gravity," NBER Working Paper No. 20787.

Alvarez, Fernando and Robert E. Lucas, Jr. (2007) "General Equilibrium Analysis of the Eaton-Kortum Model of International Trade," *Journal of Monetary Economics*, 54 (6), 1726 – 1768.

Anderson, James E. (1979) "A Theoretical Foundation for the Gravity Equation," *The American Economic Review*, 69 (1), 106–116.

——— (2011) "The Gravity Model," *Annual Review of Economics*, 3 (1), 133–160.

Anderson, James E. and J. Peter Neary (2003) "The Mercantilist Index of Trade Policy," *International Economic Review*, 44 (2), 627–649.

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

——— (2004) "Trade Costs," *Journal of Economic Literature*, 42 (3), 691–751.

Anderson, James E. and Yoto V. Yotov (2010) "The Changing Incidence of Geography," *American Economic Review*, 100 (5), 2157–86.

Arkolakis, Costas, Arnaud Costinot, Dave Donaldson, and Andrés Rodríguez-Clare (2015) "The Elusive Pro-Competitive Effects of Trade," Working Paper 21370, National Bureau of Economic Research.

Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare (2012) "New Trade Models, Same Old Gains?" *American Economic Review*, 102 (1), 94–130.

Arvis, Jean-François and Ben Shepherd (2013) "The Poisson quasi-maximum likelihood estimator: a solution to the 'adding up' problem in gravity models," *Applied Economics Letters*, 20 (6), 515–519.

Baier, Scott L., Yoto V. Yotov, and Thomas Zylkin (2016) "On The Widely Differing Effects of Free Trade Agreements: Lessons from Twenty Years of Trade Integration," Working Paper.

Cameron, A. Colin and Douglas L. Miller (2015) "A Practitioner's Guide to Cluster-Robust Inference," *Journal of Human Resources*, 50 (2), 317–372.

Cameron, A. Colin and Frank A. G. Windmeijer (1997) "An R-Squared Measure of Goodness of Fit for Some Common Nonlinear Regression Models," *Journal of Econometrics*, 77 (2), 329–342.

Chen, Natalie and Dennis Novy (2011) "Gravity, trade integration, and heterogeneity across industries," *Journal of International Economics*, 85 (2), 206 – 221.

Costinot, Arnaud, Dave Donaldson, and Ivana Komunjer (2012) "What Goods Do Countries Trade? A Quantitative Exploration of Ricardo's Ideas," *Review of Economic Studies*, 79 (2), 581–608.

Costinot, Arnaud and Andrés Rodríguez-Clare (2014) "Trade Theory with Numbers: Quantifying the Consequences of Globalization," in Elhanan Helpman Gita Gopinath and Kenneth Rogoff (Eds.), *Handbook of International Economics*, Vol. 4, Elsevier, Chap. 4, 197–261.

Eaton, Jonathan and Samuel Kortum (2002) "Technology, Geography, and Trade," *Econometrica*, 70 (5), 1741–1779.

Eaton, Jonathan and Akiko Tamura (1994) "Bilateralism and Regionalism in Japanese and U.S. Trade and Direct Foreign Investment Patterns," *Journal of the Japanese and International Economies*, 8 (4), 478 – 510.

Egger, Peter H. and Sergey Nigai (2015) "Structural Gravity with Dummies Only: Constrained ANOVA-Type Estimation of Gravity Models," *Journal of International Economics*, 97 (1), 86–99.

Fally, Thibault (2015) "Structural Gravity and Fixed Effects," *Journal of International Economics*, 97 (1), 76–85.

Fernández-Val, Iván and Martin Weidner (2016) "Individual and Time Effects in Nonlinear Panel Models with Large N, T," *Journal of Econometrics*, 192 (1), 291–312.

French, Scott (2015) "Welfare in the Eaton-Kortum Model of International Trade," Working Paper.

——— (2016) "The Composition of Trade Flows and the Aggregate Effects of Trade Barriers," *Journal of International Economics*, 98, 114 – 137.

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

Guimarães, Paulo and Pedro Portugal (2010) "A Simple Geasible Procedure to Fit Models with High-Dimensional Fixed Effects," *Stata Journal*, 10 (4), 628–649.

Hallak, Juan Carlos (2006) "Product Quality and the Direction of Trade," *Journal of International Economics*, 68 (1), 238 – 265.

Head, Keith and Thierry Mayer (2014) "Gravity Equations: Workhorse, Toolkit, and Cookbook," in Gita Gopinath, Elhanan Helpman, and Kenneth Rogoff (Eds.), *Handbook of International Economics*, Vol. 4, Elsevier, Chap. 3, 131–195.

Helpman, Elhanan, Marc Melitz, and Yona Rubinstein (2008) "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123 (2), 441–487.

Kehoe, Timothy J., Pau S. Pujolas, and Jack Rossbach (2016) "Quantitative Trade Models: Developments and Challenges," NBER Working Paper No. 22706.

Levchenko, Andrei A. and Jing Zhang (2014) "Ricardian Productivity Differences and the Gains from Trade," *European Economic Review*, 65, 45–65.

——— (2016) "The Evolution of Comparative Advantage: Measurement and Welfare Implications," *Journal of Monetary Economics*, 78, 96–111.

Manning, Willard G and John Mullahy (2001) "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics*, 20 (4), 461–494.

Mayer, Thierry and Soledad Zignago (2011) "Notes on CEPIIs distances measures (GeoDist)," CEPII Working Paper 2011-25.

Ossa, Ralph (2015) "Why Trade Matters after All," *Journal of International Economics*, 97 (2), 266–277.

Ramondo, Natalia, Andrés Rodríguez-Clare, and Milagro Saborío-Rodríguez (2016) "Trade, Domestic Frictions, and Scale Effects," *American Economic Review*, 106 (10), 3159–84.

Samuelson, Paul A. (1954) "The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments," *The Economic Journal*, 64 (254), pp. 264–289.

Santos Silva, J. M. C. and Silvana Tenreyro (2006) "The Log of Gravity," *The Review of Economics and Statistics*, 88 (4), pp. 641–658.

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

A Additional Tables

Table A1: Countries and Sources of Manufacturing Output Data

Country	Source	Country	Source	Country	Source
Albania	INDSTAT	Georgia	INDSTAT	Panama	INDSTAT(int.)
Argentina	WDI	Germany	STAN	Papua New Guinea	WDI
Australia	INDSTAT	Ghana	INDSTAT	Peru	INDSTAT
Austria	STAN	Greece	STAN	Philippines	INDSTAT
Azerbaijan	INDSTAT	Grenada	INDSTAT	Poland	STAN
Bahamas	WDI	Guatemala	WDI	Portugal	STAN
Bangladesh	WDI	Honduras	WDI	Qatar	INDSTAT
Barbados	WDI	Hungary	STAN	Rep. of Korea	STAN
Belarus	WDI	Iceland	STAN	Rep. of Moldova	INDSTAT
Belize	WDI	India	INDSTAT	Romania	INDSTAT
Benin	WDI	Indonesia	INDSTAT	Russian Federation	INDSTAT
Bolivia	WDI	Iran	INDSTAT	Rwanda	WDI
Botswana	INDSTAT	Ireland	STAN	Saint Kitts and Nevis	INDSTAT
Brazil	INDSTAT	Israel	STAN	St. Lucia	WDI
Brunei Darussalam	WDI	Italy	STAN	Samoa	WDI
Bulgaria	INDSTAT	Jamaica	WDI	Sao Tome and Princ.	WDI
Burkina Faso	WDI	Japan	STAN	Saudi Arabia	INDSTAT(int.)
Burundi	WDI	Jordan	INDSTAT	Senegal	WDI
Cambodia	WDI	Kazakhstan	INDSTAT	Slovakia	STAN
Cameroon	WDI	Kenya	INDSTAT	Slovenia	STAN
Canada	STAN	Kyrgyzstan	INDSTAT	South Africa	INDSTAT
Cape Verde	WDI	Latvia	INDSTAT	Spain	STAN
Central African Rep.	WDI	Lebanon	WDI	Sri Lanka	INDSTAT(int.)
Chile	INDSTAT	Lithuania	INDSTAT	Sudan	WDI
China	INDSTAT	Madagascar	INDSTAT	Swaziland	WDI
Colombia	INDSTAT	Malawi	WDI	Sweden	STAN
Costa Rica	WDI	Malaysia	INDSTAT	Switzerland	STAN
Cte d'Ivoire	WDI	Maldives	WDI	Syria	INDSTAT
Croatia	WDI	Malta	INDSTAT	TFYR of Macedonia	INDSTAT
Cuba	WDI	Mauritania	WDI	Thailand	INDSTAT(int.)
Cyprus	INDSTAT	Mauritius	INDSTAT	Togo	WDI
Czech Rep.	STAN	Mexico	STAN	Trinidad and Tobago	INDSTAT
Denmark	STAN	Morocco	INDSTAT	Tunisia	INDSTAT
Dominica	INDSTAT	Mozambique	WDI	Turkey	INDSTAT
Dominican Rep.	WDI	Namibia	WDI	USA	STAN
Ecuador	INDSTAT	Nepal	WDI	Uganda	WDI
Eritrea	INDSTAT	Netherlands	STAN	Ukraine	INDSTAT
Estonia	STAN	New Zealand	STAN	United Kingdom	STAN
Ethiopia	INDSTAT	Nicaragua	WDI	U. Rep. of Tanzania	INDSTAT
Fiji	INDSTAT	Niger	WDI	Uruguay	INDSTAT
Finland	STAN	Nigeria	INDSTAT	Venezuela	WDI
France	STAN	Norway	STAN	Viet Nam	INDSTAT
Gabon	WDI	Pakistan	INDSTAT(int.)	Zambia	WDI
Gambia	WDI				

Notes: INDSTAT(int.) indicates that output data were interpolated based on INDSTAT data for years before and after 2003.

Table A2: ISIC Rev. 3 Industries

ISIC	Industry Description	HS-6 Codes	Countries	Trade Share
15A	Food, beverages, and tobacco	428	76	6.6%
17	Textiles	541	63	3.3%
18	Wearing apparel; dressing and dyeing of fur	241	48	2.9%
19	Leather, leather products, and footwear	67	57	1.4%
20	Wood products, except furniture	69	75	1.3%
21	Paper and paper products	120	75	2.4%
22	Publishing, printing, reproduction of recorded media	36	78	0.9%
23	Coke, refined petroleum products, nuclear fuel	20	55	2.7%
24	Chemicals and chemical products	879	66	11.8%
25	Rubber and plastics products	121	76	3.0%
26	Non-metallic mineral products	170	79	1.5%
27	Basic metals	359	58	5.4%
28	Fabricated metal products, except mach. and equip.	221	74	2.7%
29	Other machinery and equipment	528	61	10.6%
30	Office, accounting and computing machinery	37	33	5.4%
31	Other electrical machinery and apparatus	134	62	4.7%
32	Radio, television, and communication equipment	101	48	8.5%
33	Medical, precision instruments, watches and clocks	212	53	3.9%
34	Motor vehicles, trailers and semi-trailers	54	56	13.6%
35	Other transport equipment	81	58	4.1%
36	Furniture, other manufacturing	189	64	3.2%

Notes: Column "Countries" lists number of sample countries with output data available for each ISIC industry.

Table A3: Multi-Sector Trade Cost Coefficient Estimates (Fixed-Effects Estimations)

	Food/ Tobacco	Textiles	Apparel	Leather	Wood	Paper	Printing	Fuel	Chemical	Rubber/ Plastics	Minerals	Basic Metal	Fabr. Metal	Other Mach.	Comp. Mach.	Electrical	Comm. Equip.	Medical	Vehicles	Other Transp.	Furniture
	Sector-Level Fixed Effects Log LS																				
mean($\ln \hat{d}_n$)	−2.28	−0.61	−0.87	−0.82	−1.86	−0.47	−2.91	−1.75	−0.54	−0.71	−1.88	−0.82	−1.35	−0.44	−1.46	−0.73	−1.06	−1.30	−1.57	−2.15	−1.97
Distance	−1.73	−1.97	−2.05	−1.78	−2.00	−2.42	−2.11	−2.35	−2.06	−2.18	−2.10	−2.02	−2.12	−1.85	−1.87	−1.97	−1.70	−1.48	−1.85	−1.39	−1.88
	(0.07)	(0.09)	(0.10)	(0.11)	(0.08)	(0.09)	(0.09)	(0.14)	(0.08)	(0.08)	(0.07)	(0.11)	(0.07)	(0.07)	(0.09)	(0.08)	(0.09)	(0.07)	(0.08)	(0.07)	(0.08)
Shared Border	1.32	0.97	0.82	1.32	0.81	0.83	1.07	1.32	0.87	1.08	1.43	0.80	1.11	0.92	0.55	1.00	0.85	0.95	1.06	1.14	0.96
	(0.19)	(0.20)	(0.22)	(0.22)	(0.22)	(0.22)	(0.19)	(0.25)	(0.19)	(0.19)	(0.21)	(0.18)	(0.18)	(0.18)	(0.20)	(0.23)	(0.19)	(0.19)	(0.20)	(0.19)	(0.16)
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Agg}_{PPML}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sector-Level Poisson PML																				
mean($\ln \hat{d}_n$)	−3.49	−1.95	−2.42	−3.21	−3.01	−2.65	−4.58	−2.57	−3.03	−2.61	−3.75	−1.97	−3.35	−2.59	−3.43	−2.75	−2.73	−3.38	−2.68	−4.06	−4.03
Distance	−1.10	−1.27	−1.27	−0.93	−1.26	−1.35	−1.14	−1.59	−1.03	−1.27	−1.23	−1.11	−1.20	−0.88	−1.01	−1.11	−0.81	−0.73	−1.12	−0.33	−0.89
	(0.10)	(0.11)	(0.19)	(0.21)	(0.12)	(0.09)	(0.13)	(0.14)	(0.09)	(0.10)	(0.08)	(0.09)	(0.11)	(0.07)	(0.14)	(0.09)	(0.09)	(0.07)	(0.17)	(0.15)	(0.12)
Shared Border	0.64	0.30	0.43	0.28	0.80	0.46	0.70	0.66	0.28	0.61	0.66	0.68	0.65	0.56	0.03	0.46	0.44	0.24	0.66	1.10	0.86
	(0.18)	(0.15)	(0.19)	(0.21)	(0.17)	(0.16)	(0.23)	(0.26)	(0.11)	(0.15)	(0.12)	(0.14)	(0.14)	(0.12)	(0.18)	(0.16)	(0.21)	(0.13)	(0.23)	(0.27)	(0.15)
	Produt-Level Fixed Effects Log LS																				
mean($\ln \hat{d}_n$)	−3.31	−2.19	−2.54	−2.19	−3.47	−1.81	−3.45	−2.48	−2.34	−1.87	−3.83	−2.09	−2.96	−1.69	−2.07	−1.70	−1.39	−1.97	−1.76	−2.56	−2.98
Distance	−1.14	−1.05	−1.14	−1.12	−1.12	−1.50	−1.50	−1.65	−1.11	−1.44	−1.09	−1.11	−1.22	−1.09	−1.40	−1.29	−1.19	−0.95	−1.34	−0.81	−1.09
	(0.06)	(0.07)	(0.11)	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.05)	(0.06)	(0.05)	(0.07)	(0.06)	(0.05)	(0.07)	(0.06)	(0.06)	(0.04)	(0.07)	(0.05)	(0.06)
Shared Border	1.06	0.62	0.60	0.84	0.99	0.94	0.97	0.94	0.72	0.99	1.04	0.81	0.93	0.64	0.47	0.82	0.69	0.74	0.85	0.83	0.87
	(0.12)	(0.13)	(0.16)	(0.14)	(0.14)	(0.11)	(0.13)	(0.13)	(0.09)	(0.12)	(0.11)	(0.10)	(0.13)	(0.10)	(0.13)	(0.14)	(0.12)	(0.11)	(0.12)	(0.08)	(0.14)
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Pool} = \boldsymbol{\beta}^{j,Pool}_{PPML}$	0.01	0.01	0.44	0.21	0.00	0.07	0.02	0.59	0.11	0.11	0.15	0.00	0.45	0.44	0.00	0.23	0.04	0.01	0.16	0.12	0.80
	Product-Level Poisson PML																				
mean($\ln \hat{d}_n$)	−3.31	−2.19	−2.54	−2.19	−3.47	−1.81	−3.45	−2.48	−2.34	−1.87	−3.83	−2.09	−2.96	−1.69	−2.07	−1.70	−1.39	−1.97	−1.76	−2.56	−2.98
Distance	−1.50	−1.50	−1.45	−1.14	−1.69	−1.72	−1.19	−1.78	−1.13	−1.37	−1.32	−1.51	−1.28	−0.99	−1.08	−1.19	−0.96	−0.73	−1.14	−0.71	−1.06
	(0.09)	(0.10)	(0.17)	(0.18)	(0.09)	(0.09)	(0.12)	(0.14)	(0.10)	(0.10)	(0.09)	(0.08)	(0.10)	(0.08)	(0.14)	(0.09)	(0.08)	(0.06)	(0.16)	(0.12)	(0.13)
Shared Border	0.69	0.36	0.39	0.42	0.77	0.47	0.73	0.62	0.35	0.60	0.80	0.61	0.65	0.55	−0.02	0.45	0.31	0.42	0.59	0.63	0.75
	(0.18)	(0.15)	(0.20)	(0.19)	(0.14)	(0.15)	(0.21)	(0.26)	(0.13)	(0.16)	(0.15)	(0.10)	(0.13)	(0.14)	(0.19)	(0.17)	(0.19)	(0.15)	(0.22)	(0.19)	(0.17)
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.22	0.02	0.00

Notes: Standard errors (in parentheses) are robust to multi-way clustering by both importer and exporter. Parameters reported represent $\hat{b} = -\theta\hat{\beta}$. The implied percentage effect of each coefficient on the ad valorem tariff equivalent trade cost is $100 \times (e^{-\hat{b}/\theta} - 1)$.

Table A4: Multi-Sector Trade Cost Coefficient Estimates (Sector-Level Structural Estimations)

	Food/ Tobacco	Textiles	Apparel	Leather	Wood	Paper	Printing	Fuel	Chemical	Rubber/ Plastics	Minerals	Basic Metal	Fabr. Metal	Other Mach.	Comp. Mach.	Electrical	Comm. Equip.	Medical	Vehicles	Other Transp.	Furniture
	Sector-Level Structural Log Least-Squares																				
mean($\ln \hat{d}_n$)	−1.59	0.03	−0.43	−0.32	−1.74	−0.71	−2.10	−0.76	−0.95	−0.78	−1.66	−0.11	−1.34	−0.62	−2.16	−0.58	−0.53	−1.44	−0.97	−1.06	−2.60
Distance	−1.83	−2.19	−1.95	−1.82	−2.10	−2.25	−2.19	−2.55	−1.94	−2.11	−2.18	−1.84	−2.21	−1.81	−1.69	−2.08	−1.90	−1.62	−1.89	−1.58	−1.93
	(0.12)	(0.22)	(0.21)	(0.18)	(0.13)	(0.14)	(0.17)	(0.14)	(0.11)	(0.11)	(0.10)	(0.13)	(0.11)	(0.10)	(0.12)	(0.13)	(0.13)	(0.12)	(0.14)	(0.20)	(0.12)
Shared Border	1.03	0.43	1.00	1.08	0.04	0.49	0.97	1.12	0.35	0.75	0.88	1.20	0.39	0.54	−0.15	0.46	0.16	0.12	0.85	1.14	−0.39
	(0.38)	(0.69)	(0.67)	(0.53)	(0.51)	(0.37)	(0.53)	(0.32)	(0.37)	(0.29)	(0.29)	(0.32)	(0.33)	(0.35)	(0.48)	(0.46)	(0.55)	(0.59)	(0.40)	(0.66)	(0.44)
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Agg}_{PPML}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	Sector-Level Structural Gamma PML																				
mean($\ln \hat{d}_n$)	−2.63	−1.17	−1.96	−1.79	−2.80	−1.88	−3.47	−2.58	−1.89	−1.84	−2.49	−1.64	−2.17	−1.42	−2.45	−1.61	−1.45	−1.90	−2.13	−2.64	−3.57
Distance	−1.57	−1.66	−1.61	−1.51	−1.75	−1.65	−1.74	−1.85	−1.52	−1.63	−2.14	−1.68	−1.88	−1.46	−1.44	−1.60	−1.37	−1.30	−1.33	−0.66	−1.50
	(0.09)	(0.08)	(0.10)	(0.17)	(0.10)	(0.11)	(0.08)	(0.11)	(0.08)	(0.17)	(0.09)	(0.10)	(0.10)	(0.13)	(0.09)	(0.08)	(0.11)	(0.10)	(0.10)	(0.18)	(0.07)
Shared Border	0.28	0.16	−0.03	0.10	−0.36	0.56	0.26	−0.08	0.26	0.54	−0.30	−0.51	0.12	0.38	0.10	0.39	0.37	0.36	0.77	1.74	−0.38
	(0.29)	(0.31)	(0.33)	(0.29)	(0.28)	(0.23)	(0.20)	(0.43)	(0.21)	(0.36)	(0.24)	(0.37)	(0.26)	(0.31)	(0.25)	(0.20)	(0.20)	(0.27)	(0.22)	(0.44)	(0.19)
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Agg}_{PPML}$	0.00	0.00	0.13	0.06	0.00	0.00	0.00	0.22	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.32	0.00	0.00
	Sector-Level Structural Poisson PML																				
mean($\ln \hat{d}_n$)	−3.49	−1.95	−2.42	−3.21	−3.01	−2.65	−4.58	−2.57	−3.03	−2.61	−3.75	−1.97	−3.35	−2.59	−3.43	−2.75	−2.73	−3.38	−2.68	−4.06	−4.03
Distance	−1.10	−1.27	−1.27	−0.93	−1.26	−1.35	−1.14	−1.59	−1.03	−1.27	−1.23	−1.11	−1.20	−0.88	−1.01	−1.11	−0.81	−0.73	−1.12	−0.33	−0.89
	(0.10)	(0.11)	(0.19)	(0.21)	(0.12)	(0.09)	(0.13)	(0.14)	(0.09)	(0.10)	(0.08)	(0.09)	(0.11)	(0.07)	(0.14)	(0.09)	(0.09)	(0.07)	(0.17)	(0.15)	(0.12)
Shared Border	0.64	0.30	0.43	0.28	0.80	0.46	0.70	0.66	0.28	0.61	0.66	0.68	0.65	0.56	0.03	0.46	0.44	0.24	0.66	1.10	0.86
	(0.18)	(0.15)	(0.19)	(0.21)	(0.17)	(0.16)	(0.23)	(0.26)	(0.11)	(0.15)	(0.12)	(0.14)	(0.14)	(0.12)	(0.18)	(0.16)	(0.21)	(0.13)	(0.23)	(0.27)	(0.15)
	Sector-Level Structural Nonlinear Least-Squares																				
mean($\ln \hat{d}_n$)	−3.67	−1.89	−3.00	−1.38	−3.50	−2.61	−4.75	−3.22	−3.23	−2.31	−3.90	−2.06	−3.14	−2.65	−3.39	−2.49	−2.72	−3.42	−2.30	−4.24	−3.56
Distance	−1.12	−1.31	−0.96	−1.82	−0.88	−1.45	−1.14	−1.12	−1.01	−1.51	−1.20	−1.06	−1.34	−0.82	−1.15	−1.18	−0.71	−0.78	−1.35	−0.13	−1.08
	(0.11)	(0.10)	(0.02)	(0.35)	(0.24)	(0.10)	(0.17)	(0.17)	(0.07)	(0.26)	(0.11)	(0.08)	(0.25)	(0.07)	(0.29)	(0.19)	(0.13)	(0.12)	(0.12)	(0.21)	(0.27)
Shared Border	0.36	0.24	0.73	−0.34	1.43	0.21	0.53	1.12	0.12	0.29	0.57	0.71	0.46	0.65	−0.26	0.56	0.75	0.04	0.50	1.54	0.80
	(0.23)	(0.14)	(0.26)	(0.08)	(0.41)	(0.19)	(0.25)	(0.51)	(0.08)	(0.24)	(0.16)	(0.19)	(0.22)	(0.18)	(0.27)	(0.33)	(0.37)	(0.21)	(0.22)	(0.41)	(0.11)
$\mathrm{H}_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Agg}_{PPML}$	0.57	0.93	0.21	0.00	0.04	0.64	0.75	0.03	0.51	0.43	0.80	0.91	0.31	0.93	0.58	0.64	0.53	0.79	0.47	0.68	0.25

Notes: Standard errors (in parentheses) are robust to multi-way clustering by both importer and exporter. Parameters reported represent $\hat{b} = -\theta\hat{\beta}$. The implied percentage effect of each coefficient on the ad valorem tariff equivalent trade cost is $100 \times (e^{-\hat{b}/\theta} - 1)$.

Table A5: Multi-Sector Trade Cost Coefficient Estimates (Product-Level Structural Estimations)

	Food/ Tobacco	Textiles	Apparel	Leather	Wood	Paper	Printing	Fuel	Chemical	Rubber/ Plastics	Minerals	Basic Metal	Fabr. Metal	Other Mach.	Comp. Mach.	Electrical	Comm. Equip.	Medical	Vehicles	Other Transp.	Furniture
	Product-Level Structural Log Least-Squares																				
mean($\ln \hat{d}_n$)	−2.82	−1.59	−2.29	−2.02	−2.87	−1.31	−3.07	−2.16	−1.81	−1.56	−3.26	−1.41	−2.47	−1.18	−2.21	−1.28	−1.16	−1.77	−1.33	−1.93	−2.64
Distance	−1.39	−1.34	−1.28	−1.27	−1.39	−1.74	−1.59	−1.80	−1.33	−1.55	−1.36	−1.40	−1.42	−1.33	−1.39	−1.45	−1.29	−1.08	−1.52	−1.08	−1.27
	(0.08)	(0.09)	(0.13)	(0.10)	(0.10)	(0.07)	(0.08)	(0.12)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	(0.06)	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.09)
Shared Border	0.91	0.52	0.43	0.58	0.90	0.90	1.22	1.02	0.74	1.09	0.97	0.81	0.91	0.64	0.32	0.91	0.71	0.62	0.93	0.80	0.76
	(0.13)	(0.13)	(0.19)	(0.16)	(0.16)	(0.13)	(0.19)	(0.17)	(0.13)	(0.13)	(0.11)	(0.12)	(0.14)	(0.10)	(0.19)	(0.14)	(0.16)	(0.14)	(0.16)	(0.13)	(0.15)
$H_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$H_0 : \boldsymbol{\beta}^{j,Pool} = \boldsymbol{\beta}^{j,Pool}_{PPML}$	0.47	0.08	0.48	0.18	0.01	0.00	0.00	0.09	0.00	0.00	0.25	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09
	Product-Level Structural Gamma PML																				
mean($\ln \hat{d}_n$)	−3.66	−1.32	−2.81	−2.05	−2.96	−2.46	−3.69	−3.71	−2.85	−2.14	−2.65	−2.91	−2.27	−1.75	−2.44	−1.58	−1.66	−0.54	−1.75	−1.17	−2.14
Distance	−1.04	−1.67	−0.99	−1.50	−1.53	−1.57	−1.57	−1.23	−1.08	−1.43	−1.82	−0.82	−1.82	−2.14	−1.58	−1.33	−1.12	−1.37	−1.65	−1.50	−1.25
	(0.18)	(0.06)	(0.13)	(0.17)	(0.12)	(0.16)	(0.13)	(0.26)	(0.16)	(0.10)	(0.21)	(0.17)	(0.11)	(0.21)	(0.12)	(0.09)	(0.10)	(0.19)	(0.12)	(0.11)	(0.13)
Shared Border	0.90	−0.12	0.74	−0.17	0.37	−0.10	0.45	0.43	0.20	0.67	0.41	0.68	−0.02	−2.02	−0.43	0.86	0.54	1.37	0.03	0.56	1.45
	(0.23)	(0.11)	(0.26)	(0.42)	(0.19)	(0.24)	(0.27)	(0.39)	(0.38)	(0.32)	(0.48)	(0.53)	(0.29)	(0.65)	(0.50)	(0.26)	(0.20)	(0.49)	(0.26)	(0.24)	(0.48)
$H_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	0.03	0.63	0.00	0.59	0.08	0.00	0.42	0.09	0.00	0.45	0.31	0.00	0.51	0.00	0.53	0.07	0.11	0.03	0.03	0.00	0.00
$H_0 : \boldsymbol{\beta}^{j,Pool} = \boldsymbol{\beta}^{j,Pool}_{PPML}$	0.05	0.02	0.09	0.30	0.00	0.00	0.07	0.08	0.75	0.87	0.07	0.00	0.00	0.00	0.01	0.08	0.17	0.00	0.01	0.00	0.07
	Product-Level Structural Poisson PML																				
mean($\ln \hat{d}_n$)	−2.73	−1.35	−1.91	−2.50	−2.28	−1.65	−4.40	−2.43	−2.60	−2.32	−3.48	−1.31	−3.03	−2.01	−3.14	−2.26	−2.25	−2.80	−2.43	−3.00	−3.18
Distance	−1.50	−1.50	−1.45	−1.14	−1.69	−1.72	−1.19	−1.78	−1.13	−1.37	−1.32	−1.51	−1.28	−0.99	−1.08	−1.19	−0.96	−0.73	−1.14	−0.71	−1.06
	(0.09)	(0.10)	(0.17)	(0.18)	(0.09)	(0.09)	(0.12)	(0.14)	(0.10)	(0.10)	(0.09)	(0.08)	(0.10)	(0.08)	(0.14)	(0.09)	(0.08)	(0.06)	(0.16)	(0.12)	(0.13)
Shared Border	0.69	0.36	0.39	0.42	0.77	0.47	0.73	0.62	0.35	0.60	0.80	0.61	0.65	0.55	−0.02	0.45	0.31	0.42	0.59	0.63	0.75
	(0.18)	(0.15)	(0.20)	(0.19)	(0.14)	(0.15)	(0.21)	(0.26)	(0.13)	(0.16)	(0.15)	(0.10)	(0.13)	(0.14)	(0.19)	(0.17)	(0.19)	(0.15)	(0.22)	(0.19)	(0.17)
$H_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.22	0.02	0.00
	Product-Level Structural Nonlinear Least-Squares																				
mean($\ln \hat{d}_n$)	−3.61	−1.01	−1.73	−0.18	−2.59	−1.80	−4.59	−3.15	−3.38	−2.05	−3.46	−0.93	−2.57	−2.06	−3.29	−1.70	−2.26	−2.49	−2.71	−1.81	−3.24
Distance	−1.28	−1.63	−1.41	−2.34	−1.59	−1.72	−1.25	−1.26	−0.86	−1.62	−1.34	−1.73	−1.51	−0.90	−1.08	−1.44	−0.89	−0.89	−0.98	−1.06	−0.87
	(0.31)	(0.11)	(0.08)	(0.25)	(0.16)	(0.19)	(0.19)	(0.14)	(0.15)	(0.24)	(0.06)	(0.20)	(0.24)	(0.10)	(0.23)	(0.31)	(0.12)	(0.17)	(0.20)	(0.53)	(0.48)
Shared Border	0.23	0.42	0.75	−0.22	0.65	0.27	0.33	1.11	0.23	0.22	0.77	0.47	0.53	0.76	−0.19	0.39	0.48	0.33	0.69	1.01	1.25
	(0.37)	(0.29)	(0.39)	(0.14)	(0.63)	(0.26)	(0.32)	(0.55)	(0.18)	(0.22)	(0.23)	(0.36)	(0.19)	(0.27)	(0.23)	(0.46)	(0.32)	(0.39)	(0.43)	(0.46)	(0.39)
$H_0 : \boldsymbol{\beta}^{j,Agg} = \boldsymbol{\beta}^{j,Pool}$	1.00	0.00	0.00	0.00	0.00	0.00	0.25	0.23	0.53	0.06	0.00	0.00	0.00	0.01	0.75	0.00	0.19	0.00	0.00	0.17	0.40
$H_0 : \boldsymbol{\beta}^{j,Pool} = \boldsymbol{\beta}^{j,Pool}_{PPML}$	0.00	0.32	0.42	0.00	0.34	0.66	0.42	0.22	0.11	0.16	1.00	0.45	0.23	0.91	0.92	0.13	0.78	0.70	0.64	0.00	0.51

Notes: Standard errors (in parentheses) are robust to multi-way clustering by both importer and exporter. Parameters reported represent $\hat{b} = -\theta\hat{\beta}$. The implied percentage effect of each coefficient on the ad valorem tariff equivalent trade cost is $100 \times (e^{-\hat{b}/\theta} - 1)$.

B Proofs

Proof of Proposition 1. Summing (1) over k, invoking Assumption 2, and multiplying and dividing by the terms X_n, Φ_n, and $\bar{T}_i$ yields the following expression:

$$\begin{aligned} X_{ni} &= \frac{\bar{T}_i d_{ni}^{-\theta}}{\Phi_n} X_n \sum_k (d_n^k d_i^{*k})^{-\theta} \frac{X_n^k}{X_n} \frac{\Phi_n}{\Phi_n^k} \frac{T_i^k}{\bar{T}_i} \\ &\equiv \frac{\bar{T}_i d_{ni}^{-\theta}}{\Phi_n} X_n \tilde{T}_{ni}. \end{aligned} \tag{11}$$

Imposing the market-clearing condition $Y_i = \sum_n X_{ni}$ implies that

$$\begin{aligned} Y_i &= \bar{T}_i \sum_n \frac{X_n}{\Phi_n} d_{ni}^{-\theta} \tilde{T}_{ni} \\ &\equiv \bar{T}_i \Psi_i. \end{aligned}$$

Substituting this into (11) yields (4). Given that (4) holds, imposing the market-clearing condition $X_n = \sum_n X_{ni}$, along with the definition of Ψ_i, implies that (5) must hold. Finally, note that these results hold for any positive values of $\bar{T}_i$. □

Proof of Proposition 2. Without loss of generality, we can define d_{ni} such that $\bar{d}_n = \bar{d}_i^* = 1$. If $T_i^k = T_i T^k$ for all i and k, then $\Phi_n^k = T^k \sum_i T_i d_{ni}^{-\theta} \equiv T^k \Phi_n'$. Because $\bar{T}_i$ is a homogeneous-of-degree-1 function of T_i^k, it takes the following form:

$$\bar{T}_i = T_i \bar{T}.$$

Together, these results imply that

$$\tilde{T}_{ni} = \frac{\Phi_n}{\bar{T}\Phi_n'}.$$

This implies that

$$X_{ni} = \frac{T_i d_{ni}^{-\theta}}{\bar{T}\Phi_n'} X_n.$$

Imposing the market-clearing conditions $Y_i = \sum_n X_{ni}$ and $X_n = \sum_n X_{ni}$ implies that $\bar{T}\Phi_n' = \Phi_n$, and $\tilde{T}_{ni} = 1$, for all n and i.[41] □

The proofs of Propositions 3 and 4 make use of the following lemma:

[41]Note that Φ_n and Ψ_i are only defined up to a scalar multiple, so the equality $\bar{T}\Phi_n' = \Phi_n$ implicitly assumes that the same normalization is used for both terms.

Lemma 1. *Given Assumptions 1 and 1′,*

$$\Phi_n^{\frac{\sigma-1}{\theta}} = \sum_k (\Phi_n^k)^{\frac{\sigma-1}{\theta}},$$

and

$$\frac{X_n^k}{X_n} = \left(\frac{\Phi_n^k}{\Phi_n}\right)^{\frac{\sigma-1}{\theta}}.$$

Proof. □

Proof of Proposition 3. If $d_{ni}^k = 1$, for all n, i, and k, then (1) reduces to

$$X_{ni}^k = \frac{T_i^k}{\Phi^k},$$

where $\Phi^k = \sum_i T_i^k$. Given Assumption 1′, by Lemma 1, (6) reduces to

$$\begin{aligned}\tilde{T}_{ni} &= \sum_k \left(\frac{\Phi^k}{\Phi}\right)^{\frac{\sigma-1}{\theta}-1} \frac{T_i^k}{\bar{T}_i} \\ &\equiv \tilde{T}_i,\end{aligned}$$

where $\Phi^{\frac{\sigma-1}{\theta}} = \sum_k (\Phi^k)^{\frac{\sigma-1}{\theta}}$. This implies that

$$\begin{aligned}\Psi_i &= \tilde{T}_i \sum_n \frac{X_n}{\Phi} \\ &\equiv \tilde{T}_i \Psi_i'.\end{aligned}$$

Substituting this into (4) and (5) yields the expression in Proposition 3. □

Proof of Proposition 4. The expression in Proposition 4 obtains directly by totally differentiating (6), holding constant all values of X_n, T_i^k, and d_n^k. □

C Data

C.1 Trade Data

Bilateral, product-level trade data are from the U.N. Comtrade database. The data are classified into six-digit Harmonized System (HS), 1996 revision, product codes. The sample consists of trade flows for the year 2003, which was chosen to maximize the number of countries for which both gross output data from INDSTAT and trade data from Comtrade were available. The sample consists of trade flows reported by exporters because these values are more likely to be consistent with the gross output data, which is reported by the producing country, and because exports are typically reported "free on board", as opposed to "cost, insurance, and freight", and the former is consistent with the measure of trade flows in the model.

The trade flow data were combined with manufacturing gross output data from several sources. The manufacturing output data are classified according Revision 3 of the International Standard Industrial Classification (ISIC). To match the trade and output data, the HS1996 codes were mapped to ISIC (Revision 3) codes using the concordance available from the U.N. Statistics Division.[42] All HS codes not mapped to manufacturing ISIC codes (2-digit industries 15-37) were dropped. This reduced the number of HS codes in the sample to 4,608.

C.2 Gravity Variables

The bilateral relationship variables used to estimate trade costs are from the Gravity dataset available from CEPII (see Mayer and Zignago, 2011). The estimations use the following variables: population-weighted distance (*distw*), whether countries share a common border (*contig*), whether they have a common official language (*comlang_off*), whether they have ever had a colonial link (*colony*), whether they are currently parties to a regional trade agreement (*rta*), and whether the share a common currency (*comcur*).

C.3 Manufacturing Output

Gross manufacturing output data come from three sources. First, the data are taken from the OECD STAN database, where available. If countries are not included in this database, data are from the Industrial Statistics Database (INDSTAT4), 2011 Edition, CD-ROM published by UNIDO. Where data are available for years before and after, but not including, 2003, log output is linearly interpolated based on the closest values before and after 2003. Where data are not available from either sources, output is imputed from total manufacturing value added obtained from the World Development Indicators database of the World Bank. Gross output is obtained by scaling value added by a factor of 3.04.[43]

Gross output data at the 2-digit ISIC (Revision 3) level were obtained from the INDSTAT2, 2014 Edition, CD-ROM published by UNIDO. Where countries reported data in

[42] This is available for free download from the following url:
http://unstats.un.org/unsd/cr/registry/regdntransfer.asp?f=183.

[43] This value is obtained from a cross-sectional regression of gross output on value added, omitting the constant term. The regression R^2 was equal to 0.99.

combined or aggregated categories ISIC categories, these observations were excluded. Table A2 lists the ISIC categories, their descriptions, the number of 6-digit HS codes matched to each ISIC industry, the number of countries that reported output data in each industry, and the industry's share in total world manufacturing Trade.

C.4 Constructing the Sample

To be included in the sample, data must be available for a country from the Comtrade database and at least one of the STAN, INDSTAT, or WDI databases. To avoid problems related to entrepot trade, China, Hong Kong, and Macao are merged into a single country. There were also several other cases in which there were apparent problems of entrepot trade – i.e. reported exports exceeded reported gross output – which resulted in 8 countries being dropped from the sample.[44] Once the trade and manufacturing data were merged, domestic absorbtion of domestic manufacturing output, X_{ii}, was then calculated as total manufacturing output minus total manufacturing exports to all countries (including non-reporters), and total manufacturing absorbtion, X_i, was calculated as X_{ii} plus total imports from countries in the sample, yielding an internally consistent bilateral trade flow matrix. For the industry-level sample, values of X_{ii} that were computed to be less than zero were excluded. The final sample consists of total gross manufacturing output and bilateral trade flows for 130 countries and 4,608 6-digit manufacturing HS products.[45]

C.5 Gross Domestic Product

Real GDP per worker is taken from the Penn World Tables (version 7.1).

[44] The excluded countries are Armenia, Belgium, the Federated States of Micronesia, Guyana, Luxembourg, Mali, Mongolia, and Singapore.

[45] Note that the industry-level estimations are based on the full sample of 130 countries. The lack of industry-level output only reduces the number of border cost parameters ($\bar{d}_n$) that can be identified.