The Composition of Exports and Gravity

Scott French*

December, 2012
Version 3.0

Abstract

Gravity estimations using aggregate bilateral trade data implicitly assume that the effect of trade barriers on trade flows is independent of the composition of those flows. However, I show that, in a simple framework which is consistent with generalizations of the wide class of trade models that imply an aggregate gravity equation, aggregate trade flows, in general, depend on the composition of countries' output and expenditure across products, which varies across countries in meaningful ways in the data. This implies that trade cost estimates based on aggregate data are biased and that the predictions of models based on such estimates may be misleading.

I develop a procedure to estimate a bilateral trade cost function using product-level data, which accounts for the lack of comparably disaggregated data on domestic output. This technique leads to trade cost estimates that are smaller and much more robust to distributional assumptions than estimates obtained from aggregate data, implying that failure to account for the composition of output and expenditure is a more important cause of bias than failure to properly account for heteroskedasticity. Compared to a more traditional model based on aggregate data, the model based on product-level data predicts domestic trade shares that are more consistent with the data, a smaller effect of reducing trade cost asymmetry on income differences, and very different effects of the growth of Chinese manufacturing on many countries.

*School of Economics, University of New South Wales. scott.french@unsw.edu.au. I am thankful to Dean Corbae for his guidance and support. I am also thankful for advice and comments from Kim Ruhl, Jason Abrevaya, Alan Woodland, Russell Hillberry as well as seminar participants at the University of New South Wales, University of Sydney, University of Melbourne, Monash University, and Australian National University. All errors are my own.

1 Introduction

The gravity model – which relates to bilateral trade flows to the sizes of a pair of countries and the barriers to trade that exist between them – has long been celebrated as a parsimonious yet empirically successful way to describe bilateral trade flows. It is extremely useful as a framework within which to estimate the effect of factors that determine barriers to trade on trade flows and to predict the effects of altering these factors. Since Anderson (1979), who showed that the empirical relationship is theoretically founded, it has also been useful as method to quantify trade models, allowing for serious, general equilibrium analysis of the effects of such factors on economic outcomes and welfare. Quite often, the variables of interest are aggregate country-level or bilateral quantities, and the data that is most readily available are also quite aggregated, leading researchers to estimate the parameters of gravity models using aggregate data. This practice implicitly assumes that the effect of trade barriers on trade flows is independent of the composition of those trade flows.

In this paper, I show that this assumption is not borne out in the data, which has important implications for the estimation of trade barriers and the predictions of models based on such estimates. I first develop a simple framework – which is consistent with generalizations of the wide class of trade models that imply an aggregate gravity equation – in which countries choose how to allocate their expenditure across product categories as well as across source countries for each type of product. I show that, in general, aggregate bilateral trade flows depend on the composition of countries' output and expenditure across products in addition to their overall levels of output and expenditure and bilateral trade costs, meaning that aggregate trade flows are not consistent with an aggregate gravity model. Intuitively, if a country's exports are concentrated in a set of goods for which a given importer buys a large fraction from other sources, a trade barrier between the two countries only affects the distribution of the importer's expenditure across countries within those product categories, so the effect depends on the elasticity of substitution across varieties within product categories. However, if a country is the sole provider of a large fraction of the products that it exports to a given importer, then the effect of a trade barrier between the two countries is governed by the elasticity of substitution across product categories. If varieties within product categories are more substitutable than those across product categories, then the former exporter is more effected by a given trade barrier than the latter.

I develop an index based on product-level trade flow data that indicates the degree to which a trade flow between a given pair of countries is similar to either the former or the latter of the scenarios discussed above and show that there is a great deal of heterogeneity in this index, indicating that the effect of trade barriers on aggregate trade flows varies greatly across the set of bilateral country pairs. This implies that bilateral trade barriers cannot be inferred from aggregate trade data because their effects on trade flows depend on interactions among countries' distributions of output and expenditure. As a result, trade barriers must be estimated using product-level trade data, but because output data is rarely available at such a level of disaggregation, traditional estimation methods, which rely on such data to identify the costs associated with national borders, cannot be used. Instead, I develop an estimation procedure based on a reformulation of the model that allows the estimation of trade barriers using only product-level data on international trade flows and aggregate data on domestic trade flows. Estimating trade costs in a way that is consistent with the model implies estimates of trade costs that are lower than those based on aggregate trade data and much more consistent across distributional assumptions for the error term, indicating that the bias in trade cost estimates due to ignoring the composition of trade flows is much more severe than that due to using techniques that are not robust to varying forms of heteroskedasticity.

Given the estimates of trade costs based on product-level data, I show that the composition of trade flows is quite important in explaining their magnitude. The effect of the interaction of countries' distributions of output and expenditure can more than double or halve the effect of trade barriers on trade flows between a pair of countries. And, overall, 37% of the variation in trade flows predicted by the model is due to the effect of composition.

With confidence in the trade cost estimates and the predictive power of the model, I go on to explore the implications of the composition of countries' output and expenditure on trade flows by performing several counterfactual experiments. To do so, I use the results of the estimation to parameterize a version of the model of Eaton and Kortum (2002), which is generalized to allow for differences in average productivity across product categories. I show that while a small, uniform reduction in trade barriers has a similar effect in a model based on product-level data and one based on aggregate data, the removal of the asymmetric component of trade costs leads to very different predictions in the two models. While the aggregate model predicts major gains for small, developing countries and a 17% reduction in the 90/10 ratio of real output per worker across countries, the product-level model predicts much more equal gains and just over a 1% fall in the ratio.

I also examine differences in the effects of changes in technology – which influence countries' distributions of output and expenditure across products – examining the effect of the growth of China as a major exporter of manufactured goods and finding that the product-level model makes very different prediction for many countries. Not surprisingly, the aggregate model predicts that effects of the growth of China are almost universally positive and heavily dependant on geography; the countries closest to China gain the most because they have a larger country to trade with and can do so without facing large trade barriers. The predictions of the product-level model, on the other hand, make clear the importance of accounting for the composition of output and expenditure in such an exercise. It predicts that many of the nearby countries – as well others, including many Central American countries – do not benefit nearly as much, as they produce a similar mix of products as China, for which they lose market share, making their exports more responsive to trade barriers. In fact, real output in Cambodia and Honduras is predicted to decrease. In addition, the developed countries of Europe and North America experience larger gains, as the prices of the products they tend to import fall, and many South American and Sub-Saharan African countries benefit more as demand for the basic materials they tend to export rises.

In addition to a shift in productivity for one country, I consider the effects of the removal of differences in relative productivity across countries. The model predicts large losses in real output for most countries, an average fall of 15% in the case of imposing the US pattern of productivity on the world. However, these magnitudes depend heavily on the correlation between the distribution of productivity across products in each country and exogenous determinants of demand, indicating that one may be influential in the determination of the other. Finally, I compare the predictions of the two models of the effect of removing trade imbalances, showing the product-level model predicts much larger changes in incomes as a result of the change.

This paper is related to several strands of the international trade literature. First, it is related to a large literature using theoretically founded gravity models to study the effects of trade barriers, including Anderson and van Wincoop (2003); Eaton and Kortum (2002); and Helpman et al. (2008). Recent papers have offered extensions to these models which resolve discrepancies between more traditional gravity models and the data, including Waugh (2010) and Fieler (2010).[1] This paper contributes to this literature by showing that that the effect of trade barriers on aggregate trade flows and other macroeconomic variables depends heavily on the composition of output and expenditure and by providing a tractable framework that is consistent with these models in which to study this effect.

This paper is also related to a number of other papers that address issues related to aggregation bias in the estimation of trade costs. Anderson and van Wincoop (2004) point out that estimates based on aggregate data can be biased if both the elasticity of substitution and bilateral trade costs vary across products and if the two are correlated. Hillberry (2002) shows that a similar form of aggregation bias can exist due to the specialization of countries according to comparative advantage driven by relative trade costs. This paper is largely complimentary to these studies. I show that even if trade costs and the elasticity of substitution within product categories do not differ across prod-

[1] Anderson and van Wincoop (2004) provide a survey of older papers that have extended the basic gravity framework in a number of dimensions.

ucts – so that these forms of aggregation bias do not exist – aggregate trade flows still depend on the composition of countries' output and expenditure, biasing estimates of trade costs using aggregate data. More related to this paper is Hillberry and Hummels (2002), who show that, even without variation in trade costs across industries, aggregate estimates of trade costs are biased due to the endogenous co-location of producers and suppliers of intermediate goods to minimize trade costs. In this paper, I take as given the patterns of product-level output and expenditure in the data, estimate trade costs in a model that is consistent with these patterns, and show how taking them into account has implications for the effects of trade costs and other variables on macroeconomic outcomes.

This is not the first paper to estimate trade barriers and parameterize a general equilibrium trade model using disaggregated data. Papers such as Costinot et al. (2012); Anderson and Yotov (2011); and Levchenko and Zhang (2011) perform estimations based on industry-level data for many countries. Das et al. (2007); Eaton et al. (2011); Hillberry and Hummels (2002); and many others have performed estimations using firm-level data for a single country. And, Hanson and Robertson (2010) use data from a small subset of countries and product categories. However, most stop short of disaggregating past the industry level or using data for a large number of countries and products. Presumably this is due to data limitations and computational burden. However, in this paper, I show that trade costs can be consistently estimated using the full set of available product-level data – for thousands of products and well over 100 countries – accounting for the lack of correspondingly disaggregated production data, and requiring no more computing power than is available on a modern laptop computer.

The next section develops the theoretical framework and shows that aggregate trade flows, in general, depend on the composition of countries' output and expenditure. Section 3 takes a brief look at the data, developing and computing the Elasticity Index that measures the degree to which trade between country pairs depends on the elasticity of substitution within or across product categories. Section 4 develops the estimation procedure and presents the results. Section 5 presents the results of the counterfactual experiments, and the final section concludes.

2 Model

The theoretical framework is closely related to the sector-level gravity model detailed in Anderson and van Wincoop (2004), which outlines the class of models which yield a gravity structure. As in that framework, the structure developed here is implied by a large class of underlying international trade models, which encompasses generalizations of several models that have become the workhorses of the literature.[2] The model implies bilateral trade flows that follow a gravity equation at the product level – which is meant to be analogous to a product category as defined by the agencies that collect international trade statistics. It departs from Anderson and van Wincoop (2004) by explicitly modeling the allocation of countries' total expenditure across products, which I then show implies that, in general, aggregate bilateral trade flows depend on the composition, and not simply the level, of countries' output and expenditure.

2.1 Environment

The world is made up of N countries that each produce and consume varieties of a finite number, J, of product categories. Each country, i, produces (or is endowed with) a nominal value of varieties of product j given by $Y_i^j \geq 0$, which is taken as exogenous. Each country, n, also distributes its aggregate nominal expenditure, X_n – also taken as exogenous – across product categories and source countries according to a nested-CES demand structure.[3] Specifi-

[2] It is trivially generated by an Armington model, as in Anderson and van Wincoop (2003), in which countries are each endowed with a unique variety of each product. It is also straightforward to show that it is generated by models of monopolistic competition with homogeneous firms, such as Helpman and Krugman (1987). The appendix shows that generalizations of the Ricardian model of Eaton and Kortum (2002) and the heterogeneous-firm monopolistic competition model of Chaney (2008) also imply the same structure.

[3] X_i and Y_i are allowed to differ from one another, meaning that trade may not be balanced.

cally, the nominal value of expenditure by country n on product j from source country i is

$$X_{ni}^j = \left(\frac{p_i^j d_{ni}}{P_n^j}\right)^{-\theta} X_n^j, \tag{1}$$

where X_n^j is the nominal value of expenditure by country n on all varieties of product j, given by

$$X_n^j = \beta_n^j \left(\frac{P_n^j}{P_n}\right)^{-\sigma} X_n. \tag{2}$$

Here p_i^j is the appropriate source country (or factory gate) price index for varieties of product j originating in country i. Barriers to trade, represented by $d_{ni} > 1$ take the standard "iceberg" form, meaning that for one unit of a variety to arrive in n, d_{ni} units must be shipped from i. The reduced-form elasticities of substitution across and within product categories are $1 + \sigma$ and $1 + \theta$, respectively, where the assumption $\theta \geq \sigma > 0$ is maintained, which implies that varieties within product categories are more substitutable than those across product categories.[4,5]

The product-level and aggregate destination-specific CES price indexes are, respectively,

$$P_n^j = \left(\sum_i (p_i^j d_{ni})^{-\theta}\right)^{-\frac{1}{\theta}} \tag{3}$$

and

$$P_n = \left(\sum_j \beta_n^j (P_n^j)^{-\sigma}\right)^{-\frac{1}{\sigma}}. \tag{4}$$

The parameter $\beta_n^j \geq 0$, $\sum_j \beta_n^j = 1$, allows expenditure on a particular product category to differ across countries due to factors not explicitly modeled.

[4] I refer to the elasticities of substitution as "reduced form" because, depending on the underlying framework that implies this framework, the deep parameters that underlie these parameters may have other interpretations.

[5] As there are a finite number of varieties and product categories, this assumption is not strictly necessary for the analysis of this paper. However, I maintain it because this specification is a reduced form of a wide class of underlying models, and it helps to fix the intuition for the results that follow.

2.2 Product-Level Gravity

The market clearing condition, $Y_i^j = \sum_n X_{ni}^j$, (1), and (3) imply that the set of source country prices can be expressed as

$$(p_i^j)^{-\theta} = \frac{Y_i^j}{Y^j}\left(\frac{1}{\Pi_i^j}\right)^{-\theta}, \tag{5}$$

where $Y^j = \sum_i Y_i^j$ and $(\Pi_i^j)^{-\theta} = \sum_n \left(\frac{d_{ni}}{P_n^j}\right)^{-\theta}\frac{X_n^j}{Y^j}$. Substituting this expression into (1), it can be seen that bilateral, product-level trade flows are given by the following system:

$$X_{ni}^j = \frac{X_n^j Y_i^j}{Y^j}\left(\frac{d_{ni}}{P_n^j \Pi_i^j}\right)^{-\theta} \tag{6}$$

$$(P_n^j)^{-\theta} = \sum_i \left(\frac{d_{ni}}{\Pi_i^j}\right)^{-\theta}\frac{Y_i^j}{Y^j} \tag{7}$$

$$(\Pi_i^j)^{-\theta} = \sum_n \left(\frac{d_{ni}}{P_n^j}\right)^{-\theta}\frac{X_n^j}{Y^j}. \tag{8}$$

Anderson and van Wincoop (2004) refer to the indexes P_n^j and Π_i^j as inward and outward multilateral resistance, which are functions of the set of bilateral trade barriers, $\{d_{ni}\}$, and levels of output and expenditure, $\{X_n^j, Y_i^j\}$. These terms summarize all the general equilibrium forces which affect the volume of trade between a pair of countries. Intuitively, a high value of P_n^j – the domestic price index for product j – implies that it is relatively difficult for consumers in country n to obtain varieties of good j, implying that n will import a relatively large volume of j from a given source country, i, all else equal. Likewise, a high value of Π_i^j implies that it is relatively difficult for producers in i to sell their varieties of j – either because they face high trade barriers or stiff competition – low values of P_n in potential destinations – implying that exports to a particular destination, n, will be relatively high, all else equal.

2.3 Aggregate Trade Flows

Equation (6) is a standard theoretical gravity equation. So, given data on product-level output, expenditure, and bilateral trade flows, any further analysis could proceed using standard techniques. However, the typical strategy in the literature is to consider the world to be a one-sector economy so that aggregate data can be brought to bear on (6).[6] The remainder of this section considers the implications of such a practice when non-trivial differences in output and expenditure exist across products.

2.3.1 An Aggregate Gravity Equation

Summing X_{ni}^j over all j leads to the following expression for total spending by country n on all products from country i:

$$X_{ni} = \frac{X_n Y_i}{Y} \left(\frac{d_{ni}}{P_n \tilde{\Pi}_{ni}} \right)^{-\theta}, \tag{9}$$

where

$$(\tilde{\Pi}_{ni})^{\theta} = \sum_j (\Pi_i^j)^{\theta} \frac{Y_i^j / Y_i}{Y^j / Y} \beta_n^j \left(\frac{P_n^j}{P_n} \right)^{\theta - \sigma}, \tag{10}$$

and where $Y_i = \sum_j Y_i^j$, $Y = \sum_i Y_i$, and P_n, defined in (4), can also be given by

$$(P_n)^{-\theta} = \sum_i \left(\frac{d_{ni}}{\tilde{\Pi}_{ni}} \right)^{-\theta} \frac{Y_i}{Y}.$$

Equation (9) has a gravity-like structure similar to (6), relating aggregate bilateral trade flows to aggregate output and expenditure and bilateral trade costs. Unlike the product-level gravity equation, though, the aggregate equation features an outward multilateral resistance term, $\tilde{\Pi}_{ni}$, which varies by destination as well as source country. From (10), it is clear that the value of

[6] While most studies use aggregate trade data, there are many that use disaggregated data, for example Hummels (1999), Anderson and Yotov (2011), and Hanson and Robertson (2010). However, these papers typically select a few countries and products or use data disaggregated to at most a couple dozen sectors, which still masks great deal of heterogeneity across thousands of product categories.

$\tilde{\Pi}_{ni}$ depends product-level output and multilateral resistance variables, meaning that aggregate trade flows cannot be expressed as a function of aggregate data alone.

This new term is an index of product-level outward resistance terms, which depends not only by the relative concentration of a source country's output in each product but also by a function of the variables determining expenditure on the product by the destination country. So, $\tilde{\Pi}_{ni}$ is not a simple summary index of the set of Π_i^j terms, as P_n is of the set of P_n^js. Rather, it also summarizes the interaction between the distribution of output across products in i with the distribution of prices and demand conditions in n. This implies exports from i to n will be greater if country i's output is relatively concentrated in the products for which either β_n^j is higher or prices are relatively high. The intuition behind the first effect is straightforward; a source country whose output is relatively concentrated in the products that given destination country prefers to consume will export relatively more to that destination.

The intuition behind the second is a bit more nuanced. First, note that X_{ni}^j can be broken into two components: the fraction of total expenditure that is spent on product j, $\beta_n^j \left(\frac{P_n^j}{P_n}\right)^{-\sigma}$, and the fraction of expenditure on product j that is is spent on varieties originating in country i, $\left(\frac{p_i^j d_{ni}}{P_n}\right)^{-\theta}$. The first is decreasing in P_n^j with elasticity σ, while the second is increasing in P_n^j with elasticity θ. In other words, all else equal, a higher price of product j implies that consumers in n spend relatively less on that product, but, holding $p_i^j d_{ni}$ constant, consumers in n spend relatively more of that amount on country i's varieties. Since θ is greater than σ – meaning varieties within product categories are more substitutable than varieties across categories – the latter effect dominates, and sales of product j from country i are increasing in P_n^j.

2.3.2 The Trade Cost Elasticity

The fact that $\tilde{\Pi}_{ni}$ is a function of the distribution of output and expenditure across products implies that, in general, the effect of trade barriers on aggregate trade flows will also depend on these values. To see how, note that the partial elasticity of X_{ni} with respect to d_{ni} – holding constant the source country prices and aggregate expenditure – is equal to

$$\varepsilon_{ni} = \frac{d\ln(X_{ni})}{d\ln(d_{ni})} = -\theta\left(1 - \sum_j \frac{X^j_{ni}}{X^j_n}\frac{X^j_{ni}}{X_{ni}}\right) - \sigma\left(\sum_j \frac{X^j_{ni}}{X^j_n}\frac{X^j_{ni}}{X_{ni}} - \frac{X_{ni}}{X_n}\right). \tag{11}$$

This elasticity is decreasing in absolute value as the term $\sum_j \frac{X^j_{ni}}{X^j_n}\frac{X^j_{ni}}{X_{ni}}$ increases. This term is a weighted average across all products of the contribution of country i's varieties of product j to country n's expenditure on j, weighted by the fraction of total bilateral trade between i and n accounted for by product j. Thus, it is increasing in the degree to which exports from i to n are concentrated in product categories for which country i has a relatively large market share for product j in n.

The summation term lies in the range $\left[\frac{X_{ni}}{X_n}, 1\right]$, which implies that ε_{ni} lies in $\left[-\theta(1 - \frac{X_{ni}}{X_n}), -\sigma(1 - \frac{X_{ni}}{X_n})\right]$. Thus, given $\frac{X_{ni}}{X_n}$, ε_{ni} depends, at one extreme, only on θ, and at the other, only on σ. The first case occurs when X^j_{ni} is a constant fraction of X^j_n for all products. This implies that that a change in d_{ni} does not affect relative prices of different products in n, so there is no reallocation in expenditure across products, only across sources within products. As a result, the change in trade flows is governed by the elasticity of substitution across varieties within product categories, θ. The second case occurs when $X^j_{ni} = X^j_n$ for every product for which X^j_{ni} is positive. This implies that country i supplies a unique set of products to country n. In that case, a change in d_{ni} cannot affect the allocation of expenditure within products across sources because no other source is producing those products, so the change in bilateral trade flows depends only on the elasticity of substitution across products, σ.

In every other case, the trade cost elasticity is a convex combination of these two extremes, where the weight on each extreme depends non-trivially on the distribution of output and expenditure across products. One might surmise that in the two extreme cases, where the responsiveness of aggregate bilateral trade depends only on a single parameter and the exporter's aggregate market share in the importing country, that it would be possible to express aggregate trade flows as a function of aggregate data and trade barriers. That turns out to be correct, as the following proposition illustrates.

2.3.3 Some Special Cases

Proposition 1 lists four cases in which aggregate trade flows depends only on aggregate variables.

Proposition 1. *Suppose that $\beta_n^j = \beta^j$, for all j, and any of the following hold:*

1. $\frac{Y_i^j}{Y_i} = \alpha^j, \forall i, j,$
2. $\frac{Y_i^j}{Y^j} \in \{0, 1\}, \forall i, j,$
3. $\theta = \sigma,$
4. $d_{ni} = 1, \forall n, i.$

The value of aggregate trade flows from a given source, i, to a given destination, n, is given by the following system of equations:

$$X_{ni} = \frac{X_n Y_i}{Y} \left(\frac{d_{ni}}{P_n \Pi_i} \right)^{-\eta} \tag{12}$$

$$(P_n)^{-\eta} = \sum_i \left(\frac{d_{ni}}{\Pi_i} \right)^{-\eta} \frac{Y_i}{Y} \tag{13}$$

$$(\Pi_i)^{-\eta} = \sum_n \left(\frac{d_{ni}}{P_n} \right)^{-\eta} \frac{X_n}{Y}, \tag{14}$$

where the value of η in each case is

1. $\eta = \theta,$
2. $\eta = \sigma,$
3. $\eta = \theta = \sigma,$
4. $\eta = 0.$

The proof of Proposition 1 is in the appendix, but I will discuss the intuition for each in turn. The first two cases correspond directly to the extreme cases discussed above in which ε_{ni} depends only on aggregate variables. In the first case, each country's output and expenditure are distributed identically across all product categories, meaning that countries differ only in their overall level of aggregate output and expenditure as well as the bilateral trade costs that they face. As a result, relative source country prices are identical across products for each source country ($\frac{p_i^j}{p_i^{j'}} = \frac{p_{i'}^j}{p_{i'}^{j'}}$), which implies that relative values of P_n^j are also the same in each country, and each country spends the same fraction of its total expenditure on each product. Thus, trade costs affect the allocation of expenditure in each destination over each of a source country's products in the same way – governed by θ – so bilateral trade flows can be expressed as a function of aggregate output, expenditure, and multilateral resistance terms. Further, this implies that the each source country has the same market share for every product in a particular destination, so the summation term in (11) reduces to $\frac{X_{ni}}{X_n}$, meaning that the trade cost elasticity reduces to $-\theta(1 - \frac{X_{ni}}{X_n})$.

In the second case, each country produces a unique set of products. In this case, since each product is provided by only one source country, P_n^j in each country is equal to $d_{ni}p_i^j$. As a result, as in the first case, trade costs have the same proportional effect on the level of expenditure on each of a country's products, in this case governed by σ. The summation term in (11) then reduces to 1, and the trade cost elasticity reduces to $-\sigma(1 - \frac{X_{ni}}{X_n})$.

In the third case, the elasticities of substitution within and between product categories are equal, so varieties in a particular product category are indistinguishable from those in another. This is essentially a special case of case 2 in which there is a single product of which each country produces a unique set of varieties.

The final case, frictionless trade, implies that a product's point of origin is irrelevant to its price because it can reach any destination costlessly. As a result, product-level price indexes and relative expenditure on each source country's varieties of each product are identical in every country, and the revenue an exporter receives from each country for a particular product is proportional to that country's total expenditures. So, only the overall economic size of each country matters in determining aggregate bilateral trade flows.

3 Comparative Advantage in the Data

In general, if product-level trade flows are characterized by a product-level gravity relationship, it is not possible to express aggregate trade flows as a function of only aggregate variables. However, Proposition 1 shows that there are cases in which trade flows are consistent with an aggregate gravity equation, which leads to the question of whether any of the cases of Proposition 1 is a reasonable approximation to the data and, if not, then how trade cost estimates and the implications of models based on aggregate trade data are affected. This section will take a brief look at the features of the product-level trade data to gauge the reasonableness of aggregate gravity estimations and to gain some intuition into how the implications aggregate and product-level gravity models might differ.

That trade is not frictionless is taken as evident given the existence of tariffs and non-tariff barriers, an international shipping industry that makes up a significant fraction of world GDP, and the success of gravity models featuring trade barriers in rationalizing observed trade flows. Likewise, that the reduced-form elasticity of substitution within product categories is greater than that across product categories is also taken to be true in the data. This is both intuitively appealing, given the aim of product classifications to group similar products together, and consistent with the available evidence.[7] Whether either case 1 or case 2 provides a generally valid description of the data, however, is less clear. To evaluate whether that is the case, I turn to an insight from the model.

3.1 A Trade Cost Elasticity Index

The formula for ϵ_{ni} provides a useful way to summarize the degree to which the patterns of product-level output and consumption differ from the extreme

[7]See, for example, Broda and Weinstein (2006) and Eaton et al. (2011).

cases of Proposition 1. Toward this end, I define the following Elasticity Index:

$$\mathrm{EI}_{ni} = \frac{\chi_{ni}\frac{X_{nf}}{X_n} - \frac{X_{ni}}{X_n}}{1 - \frac{X_{ni}}{X_n}}$$

where

$$\chi_{ni} = \sum_j \frac{X_{ni}^j}{X_{nf}^j}\frac{X_{ni}^j}{X_{ni}}.$$

This index corresponds to the term multiplying σ in (11), which has been scaled by $1 - \frac{X_{ni}}{X_n}$ so that it is independent of an exporter's size and lies in the interval $[0,1]$. $\mathrm{EI}_{ni} = 0$ corresponds to case 1, where trade barriers affect only the allocation of expenditure within product categories, and substitution across sources is governed only by θ. Conversely, $\mathrm{EI}_{ni} = 1$ corresponds to case 2, where trade barriers affect only the allocation of expenditure across product categories, and σ governs substitution across sources. Since product-level data on domestic consumption are not available, I substitute X_{nf}^j – the value of total imports of j by n – for X_n in the computation of the χ_{ni} term. To ensure that EI_{ni} remains within $[0,1]$, the correction term $\frac{X_{nf}}{X_n}$ is added. However, omitting it does not significantly effect the results that follow.

3.2 Data

To construct the index, I use data from the United Nations Comtrade database. I focus on a cross section of product-level bilateral trade data flows from the year 2000, classified according to the 1996 revision of the Harmonize System.[8] The analysis is restricted to manufactured goods and countries for which data on manufacturing output is available, resulting in a sample of 148 countries and 4,612 products. Details are in the appendix.

Figure 1 is a histogram of the values of EI for each country pair for which aggregate trade flows are positive. Its most striking feature is that the many of values are very close to zero, meaning that for many country pairs, the world

[8] The year and classification system are chosen to maximize the number and countries and products for which data were reported in a common classification system. The results that follow are not qualitatively affected by the choice of year or HS revision.

Figure 1: Histogram of Elasticity Index Values

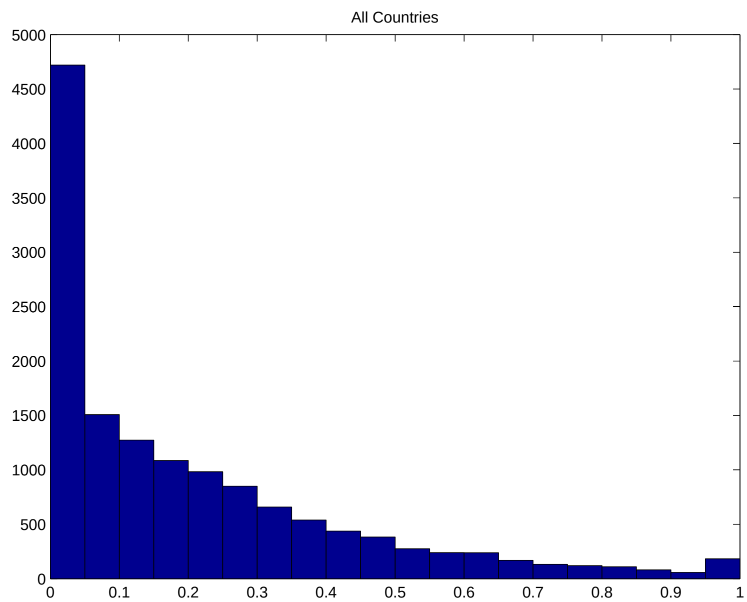
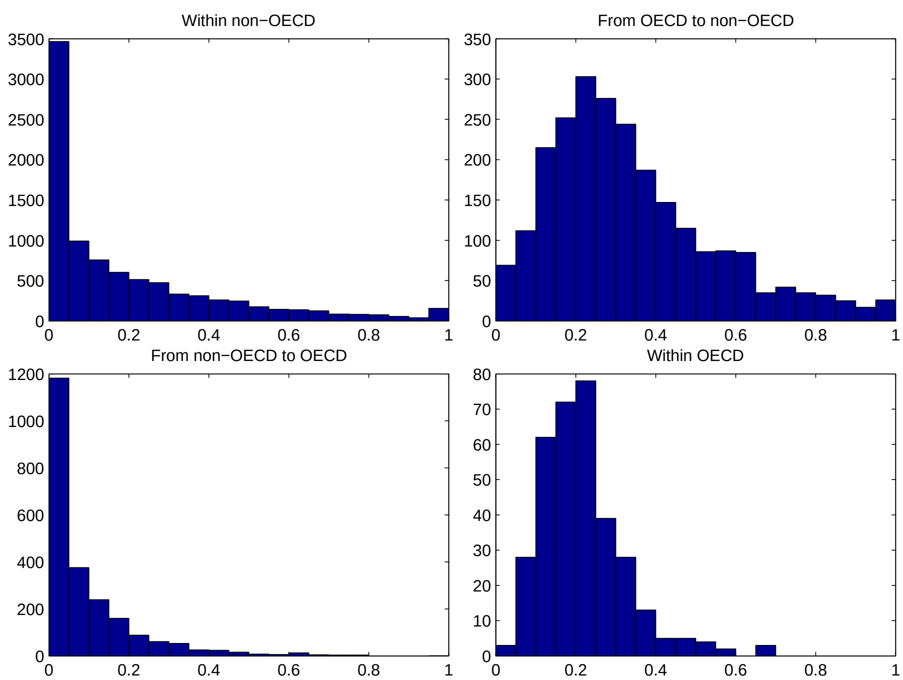


Figure 2: Histograms of Elasticity Index Value by Group



looks very similar to case 1. However, there is some heterogeneity, with EI being small but significantly positive for many countries and EI being very close to 1 for a non-negligible number of country pairs.

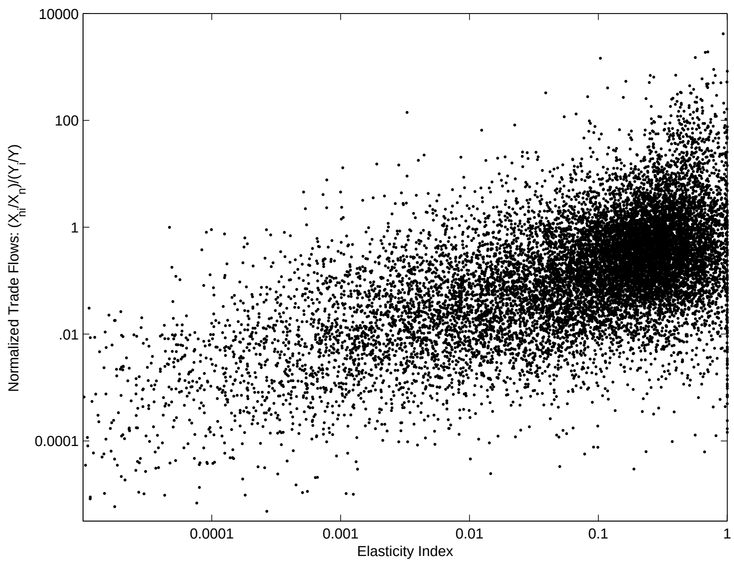
Figure 2 explores this heterogeneity further, sorting values of EI by whether they correspond to trade flows originating or arriving in OECD or non-OECD countries. It is evident that OECD exports tend to have a higher EI, while OECD imports tend to have a lower EI, which implies that there are systematic differences in the set of products produced by developed and developing countries which affects the degree to which their imports and exports respond to trade barriers. Notably, this implies that trade barriers can have an asymmetric effect on trade flows between a given pair countries depending on the direction of trade, with exports from OECD to non-OECD countries facing a lower trade cost elasticity on average than exports from non-OECD to OECD countries.

It is clear that the elasticity index varies substantially across country pairs and even by direction of trade. Figure 3 takes the analysis a step further in evaluating whether the composition of countries' output and expenditure has a substantial effect on the relationship between trade barriers and aggregate bilateral trade flows. The figure plots bilateral trade flows, normalized by importer and exporter size – $\frac{X_{ni}}{X_n}/\frac{Y_i}{Y}$ – against the associated value of EI. In an aggregate gravity model consistent with (12), these values would be unrelated. However, there is a clear positive relationship between the size of a flow and the corresponding value of EI, indicating that trade flows for which the model predict trade barriers to have a smaller effect, trade flows are larger.

4 Estimation

The evidence indicates that trade barriers affect trade flows between countries to different degrees depending on the composition, and not simply the level, of countries' output and expenditure. However, to effectively quantify the importance of accounting for the composition of trade flows, it is necessary to formally estimate a trade cost function in a way that is consistent with a

Figure 3: Elasticity Index and Trade Flows



generic set of product-level output and expenditure values and to parameterize a gravity model that is consistent such patterns. This section develops and implements such an estimation strategy.

4.1 Empirical Framework

I assume that trade costs, d_{ni}, are a semi-parametric function of a set of bilateral relationship variables commonly used in the gravity literature, taking the following functional form:

$$\log\left(d(\boldsymbol{I}_{ni};\boldsymbol{\alpha})\right) = \begin{cases} \alpha_i + \sum_k(\alpha^k I^k_{ni}) + \alpha^b I^b_{ni} + \alpha^l I^l_{ni} + \alpha^c I^c_{ni} & \text{if } n \neq i \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

where I^k_{ni} is an indicator that the distance between n and i lies in the interval k, I^b_{ni} that n and i share a border, I^l_{ni} that they share a common language, and I^c_{ni} that they share a colonial relationship.[9] The parameter α_i the cost

[9]The six distance intervals are (in kilometers) [0, 625); [625, 1,250); [1,250, 2500); [2,500, 5,000); [5,000, 10,000); and [10,000, max]. In the estimations that follow, the dummy variable associated with the interval [0, 625) is the one omitted to avoid multicollinearity with the exporter-specific effect, meaning the total cost associated with shipping a good within distance category k is $e^{(\alpha_i+\alpha_d^k)}$.

associated with crossing a national border. That it varies by country implies that trade costs can be asymmetric. The effect of this component of the trade cost function on trade flows is identical regardless of whether it is varies by importer (specified as α_n) or by exporter, as it is here. I follow Waugh (2010), which finds that relationship between income per worker and relative prices or tradable goods in the data is more consistent with trade costs that vary by exporter, in choosing the latter specification. Data on bilateral relationships are from CEPII. Details are in the appendix.

With this specification of trade costs, the stochastic form of (6) is

$$X_{ni}^j = \frac{X_n^j Y_i^j}{Y^j} \left(\frac{d(\boldsymbol{I}_{ni}; \boldsymbol{\alpha})}{P_n^j \Pi_i^j} \right)^{-\theta} + \epsilon_{ni}^j, \tag{16}$$

where

$$E(\epsilon_{ni}^j | \boldsymbol{X^j}, \boldsymbol{Y^j}, \boldsymbol{I}) = 0,$$

and P_n^j and Π_i^j are given by (7) and (8), respectively. The error term, which can be thought of as measurement error, is simply appended to the product-level gravity equation in keeping with the typical practice of aggregate gravity specifications. Of course, there are likely many other sources of variation in trade flows, such as unobservable components of the trade cost function. Treatment of such sources of variation – which would imply that P_n^j and Π_i^j are functions of the errors – is beyond the scope of this paper. Eaton et al. (2012) is a recent attempt to deal with such issues, and Anderson and van Wincoop (2003) discuss why biases from some such variation are likely to be small.

Equation (16) expresses the expected value of product-level bilateral trade flows as a function of data on product-level output and expenditure and the set of trade cost parameters, $\boldsymbol{\alpha}$. Given such data, it is straightforward to estimate the value of the trade cost parameters from product-level trade data using standard techniques. However, data on output or expenditure are not available at a level of disaggregation comparable to the product-level trade data, making estimation in such a way impossible. As a result, another method of estimating $\boldsymbol{\alpha}$ must be employed.

4.2 Aggregate Estimation

The strategy typically employed in the literature is to estimate the trade costs parameters using data on aggregate output and expenditure (or GDP) along with aggregate bilateral trade flows. As has been discussed above, this makes the implicit assumption that the volume of trade flows is independent of the distribution of output and expenditure across products. For the sake of completeness, suppose that one of cases 1 - 3 of Proposition 1 describes the world economy. Equation (16) then reduces to

$$X_{ni} = \frac{X_n Y_i}{Y} \left(\frac{d(\boldsymbol{I}_{ni}; \boldsymbol{\alpha})}{P_n \Pi_i} \right)^{-\eta} + \epsilon_{ni}, \tag{17}$$

where

$$E(\epsilon_{ni} | \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{I}) = \sum_j E(\epsilon^j_{ni} | \boldsymbol{X^j}, \boldsymbol{Y^j}, \boldsymbol{I}) = 0$$

and P_n and Π_i are given by (13) and (14), respectively.

Equation (17) is typically estimated in one of two ways: 1) by using (13) and (14) to solve for the multilateral resistance terms, making (16) a nonlinear function of data and parameters and estimating via nonlinear least squares, as in Anderson and van Wincoop (2003); and 2) by using importer and exporter fixed effects to control for the multilateral resistance terms, making a (17) a log-linear function of data and parameters, which can be estimated via OLS. The second technique has the advantage of simplicity and robustness to country-specific unobserved heterogeneity, while the first is more efficient, as it imposes more of the model's structure. The first technique also has the advantage that the multilateral resistance terms estimated are consistent with the underlying trade model of interest, and the predicted trade flows are consistent with observed output and expenditure.

Both methods, when estimated in logs, present two major drawbacks, which are discussed by Santos Silva and Tenreyro (2006): 1) all observations for which bilateral trade is equal to zero are dropped from the estimation, and 2) the estimates are potentially biased in the presence of heteroskedasticity. Santos Silva and Tenreyro (2006) propose using pseudo-maximum-likelihood

(PML) estimation – especially Poisson PML – to estimate (17) in its multiplicative form using the fixed effects approach. However, the first technique can also be employed in a PML framework using the multiplicative form of (17) and is only marginally more difficult to implement, so I employ both techniques with variety of PML estimators below.

4.3 Product-Level Estimation

However, estimation based on aggregate data is not valid if the composition of output and expenditure does not satisfy the cases of Proposition 1, so the lack of product-level output data must be overcome in another way. One solution is to reformulate the model so that the expected value of product-level bilateral trade flows are expressed as a function total product-level *exports* and *imports* – that is, output and expenditure net of the value of domestic trade – rather than total output and expenditure. It turns out that the model readily admits such a formulation, which is given by Proposition 2.

Proposition 2. *Given a set of bilateral trade flows that satisfy (6), (7), and (8), the same set also satisfy the following*

$$X_{ni}^j = \frac{X_{nf}^j Y_{if}^j}{Y_f^j} \left(\frac{d_{ni}}{P_{nf}^j \Pi_{if}^j} \right)^{-\theta} + \tilde{\epsilon}_{ni}^j \tag{18}$$

$$(P_{nf}^j)^{-\theta} = \sum_{i \neq n} \left(\frac{d(\boldsymbol{I}_{ni}; \boldsymbol{\alpha})}{\Pi_{if}^j} \right)^{-\theta} \frac{Y_{if}^j}{Y_f^j} \tag{19}$$

$$(\Pi_{if}^j)^{-\theta} = \sum_{n \neq i} \left(\frac{d(\boldsymbol{I}_{ni}; \boldsymbol{\alpha})}{P_{nf}^j} \right)^{-\theta} \frac{X_{nf}^j}{Y_f^j}, \tag{20}$$

where $X_{nf}^j = \sum_{i \neq n} X_{ni}^j$, $Y_{if}^j \sum_{n \neq i} X_{ni}^j$, *and* $Y_f^j = \sum_i Y_{if}^j$.

The stochastic form of (18) is

$$X_{ni}^j = \frac{X_{nf}^j Y_{if}^j}{Y_f^j} \left(\frac{d(\boldsymbol{I}_{ni}; \boldsymbol{\alpha})}{P_{nf}^j \Pi_{if}^j} \right)^{-\theta} + \tilde{\epsilon}_{ni}^j, \tag{21}$$

where

$$E(\tilde{\epsilon}_{ni}^j | \boldsymbol{X_f^j}, \boldsymbol{Y_f^j}, \boldsymbol{I}) = 0,$$

and the value of the error term differs from that in (16) because in this specification, total product-level imports and exports rather than output and expenditure are taken as given.

Equation (18) can be estimated in its nonlinear form in exactly the same way as discussed above except that no data on product-level domestic trade is necessary.[10] I refer to this as the "conditional" estimation strategy, as the procedure computes expected bilateral trade flows conditional on total imports and exports.

Employing this strategy is not entirely costless, however. Since it does not make use of data on the value of domestic trade, it is not possible to identify the exporter-specific border costs. This is due to the property of gravity models made clear by Anderson and van Wincoop (2003) that bilateral trade flows depend only on relative trade costs. Since border costs only vary between domestic and foreign sales and not across foreign destinations, they have no effect on the bilateral trade flows, given total imports and exports. More formally, in (18) - (20) it is straightforward to show that a change in α_i simply causes a change in Π_{if}^j for all j which is proportional to the change in d_{ni} for all n, such that there no change in X_{ni}^j.

It is still possible, though, to obtain a value for α_i that is consistent with data on the value of aggregate domestic trade, which is available. Given a set of parameter estimates for bilateral component of the trade cost function, the model's predicted value of aggregate domestic trade for a given country is

$$\hat{X}_{nn} = \sum_j \hat{X}_{nn}^j = \sum_j \frac{X_{nf}^j Y_{if}^j}{Y_f^j} \left(\frac{1}{\hat{P}_{nf}^j \hat{\Pi}_{nf}^j} \right)^{-\theta}, \tag{22}$$

where $\hat{P}_{nf}^j$ and $\hat{\Pi}_{nf}^j$ are the respective values of P_{nf}^j and Π_{nf}^j as functions of the estimated trade cost parameters as well as the exporter-specific trade cost

[10] Estimation of (18) using fixed effects is also theoretically possible. However, as it would involve the estimation of 2*(N-1)*(J-1) = 1,355,634 coefficients on the set of country-product dummy variables, it is practically infeasible given current computational power.

parameter. Since domestic trade is assumed to be costless, and the value of α_i affects only the value of $\hat{\Pi}_i^j$, the value of α_i can be chosen to equate the value of $\hat{X}_{ii}$ with its counterpart in the data for each source country.

4.4 Results

The coefficient estimates from the four sets of estimation strategies discussed above are presented in Table 1. These include the reduced form estimation with fixed effects and the nonlinear estimation using aggregate data as well as the estimation conditional on total imports and exports using both aggregate and product-level data. The first column reports the estimates from the logged gravity equation obtained via least squares.[11] Columns 2 - 4 present the results of PML estimations based on three different probability distributions: gamma, Poisson, and Gaussian (which implies non-linear least squares in levels). As is shown in Gourieroux et al. (1984), all three produce coefficient estimates which are asymptotically consistent but make different assumptions about the form of heteroskedasticity, meaning observations are weighted differently by each objective function.

While the Poisson specification has become the most widely used in recent years, it is informative to include the estimates from the gamma specification, as it assumes the same form of heteroskedasticity as log least squares but does not omit the zero-valued observation or suffer from bias under other forms of heteroskedasticity. Least squares in levels is included for completeness despite criticism by Santos Silva and Tenreyro (2006) and others that it is generally unreliable due to its placing excessive weight on large, noisy observations. In fact, the nonlinear least squares estimation in levels was numerically unstable, so the parameter estimates are omitted.

The results from the aggregate estimations are roughly in line with the literature. Bilateral trade is generally decreasing in distance and higher if countries share a border, language, or colonial ties. The average ad valorem tariff equivalent of the trade costs implied by these estimates is higher than

[11] OLS in reduced form estimation with fixed effects and nonlinear least squares otherwise.

Table 1: Trade Cost Estimates

Variable	Log LS		Gamma PML		Poisson PML		Least Squares	
	Aggregate Log-Linear Estimation							
<625 km	−5.31	(0.24)	−3.39	(0.37)	−5.92	(0.19)	−6.89	(0.36)
625 – 1,250 km	−6.42	(0.12)	−4.93	(0.22)	−6.29	(0.16)	−7.31	(0.32)
1,250 – 2,500 km	−7.55	(0.09)	−6.44	(0.17)	−6.61	(0.12)	−7.57	(0.21)
2,500 – 5,000 km	−8.86	(0.06)	−7.75	(0.10)	−7.21	(0.11)	−8.08	(0.22)
5,000 – 10,000 km	−9.94	(0.04)	−8.88	(0.06)	−8.10	(0.12)	−9.04	(0.28)
>10,000 km	−10.60	(0.07)	−9.68	(0.08)	−8.13	(0.09)	−8.74	(0.20)
Shared Border	0.94	(0.15)	0.99	(0.26)	0.55	(0.10)	0.37	(0.15)
Common Language	1.00	(0.10)	1.02	(0.12)	0.18	(0.09)	0.18	(0.17)
Colonial Ties	0.88	(0.14)	1.33	(0.24)	0.01	(0.12)	−0.16	(0.11)
	Aggregate Non-Linear Estimation							
<625 km	−6.40	(0.33)	−5.02	(0.21)	−5.92	(0.19)	–	–
625 – 1,250 km	−6.69	(0.13)	−5.34	(0.12)	−6.29	(0.16)	–	–
1,250 – 2,500 km	−7.64	(0.10)	−6.31	(0.11)	−6.61	(0.12)	–	–
2,500 – 5,000 km	−8.94	(0.09)	−7.53	(0.09)	−7.21	(0.11)	–	–
5,000 – 10,000 km	−9.96	(0.05)	−8.31	(0.07)	−8.10	(0.12)	–	–
>10,000 km	−10.59	(0.09)	−9.11	(0.08)	−8.13	(0.09)	–	–
Shared Border	0.32	(0.15)	0.48	(0.18)	0.55	(0.09)	–	–
Common Language	1.05	(0.12)	0.86	(0.09)	0.18	(0.09)	–	–
Colonial Ties	0.98	(0.20)	0.53	(0.14)	0.01	(0.12)	–	–
	Aggregate Conditional Estimation							
<625 km	−5.08	–	−4.84	–	−5.92	–	−6.30	–
625 – 1,250 km	−6.05	(0.59)	−5.39	(0.30)	−6.29	(0.18)	−6.60	(1.28)
1,250 – 2,500 km	−6.25	(0.45)	−6.21	(0.35)	−6.61	(0.24)	−6.68	(1.56)
2,500 – 5,000 km	−8.21	(0.48)	−7.55	(0.40)	−7.21	(0.27)	−7.15	(1.23)
5,000 – 10,000 km	−9.62	(0.50)	−8.23	(0.43)	−8.10	(0.28)	−8.28	(2.43)
>10,000 km	−10.10	(0.59)	−8.87	(0.46)	−8.13	(0.36)	−7.93	(2.22)
Shared Border	−0.44	(0.43)	0.48	(0.22)	0.55	(0.14)	0.55	(0.67)
Common Language	1.83	(0.26)	0.60	(0.14)	0.18	(0.08)	−0.01	(0.34)
Colonial Ties	1.64	(0.33)	0.39	(0.18)	0.01	(0.13)	−0.08	(0.38)
	Product-Level Conditional Estimation							
<625 km	−4.91	–	−4.97	–	−5.55	–	−5.91	–
625 – 1,250 km	−5.79	(0.11)	−5.92	(0.40)	−5.95	(0.24)	−6.21	(2.78)
1,250 – 2,500 km	−6.34	(0.13)	−6.47	(0.40)	−6.32	(0.33)	−6.34	(2.81)
2,500 – 5,000 km	−7.36	(0.13)	−7.46	(0.38)	−6.97	(0.38)	−6.83	(3.34)
5,000 – 10,000 km	−8.18	(0.13)	−8.26	(0.43)	−8.07	(0.38)	−8.17	(5.22)
>10,000 km	−8.80	(0.13)	−8.81	(0.48)	−8.29	(0.46)	−7.83	(5.01)
Shared Border	0.57	(0.08)	0.37	(0.24)	0.58	(0.17)	0.55	(1.71)
Common Language	0.54	(0.05)	0.96	(0.17)	0.17	(0.08)	−0.11	(0.28)
Colonial Ties	0.77	(0.07)	0.42	(0.34)	0.07	(0.11)	−0.36	(0.99)

Notes: Parameters reported, $\hat{\beta}$, represent $-\theta\hat{\alpha}$. The implied percentage effect of each coefficient on ad valorem tariff equivalent trade cost is $100 \times (e^{\hat{\beta}/\theta} - 1)$. Distance coefficients are normalized so that $\sum_i \theta\hat{\alpha}_i = 0$.

that from many previous estimations.[12,13] However, this is largely due to the use of a large sample of countries, including many small and less developed countries, which are estimated to have higher border costs, whereas most previous studies have focused on trade among smaller sets of mostly developed countries.

There are two features of the estimates based on product-level data that stand out when compared with those based on aggregate data. The first is their robustness to distributional assumptions. The second is that the implied trade costs are generally lower than those based on estimates from aggregate data. These features can more easily be seen in Figure 4, which plots the estimated percentage effect on trade costs of being in each distance category for each of the estimation strategies. While both the overall slopes and the intercepts of the functions vary significantly in the three panes plotting the estimates based on aggregate data, the functions based on estimates from product-level data are nearly indistinguishable. Particularly interesting is that the estimated effects of distance on trade flows from the log least squares and gamma PML estimations are nearly identical, implying that biases of log least squares resulting from heteroskedasticity and the omission of zero-valued observations are negligible once the composition of output and expenditure has been accounted for.

Further, it is apparent that overall trade costs are estimated to be higher on average in estimations based on aggregate data, as the functions in the lower-right pane of Figure 4 generally lie below those in the other panes. For example, in the aggregate reduced form estimation, which is the standard practice in the literature, the trade costs associated with a pair of countries between 1,250 and 2,500 km apart ranges from a tariff equivalent of 195% to 278%, depending on the specification, while in the estimation based on

[12]The ad valorem tariff equivalent trade costs is given by $100 \times (d_{ni} - 1) = 100 \times (e^{\hat{\boldsymbol{\alpha}} \boldsymbol{I}_{ni}} - 1)$. Since the estimated coefficients in Table 1 represent $\theta\hat{\boldsymbol{\alpha}}$, and the value of θ is not separately identifiable using trade data from the trade cost parameters, I use the value of θ estimated from price data by Waugh (2010) of 5.5 calculate trade costs.

[13]See Anderson and van Wincoop (2004) for a list of tariff equivalent trade costs estimated in a gravity framework in other studies.

Figure 4: Estimated Increase in Cost Due to Distance

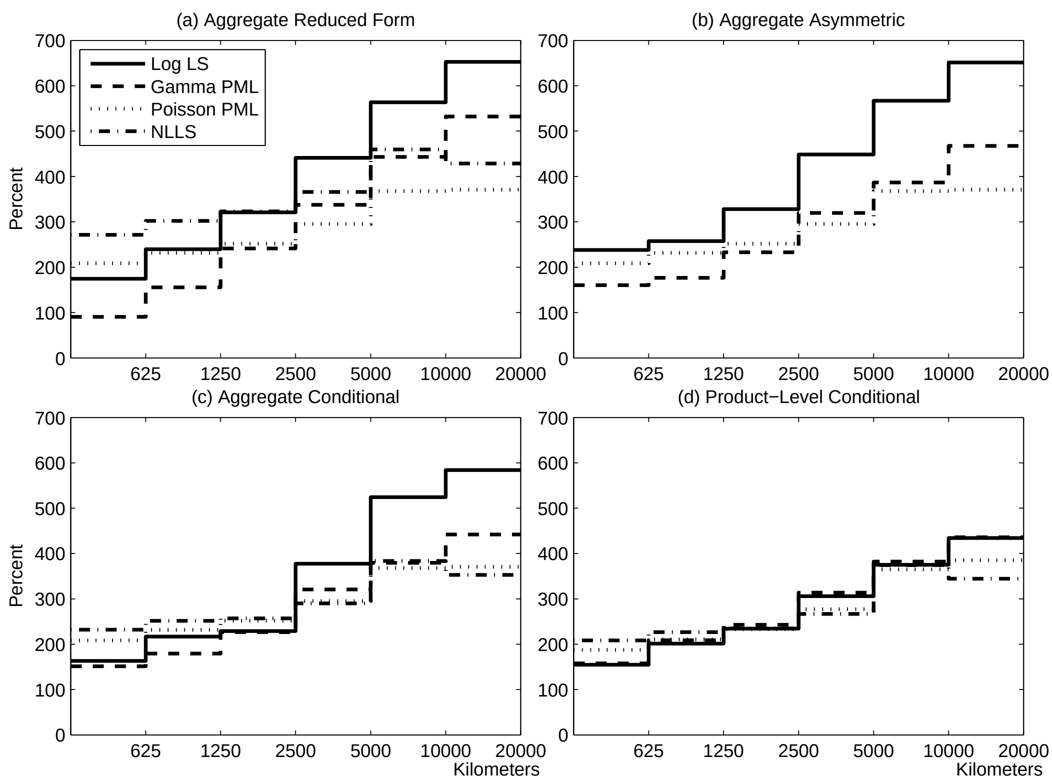
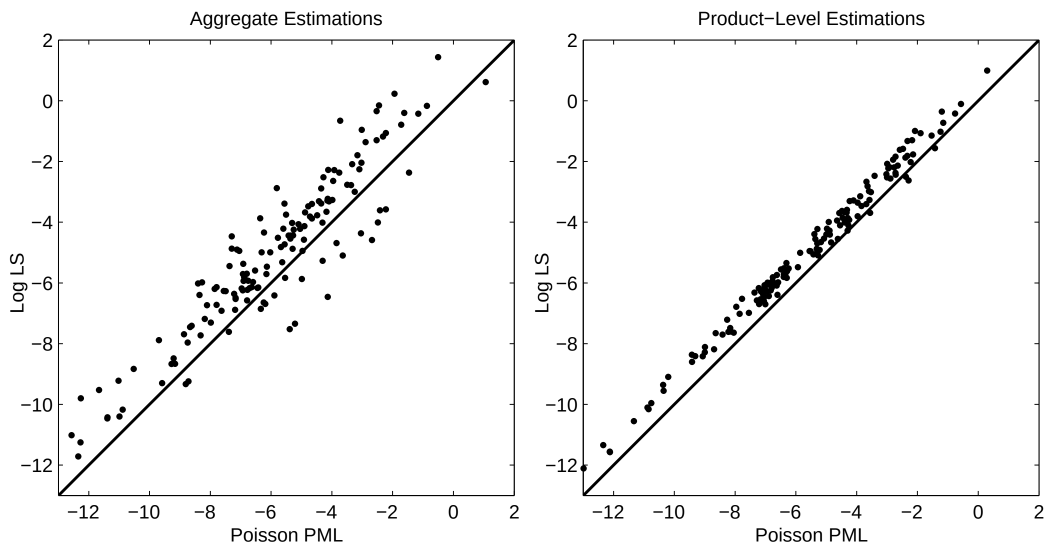


Figure 5: Fixed Effect Coefficients



product-level data, it ranges from 178% to 206%, with a difference between the two of between 11 and 76 percentage points.

Figure 5 demonstrates that the same pattern is also present in the estimates of the exporter-specific border costs. The left pane of the figure plots the estimates $-\theta\hat{\alpha}_i$ from the log least squares specification of the conditional estimation using aggregate data against those from Poisson PML, while the right set of axes plots the corresponding coefficients from the estimations using product-level data. It is clear the coefficients of the product-level estimations fall much closer to the 45 degree line, indicating that the differences in coefficient estimates across specifications are much smaller when estimated from product-level data. The pattern is similar for all other pairs of specifications and for the other estimation techniques using aggregate data.

Looking deeper at the differences in coefficient estimates obtained under different distributional assumptions reveals a clear pattern. As we move from gamma PML to Poisson PML to least squares, the estimated average border effect (the coefficient on the first distance category) gets larger, while the additional effect of distance shrinks. Santos Silva and Tenreyro (2006) attribute such differences to low levels of efficiency in finite samples of PML estimators that give excessive weight to particularly noisy observations. However, it is still useful to consider why such a patter might emerge. To this end, note that gamma PML (as well as log least squares) assumes that the variance of the error term is proportional to the expected value of the trade flow, meaning that the weights implied by the likelihood function are inversely proportional to the volume of trade. Poisson PML, assumes that the variance of the error is independent of the volume of trade and weights all observations equally, while least squares in levels assumes that variance of the error is inversely proportional to the volume of trade and weights observation proportionally to their size.

In a world in which trade flows follow a product-level gravity equation, the effect of trade costs on trade flows depends on the composition of output and expenditure across products, and as Figure 3 indicates, trade flows between pairs of countries for which the model would predict a smaller effect of trade costs on trade flows are larger. Therefore, in an estimation which imposes a constant elasticity of trade flows with respect to trade costs, an estimator that places more weight on larger observations will estimate a smaller effect of distance on trade. This is exactly what we find in all the estimations based on aggregate data. Why then do we see the opposite pattern in the estimates of the effect of borders? This is because α_i is identified by domestic trade flows. So, when the estimator estimates a large trade cost due to distance, it must also estimate a low border cost in order to keep domestic trade from being too large, and vice versa for small distance related trade costs.

In contrast to the estimates based on aggregate data, the effects of borders and distance are estimated to be much more similar based on product-level data. In fact, returning to Figure 4, it is evident that the trade costs implied by borders and distance for the four mid-range distance categories – in which 95% of country pair lie – are virtually identical.

Taken together, the evidence is strongly suggestive that the bias introduced by ignoring the effect of the composition of output and expenditure on aggregate bilateral trade flows is significant and that it is likely more severe than the bias due to omitting zero-valued observations and using estimation techniques that are not robust to different forms of heteroskedasticity. In fact, much of differences in estimates of the distance elasticity base on different distributional assumptions appears to be the result of the inability of estimations based on aggregate data to simultaneously fit the trade flows of a large set of countries with differing patterns of output and expenditure.

4.5 Implications

In addition to examining the parameter estimates based on aggregate and product-level data, there are several other implications of the estimations that shed light on the effect of the composition of output and expenditure on trade flows. I focus on two: the estimated values of the outward multilateral resistance terms and the ability of the aggregate and product-level models to predict domestic trade shares when export costs are not chosen specifically to match them.

4.5.1 Outward Multilateral Resistance

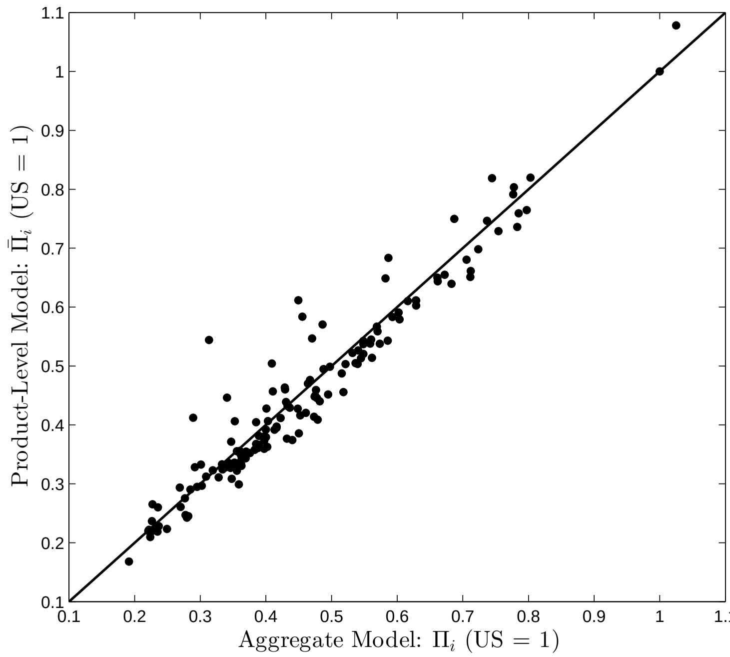
The key difference between the aggregate and product-level gravity models is that the latter allows the outward multilateral resistance term to vary by importer. To evaluate the degree to which this variance is important in accounting for aggregate bilateral trade flows, I first define a measure of average multilateral resistance for each exporter in order to separate variance across exporters from variance within each exporter across importers. Let

$$\bar{\Pi}_i = \sum_n \left(\frac{d_{ni}}{P_n}\right)^{-\theta} \frac{X_n}{Y},$$

which corresponds to the definition of Π_i in the aggregate model. Figure 6 shows that the values of $\bar{\Pi}_i$ estimated using product-level data are quite similar to the value of Π_i estimated using aggregate data, indicating that it is a sensible measure of the average value of Π_{ni} across all destinations.

To explore the degree to which Π_{ni} typically differs from $\bar{\Pi}_i$ for a given pair of countries, Figure 7 plots the histogram of percent deviations of Π_{ni} from $\bar{\Pi}_i$ for all country pairs. Note that, all else equal, a value of Π_{ni} that

Figure 6: Outward Multilateral Resistance Terms



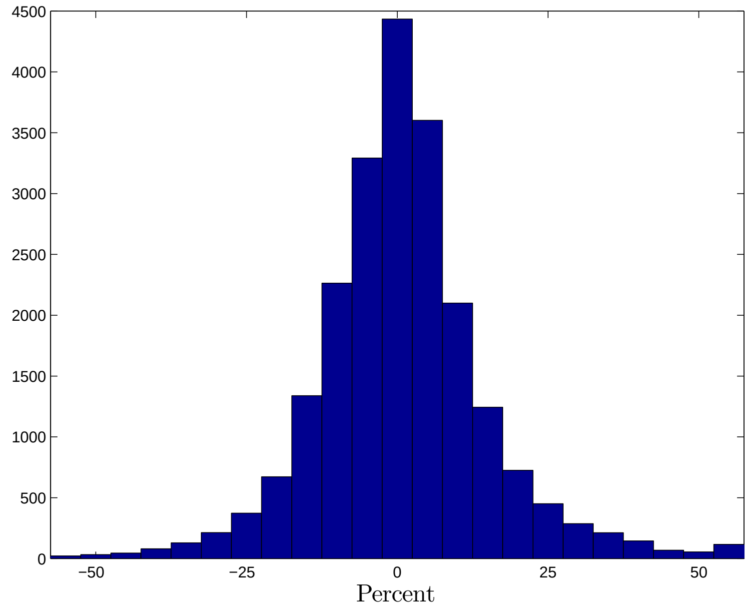
is 25% below $\bar{\Pi}_i$ has the same effect on bilateral trade as a 25% increase in d_{ni}. This indicates that the variation in Π_{ni} across country pairs has a significant effect on bilateral trade flows. To further quantify the degree to which differences in Π_{ni} across destinations matter for aggregate trade flows, consider the following identity relating predicted trade flows to trade costs and the multilateral resistance terms.

$$\log\left(\frac{\hat{X}_{ni}}{X_n}\Bigg/\frac{Y_i}{Y}\right) = \log(d_{ni}^{-\theta}) + \log(P_n^{\theta}) + \log(\bar{\Pi}_i^{\theta}) + \log\left(\left(\frac{\Pi_{ni}}{\bar{\Pi}_i}\right)^{\theta}\right) \tag{23}$$

A regression the left-hand side of (23) on all the right-hand side variables except the last has an R^2 of .63, indicating that 37% of the variation in predicted aggregate bilateral trade flows is due to the variation in Π_{ni} relative to $\bar{\Pi}_{ni}$, which is to say that it is due to the interaction between countries' distributions of output and expenditure across products.

The degree to which Π_{ni} differs from $\bar{\Pi}_i$ also varies across source countries. For instance, Figure 8 plots separately the histograms for the percentage de-

Figure 7: Histogram of $\frac{\tilde{\Pi}_{ni}}{\bar{\Pi}_i}$



viation of Π_{ni} from $\bar{\Pi}_i$ for OECD and non-OECD countries. The figure shows that Π_{ni} varies much more relative $\bar{\Pi}_i$ for non-OECD countries, indicating that the composition trade is much more important for explaining the the volume of exports from non-OECD countries.

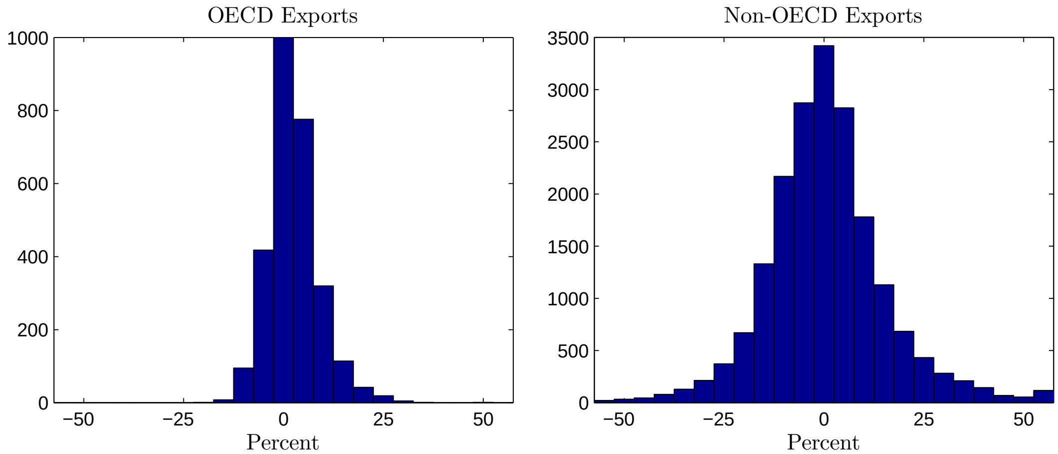
4.5.2 Predicted Home Trade Share

One well-documented area in which the predictions of gravity models are at odds with the data is the relationship between the fraction of countries' output that is traded and its level of per-capita income. The relationship implied by gravity models can be seen quite clearly by examining the expression for the fraction of country's expenditure that is produced domestically. Dividing both sides of equation (12) in the case where $n = i$ gives

$$\frac{X_{ii}}{X_i} = \frac{Y_i}{Y}\left(\frac{1}{P_i \Pi_i}\right)^{-\eta}.$$

The domestic share of expenditure is strongly related to a country's level of output. However, it is only related to per-capita income inasmuch as income

Figure 8: Histogram of $\frac{\tilde{\Pi}_{ni}}{\bar{\Pi}_i}$ by Group



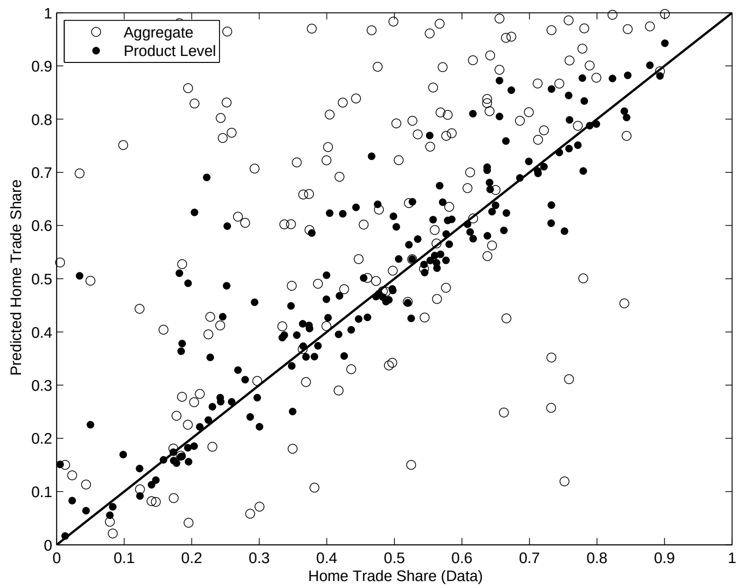
is correlated with total output or trade costs. However, in the data, the home share of expenditure is essentially unrelated to per-capita income and has a weaker relationship with total output than the model predicts.[14]

Of course, if trade costs are specified as having a country-specific component, as in many theoretically-founded gravity estimations, the trade costs that are systematically higher for low-income countries would reconcile the model with the data, which Waugh (2010) documents is precisely the case for trade costs estimated from trade data in such a framework. However, since such a specification can mechanically replicate such a feature of the data, examining its ability to reproduce it is not a particularly useful exercise. So, instead, I evaluate the predictions of the model under the restriction that α_i is constant across exporters.

I use the following procedure to obtain the predicted value of expenditure on domestic output given a common border cost. First, I set $\alpha_i = \bar{\alpha}$, for all i, where $\bar{\alpha} = \frac{1}{N}\sum_i \hat{\alpha}_i$. In the aggregate case, it is straightforward to solve for domestic trade using (17), (13), and (14) along with the remaining trade cost parameters and data on X_n and Y_i. In the product-level case, however, in order to perform the same exercise using equation (16), information on X_n^j and

[14] Fieler (2010) is a recent example which documents the relationship in the data and attempts to reconcile it with a gravity-type model.

Figure 9: Predicted versus Actual Home Trade Share



Y_n^j is required. For this, I employ the values, $\hat{X}_{ii}^j$ from equation (22) evaluated at $\alpha_i = \hat{\alpha}_i$ along with the values of X_{nf}^j and Y_{if}^j from the data. Then, given $\hat{X}_i^j = \hat{X}_{ii}^j + X_{if}^j$ and $\hat{Y}_i^j = \hat{X}_{ii}^j + Y_{if}^j$, it is straightforward to predict domestic trade, given $\alpha_i = \bar{\alpha}$, using (16), (7), and (8).

Figure (9) illustrates the predictions of the aggregate and product-level models, based on the conditional Poisson PML specification.[15] Both are plotted against the actual home shares from the data. It is clear that the predictions from the product-level model lie much closer to the 45 degree line than those of the aggregate model, indicating that taking into account the composition of output and expenditure goes quite far in explaining how much of a country's output is exported or consumed domestically even without the inclusion of country-specific trade costs.

Table (2) shows that the product-level model also is much better at predicting the relationship between the domestic expenditure share and the country characteristics discussed above. The first column presents the coefficients from

[15]The predictions based on the other aggregate estimation techniques and specifications are qualitatively similar.

Table 2: Correlation of Home Trade Share and Income Components

		Log LS		Gamma PML		Poisson PML		Least Squares	
Variable	Data	Agg	PL	Agg	PL	Agg	PL	Agg	PL
Mfg. Output	0.18	0.26	0.20	0.26	0.20	0.25	0.20	0.25	0.20
	(0.03)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
GDP/Worker.	−0.22	−0.13	−0.13	−0.10	−0.13	−0.08	−0.12	−0.07	−0.12
	(0.05)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)

a regression of the log of the domestic expenditure share in the data on a constant, manufacturing output, and GDP per worker. The remaining columns indicate that all four estimations of the aggregate conditional specification over-predict the magnitude of the elasticity of the domestic expenditure share with respect to output and under-predict the magnitude of the elasticity with respect the GDP per worker. The product-level model, on the other hand, significantly reduces the distance between the predicted and actual elasticities.

5 Counterfactual Predictions

Using product-level data to take into account the effect of the composition of output and expenditure on bilateral trade flows leads not only to different estimates of trade costs but also to differing predictions about the effects of changes in trade barriers or the state of technology. To explore how the predictions based on estimates using product-level data differ from those using aggregate data, I perform several counterfactual experiments.

The data on trade flows and aggregate output are all the data that are necessary to perform counterfactual experiments, but it is necessary to set values of the parameters σ and θ and to close the factor markets. To do the latter, I employ the Ricardian model of Eaton and Kortum (2002), generalized to allow for deterministic differences in productivity across product categories, which is described in Appendix C.1. From (31) and (5), we have that

$$T_i^j = (p_i^j)^{-\theta} w_i = \frac{Y_i^j}{Y^j} \left(\frac{1}{\Pi_i^j} \right)^{-\theta},$$

where T_i^j is the parameter governing the average productivity of i in producing j, and w_i is the price of a bundle of inputs in i. It is useful to work with the following transformed value of T_i^j,

$$\tilde{T}_i^j = T_i^j L_i^\theta = \frac{(Y_i^j)^{1+\theta}}{Y^j} \left(\frac{1}{\Pi_i^j} \right)^{-\theta},$$

where L_i is the effective supply of inputs devoted to the manufacturing sector of i, and the last equality follows from the identity $Y_i = w_i L_i$. To close the factor markets, I assume that labor is the only factor of production and that it is perfectly mobile within countries across product categories but perfectly immobile across countries and between the manufacturing and non-manufacturing sectors.[16] Finally, as in Dekle et al. (2008), I take each country's nominal trade deficit, D_n as given, which implies that $X_n = Y_n + D_n$. Given a set of values, $\{\tilde{T}_i^j\}$, $\{\beta_n^j\}$, σ, and θ, the equilibrium of this model is the set of values of nominal output, $\{Y_i\}$, which solves the following system of equations:

$$Y_i = \sum_n \pi_{ni} \left(Y_n + D_n \right), \quad \forall i,$$

where

$$\pi_{ni} = \sum_j \frac{\tilde{T}_i^j (Y_i d_{ni})^{-\theta}}{(P_n^j)^{-\theta}} \beta_n^j \left(\frac{P_n^j}{P_n} \right)^{-\sigma},$$

and P_n^j and P_n are as given by (31). For counterfactual predictions based on aggregate data, I impose that $T_i^j = T_i$ and $\beta_n^j = 1$, for all j, so that the model is consistent with an aggregate gravity equation.

For the following counterfactual experiments, I use a value of θ of 5.5, which is value estimated by Waugh (2010) based on price data and lies in the range of estimates from Eaton and Kortum (2002) and other papers, and a value of σ of 2.25, which is taken from the value $\frac{\theta}{\sigma} = 2.46$, estimated by Eaton et

[16] That labor is the only factor of production is maintained for convenience of interpretation. L_i can be also be interpreted as a composite of multiple inputs without changing any of the results that follow. That labor is immobile between manufacturing and non-manufacturing sectors is very useful in that it makes it possible to explore the effect of changes in trade barriers and technology on the manufacturing sector – which makes up the majority of international trade – without requiring data on other sectors, which is less readily available. However, it is straightforward to extend the analysis to multi-sector economies by following the methodology of Dekle et al. (2008) and Waugh (2010).

al. (2011) based on the exporting behavior of French firms.[17] In what follows I present the predictions of the models based on estimates from the Poisson PML specification, as this is the specification that has been employed most often in recent years.[18]

5.1 Changes in Trade Costs

The first set of counterfactual experiments involve changes in trade barriers. The first addresses perhaps the most common question asked in the international trade literature: what would be the effect of a decrease in barriers to trade? To address this question, I consider a uniform reduction in all trade barriers. However, many argue that this is not a realistic experiment, as a large component of trade barriers is made up of natural impediments, such as physical distance, so I also perform a second experiment, the elimination of the asymmetric component of trade barriers. As was recently argued by Waugh (2010), most natural barriers to trade, like distance, are inherently symmetric, so to the degree that the barriers are estimated to be asymmetric, they may also be policy related.

5.1.1 Uniform Fall in Trade Barriers

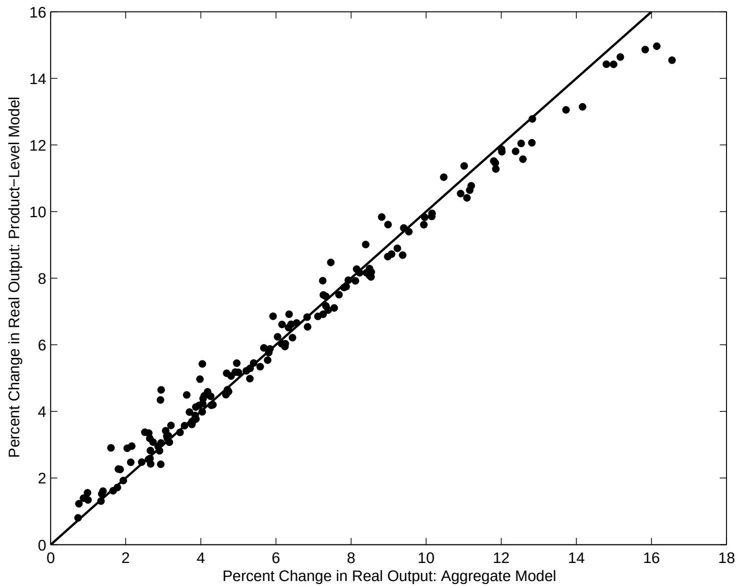
I first consider a uniform 10% reduction in all trade barriers. The first two columns of Table 4 present the predicted changes in real output for each country by the aggregate and product-levels models, respectively. The results are also summarized in Figure 10.

As can be seen from the figure, the gains from the fall in trade barriers are quite similar in both models, with growth in real output ranging from slightly less than 1% in both to about 15% in the product-level model and about 16%

[17] These parameters are important for the magnitude of changes in the variables of interest discussed below, but choosing different values within a reasonable range does not have a significant effect on the differences in the predictions of the aggregate and product-level models, which are the focus of this paper.

[18] The results are not qualitatively different when based on other specifications, particularly for the product-level model, for which the estimates of trade costs are very similar across specifications.

Figure 10: Percent Change in Welfare From a 10% Fall in Trade Barriers



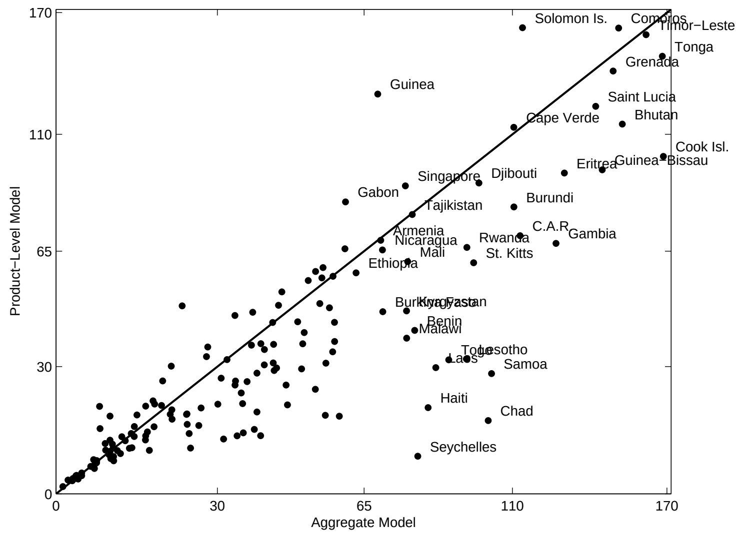
in the aggregate model, with a median of about 6% for both. This indicates that the effects of small, uniform changes in trade barriers, do not depend heavily on the composition of countries' output and expenditure.

5.1.2 Elimination of Asymmetric Trade Barriers

Next, I the remove the asymmetric component of trade barriers, setting all trade costs to be equal to the smaller value of the two directions of trade between a given pair of countries, i.e. $d'_{ni} = \min\{d_{ni}, d_{in}\}$. Table 4 presents the predicted changes in real manufacturing output for each country based on the aggregate and product-level models. Figure 11 summarizes these predictions, plotting the percentage change in real manufacturing output predicted by the product-level model against that of the aggregate model.

The figure shows that, unlike in previous experiment, there are substantial differences between the two models, particularly that the countries that benefited the most in the aggregate model are predicted to benefit much less by the product-level model. Further, it appears that the countries for which the difference is greatest tend to be developing countries, suggesting that ac-

Figure 11: Percent Change in Welfare From Elimination of Asymmetric Trade Barriers



counting for the composition of trade is important for predicting the effect of asymmetric trade barriers on income inequality.

To explore this phenomenon in more depth, I incorporate data on the size of the manufacturing workforce, obtained from UNIDO's INDSTAT database for 80 of the 148 countries in the sample. Figure 12 plots the changes in real manufacturing output per worker predicted by each model against its initial value. While both models predict larger increases in output per worker for those countries with low initial levels, this pattern is much weaker in the product-level model's predictions.

In addition, Table 3 shows that the difference is quantitatively important. While the aggregate model predicts a 12.3% fall in the variance of log output per worker, the product-level model predicts only a 9.7% fall. When inequality is measured by the 90/10 ratio in output per worker, the difference is even more striking. Compared to the 16.8% fall in the ratio predicted by the aggregate model, the product-level model predicts only a 1.2% fall.

For comparison, the table also includes the predicted measures of inequality in autarky and in a world of frictionless trade. With frictionless trade, the

Figure 12: Welfare Gains and Initial Output Per Worker

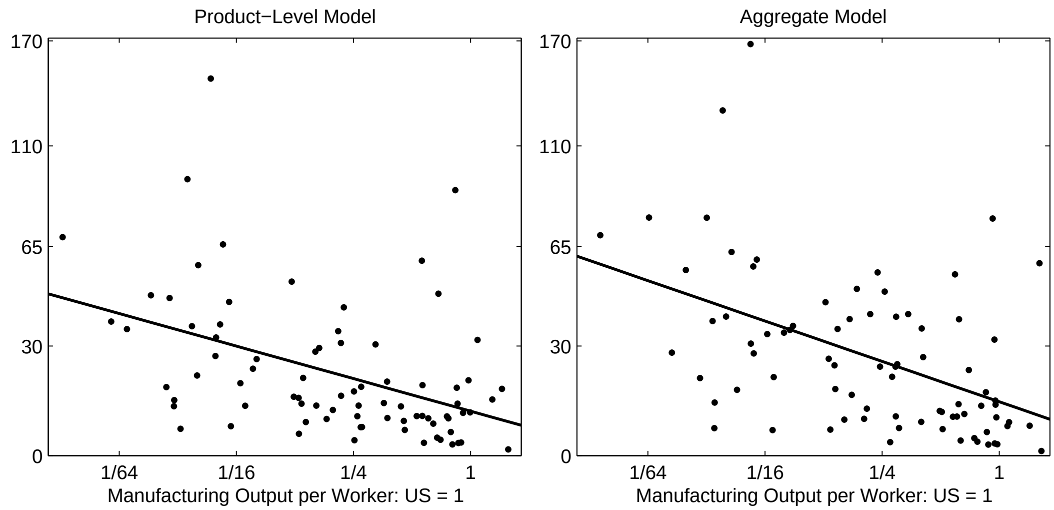


Table 3: Variance in Log Real Output per Worker

		Baseline	Autarky	$\min(d_{ni}, d_{in})$	$d_{ni} = 1$
var $(\log(y))$	Aggregate	1.65	1.78	1.45	1.21
	Product-Level	1.67	1.98	1.51	1.23
y_{90}/y_{10}	Aggregate	27.9	34.5	23.5	18.3
	Product-Level	29.9	37.4	29.5	19.2

measures are nearly identical, which is not surprising since Proposition 1 shows that the product-level and aggregate models are consistent with one other in this case, given uniform demand shifters across countries. In other words, differences in demand conditions across countries, alone, contribute little to income inequality. More interestingly, Table 3 shows that autarky is much worse for developing countries in the product-level model than in the aggregate model, which reflects the fact that, as the elasticity index suggested, developed countries dominate the production of many of the world's products, which the other countries lose the ability to consume in autarky.

5.2 Changes in Technology

In addition to predicting the effects of changes in trade barriers, the model is also capable of making predictions about the effects of changes in countries' levels productivity. Further, while the aggregate model is only able to make predictions about changes in countries' overall levels of productivity, the product-level model is also able to make predictions about changes in countries' relative productivity across products. In this section, I discuss two such experiments: one involving a change in a country's overall level of productivity – which allows for a comparison between the predictions of the aggregate and product-level models – and one involving a change in countries' relative productivity – highlighting the contribution of differences in product-level productivity differences to world trade flows and welfare.

5.2.1 Rise of China

First, I consider a positive shock to the overall productivity level of China which increases China's nominal manufacturing output as a fraction of total world manufacturing output by 300%, roughly the magnitude that was observed between 2000 and 2007. Table 4 presents the results for both models.

While both models predict a modest median gain in real output of about one-half of one percent, there are substantial differences in the gains for many countries across the models. The gains predicted by the aggregate model are heavily influenced by geography, with the greatest beneficiaries being nearby Asian countries such as Mongolia, Sri Lanka, Singapore, Nepal, Laos, and Vietnam. Other than Lesotho – the only country to have a loss in real output – the countries that benefit the least are mostly large European countries, including Germany, Italy, and France.

The picture that emerges from the product-level model is much different, demonstrating the importance of taking into account the composition of output and expenditure across countries. The product-level model predicts that the list of countries whose real output falls also includes Honduras and Cambodia, which is not surprising, given that, like China, a large fraction of the exports of these countries are simple manufactures, such as textiles, that are bound for developed countries. The set of countries that are predicted to benefit the most by the product-level model, in addition to some of the same Asian countries as with the aggregate model, includes several countries from sub-Saharan Africa, such as Angola, Guinea-Bissau, and Gambia, reflecting the increased demand by China for basic materials.

Looking at countries that made large moves through the distribution of the gains from the growth of China reveals a similar story. The set of countries with the largest fall in gains relative to the other countries is mostly made up of Asian and Central American countries known for exporting simple manufactures to developed countries. The set that experienced the largest upward move are a mixture of developed countries, such as the Netherlands, Belgium, and the United States, which benefit largely from a fall in prices, and countries that export a large amount of basic materials, such as Peru, Chile, Kuwait, Saudi Arabia, and Algeria, which benefit from increased demand for their products.

5.2.2 Removal of Comparative Advantage Across Products

To assess the degree to which trade driven by comparative advantage across products contributes to world welfare, I perform the counterfactual experiment of removing countries' deterministic differences in productivity across products. In other words, I impose that $T_i^j = T_i T^j$ for all i and j. I determine the values of T_i and T^j by first selecting a reference country, i_0, setting $T_{i_0} = 1$, and then setting $T^j = T_{i_0}^j$ for all j. I choose T_i for all $i \neq i_0$ in order keep the level of output of each country equal to its value in the data (and the baseline case).[19]

I choose the United States as the initial reference country due to the fact that it exports varieties of every product in the data and its position as the largest producer and exporter in the world. Table 5 presents the effect of this change on the real output of each country. Imposing that all countries have the same pattern of productivity across products as the US results in a 15% decrease in real output, on average. Table 5 also presents the effects of the same experiment with Germany and China as the reference countries, showing

[19] This is essentially the experiment conducted by Costinot et al. (2012), with the major difference being that this experiment is harmonizing productivity across 4,612 product categories instead of 13 much more aggregated industries.

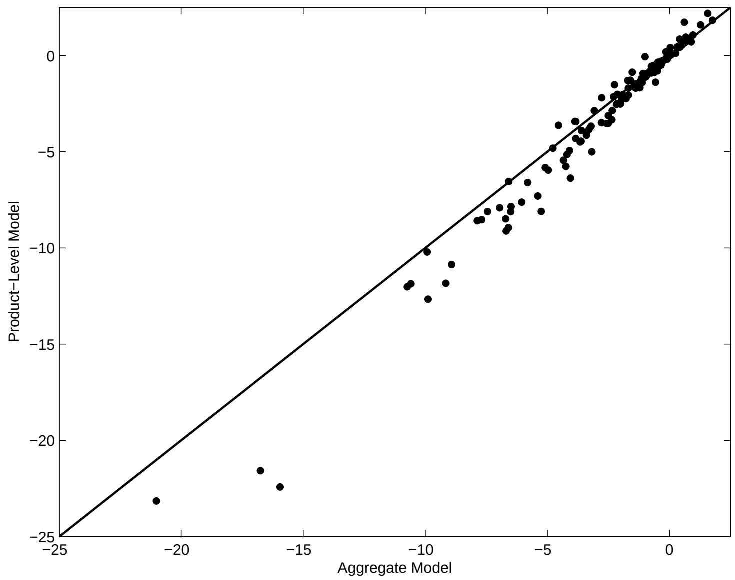
relatively similar results, with average losses in real output of 8% and 18%, respectively.

Interestingly, it seems to be the relatively isolated countries, such as Lesotho, Myanmar, and Iran, that are most hurt by the removal of comparative advantage. This turns out to be largely a function of the differences in preferences across countries. Imposing that every country also has the base country's pattern of β_n^j leads to the opposite prediction, that these countries that trade very little are the biggest beneficiaries of the removal of comparative advantage. Also, in this experiment, the reference countries' losses are more severe, with the the US, Germany, and China losing 3%, 15%, and 4%, respectively, of their real output. All this indicates that there is positive correlation between preferences and technology within countries. This finding is consistent with that of Costinot et al. (2012), who point out that the phenomenon could be due to endogenous preference formation, as in Atkin (2010), or due to directed technical change, as in Acemoglu (2003). It may also be evidence of the endogenous location of production to places of high demand to avoid trade costs, as in Hillberry and Hummels (2002).

5.3 Global Rebalancing

Finally, I consider a change in countries' trade imbalances. As in Dekle et al. (2008), I impose balanced trade on the world, setting $D_n = 0$ for every n. The resulting changes in real output are listed in Table 4, and Figure 13 illustrates the differences in the predictions of the two models. Generally, the direction and level of change in real output for a country depends on the sign and size of its initial imbalance. However, the figure shows that the predicted changes are generally larger in absolute value in the product-level model than in the aggregate model. More precisely, for all but Chile the predicted sign of the change is the same for both models, and for two-thirds of the countries, the magnitude of the change is larger for the product-level model.

Figure 13: Percent Change in Welfare From Global Rebalancing



6 Conclusion

This paper demonstrates that accounting for composition of countries output and expenditure is important for understanding the magnitudes of their aggregate bilateral trade flows. If trade flows can be characterized by a product-level gravity equation, only under extreme circumstances can aggregate trade flows be characterized by a gravity equation that depends only on aggregate output and expenditure and bilateral trade costs. As a result, even if trade costs do not vary across products, they cannot be inferred from aggregate trade data and must be estimated using product-level data.

I have shown that trade costs can be consistently estimated using product-level trade data even when no comparably disaggregated output data is available. Estimation based on this procedure produces estimates of trade costs that are generally lower than those from estimations based on aggregate data, and they are more robust to assumptions on the form of heteroskedasticity present in the data, which indicates that it is the failure to account for the composition of trade flows – and not incorrect assumptions about the form of heteroskedasticity – that is the major cause of bias in aggregate gravity estimations.

Counterfactual experiments conducted using a gravity model fitted to the estimates using product-level data, in many case, predict very different outcomes as a result of changes in trade costs and technology. The removal of asymmetric trade barriers has a much smaller effect on income inequality than is predicted by an aggregate model, and the effect of the rise of China as a major manufacturer and exporter has very different on a number of countries under the two models. Countries in Asia and Central America that export a similar set of products to a similar set of countries as China benefit much less – and in some cases lose – once the composition of output and expenditure is taken into account, and countries in the global north that intensively import the set of goods China produces and countries in the South America and Sub-Saharan Africa that intensively export the set of goods China that China demands benefit much more.

References

Acemoglu, Daron, "Patterns of Skill Premia," *Review of Economic Studies*, 04 2003, *70* (2), 199–230.

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

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

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

_ **and Yoto V. Yotov**, "Specialisation: Pro and Anti-Globalizing 1990-2002," *Working Paper*, 2011.

Atkin, David, "Trade, tastes and Nutrition in India," April 2010, (986).

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety*," *Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Chaney, Thomas, "Distorted Gravity: the Intensive and Extensive Margins of International Trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

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

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, 05 2007, *75* (3), 837–873.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Global Rebalancing with Gravity: Measuring the Burden of Adjustment," *IMF Staff Papers*, July 2008, *55* (3), 511–540.

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

_ , _ , **and Francis Kramarz**, "An Anatomy of International Trade: Evidence From French Firms," *Econometrica*, 2011, *79* (5), 1453–1498.

_ , _ , **and Sebastian Sotelo**, "International Trade: Linking Micro and Macro," 2012, *NBER Working Paper 17864.*

Fieler, Anna C., "Nonhomotheticity and Bilateral Trade: Evidence and a Quantitative Explanation," 2010, *Working Paper.*

French, Scott, "Innovation and Product Level Trade," 2009, *Working Paper.*

Gourieroux, C., A. Monfort, and A. Trognon, "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, 1984, *52* (3), pp. 681–700.

Hanson, Gordon H. and Raymond Robertson, "China and the Manufacturing Exports of Other Developing Countries," in "China's Growing Role in World Trade" NBER Chapters, National Bureau of Economic Research, Inc, 2010, pp. 137–159.

Helpman, Elhanan and Paul Krugman, *Market Structure and Foreign Trade: Increasing Returns, Imperfect Competition, and the International Economy*, Vol. 1 of *MIT Press Books*, The MIT Press, December 1987.

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

Hillberry, Russell and David Hummels, "Explaining Home Bias in Consumption: The Role of Intermediate Input Trade," June 2002, (9020).

Hillberry, Russell H., "Aggregation Bias, Compositional Change, and the Border Effect," *The Canadian Journal of Economics / Revue canadienne d'Economique*, 2002, *35* (3), pp. 517–530.

Hummels, David, "Toward a Geography of Trade Costs," *GTAP Working Paper*, 1999.

Levchenko, Andrei A. and Jing Zhang, "The Evolution of Comparative Advantage: Measurement and Welfare Implications," NBER Working Papers 16806, National Bureau of Economic Research, Inc February 2011.

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

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

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

Table 4: Counterfactual Percent Changes in Real Income

Country	10% Reduction		Symmetric		Rise of China		Rebalancing	
	Agg.	P.L.	Agg.	P.L.	Agg.	P.L.	Agg.	P.L.
Albania	15.17	14.64	38.04	14.19	0.74	0.35	-1.52	-0.87
Algeria	11.16	10.64	18.92	26.24	0.38	0.74	-0.11	-0.21
Angola	7.24	7.92	43.12	29.65	1.43	2.17	-7.69	-8.53
Argentina	2.65	2.59	6.55	6.35	0.23	0.32	-0.55	-0.65
Armenia	10.47	11.03	69.53	68.72	0.88	0.81	-3.08	-2.86
Australia	3.77	3.70	4.26	3.81	0.46	0.71	-1.44	-1.48
Austria	7.81	7.72	10.47	9.33	0.25	0.34	-0.70	-0.74
Azerbaijan	4.06	4.38	57.29	36.92	0.49	0.70	-2.14	-2.50
Bangladesh	3.76	3.61	20.60	30.14	0.41	0.01	0.25	0.11
Belarus	3.87	4.13	33.79	25.14	0.45	0.60	-2.17	-2.53
Belgium	11.80	11.52	7.33	19.77	0.25	0.40	-0.36	-0.46
Benin	2.61	3.35	79.14	40.10	0.99	1.42	-6.49	-7.85
Bhutan	2.66	2.83	151.09	114.48	0.63	0.39	-0.57	-0.56
Bolivia	4.09	4.47	40.33	34.68	0.63	0.84	-2.79	-3.49
Bosnia	9.96	9.83	56.88	56.64	0.80	1.06	-3.84	-3.43
Brazil	1.78	1.72	3.27	3.75	0.13	0.10	-0.10	-0.06
Brunei	2.73	3.08	56.79	34.05	0.65	0.70	-2.36	-3.34
Bulgaria	8.54	8.18	30.79	26.95	0.36	0.33	-0.67	-0.74
Burkina Faso	1.00	1.34	70.10	45.62	0.43	0.63	-2.51	-3.53
Burundi	0.98	1.56	110.50	80.75	0.76	1.36	-5.25	-8.10
C.A.R.	2.13	2.47	112.66	70.33	0.39	0.52	-1.78	-2.24
Cambodia	5.31	4.98	55.97	46.79	0.38	-0.25	0.89	0.71
Canada	8.49	8.09	13.00	13.25	0.13	0.23	-0.07	-0.04
Cape Verde	3.62	4.49	110.49	113.06	1.26	1.73	-8.92	-10.86
Chad	0.87	1.39	101.89	16.30	0.50	0.87	-3.18	-5.01
Chile	4.67	4.50	9.17	11.70	0.30	0.71	-0.14	0.19
China	2.66	2.43	1.96	2.87	277.57	284.01	0.86	0.89
Colombia	3.13	3.27	23.73	17.91	0.30	0.30	-1.18	-1.37
Comoros	4.04	5.43	149.60	161.53	1.43	2.13	-9.89	-12.66
Cook Isl.	11.01	11.37	168.44	100.64	1.51	1.91	-6.58	-6.55
Croatia	6.04	6.24	35.43	20.44	0.45	0.55	-1.96	-2.27
Cyprus	6.34	6.52	39.46	12.74	0.93	1.19	-4.96	-5.95
Czech Rep.	7.87	7.75	15.68	12.70	0.19	0.17	-0.15	-0.17
Cte d'Ivoire	3.09	3.26	30.10	20.31	0.27	0.35	-0.63	-0.80
Denmark	7.33	7.17	13.12	9.95	0.21	0.30	-0.33	-0.34
Djibouti	3.98	4.97	98.86	89.94	1.98	2.93	-21.02	-23.15
Dom. Rep.	8.11	7.92	31.30	11.96	0.53	0.31	-1.36	-1.69
Ecuador	3.11	3.19	40.30	30.49	0.27	0.28	-0.96	-1.10
Egypt	1.40	1.61	13.58	14.88	0.37	0.45	-2.01	-2.52
El Salvador	4.04	3.99	42.56	28.97	0.17	0.02	-0.08	-0.15
Eritrea	2.04	2.89	128.53	93.88	0.97	1.56	-6.69	-9.11
Ethiopia	2.51	3.38	62.87	57.79	0.87	1.39	-5.39	-7.30
Finland	4.72	4.58	9.61	10.76	0.00	0.09	0.96	1.06
France	3.87	3.77	2.77	3.18	0.11	0.16	-0.13	-0.16
Gabon	5.92	6.86	60.09	82.66	1.24	1.64	-6.96	-7.92
Gambia	8.15	8.27	125.42	67.68	1.28	1.75	-7.45	-8.11
Germany	4.32	4.20	4.26	4.39	0.04	0.11	0.44	0.44
Ghana	9.54	9.40	42.23	42.40	1.03	1.27	-3.87	-3.42
Greece	4.05	4.23	15.65	11.74	0.62	0.76	-3.39	-3.98
Grenada	8.98	9.61	147.40	139.31	1.28	1.84	-7.87	-8.58

Table 4 – Counterfactual Percent Changes in Real Income (cont.)

Country	10% Reduction		Symmetric		Rise of China		Rebalancing	
	Agg.	P.L.	Agg.	P.L.	Agg.	P.L.	Agg.	P.L.
Guatemala	6.15	6.03	42.36	30.99	0.39	0.15	-1.00	-1.13
Guinea	9.94	9.60	68.72	128.22	1.13	1.64	-4.55	-3.62
Guinea-Bissau	2.93	4.34	143.01	95.17	1.25	2.15	-9.16	-11.83
Haiti	6.55	6.66	83.10	19.47	0.75	0.60	-3.40	-4.14
Honduras	12.58	11.57	37.39	35.89	0.47	-0.38	0.48	0.75
Hungary	10.92	10.54	9.20	9.17	0.36	0.23	-0.32	-0.30
Iceland	10.16	9.95	38.62	18.39	0.51	0.84	-0.74	-0.56
India	1.66	1.62	6.43	5.35	0.32	0.14	0.51	0.60
Indonesia	7.25	6.92	6.31	7.30	0.83	0.52	1.28	1.59
Iran	0.73	0.80	9.81	7.97	0.15	0.19	-0.74	-0.91
Iraq	6.16	6.61	52.44	24.05	1.47	1.62	-10.74	-12.02
Ireland	12.01	11.87	7.41	14.39	0.19	0.24	0.69	0.75
Israel	7.26	7.50	9.75	9.94	0.31	0.33	-0.16	-0.21
Italy	3.16	3.07	2.68	2.72	0.09	0.09	-0.01	-0.03
Jamaica	8.41	8.16	58.49	17.35	0.56	0.82	-1.68	-1.69
Japan	1.34	1.31	1.12	1.49	0.32	0.35	0.40	0.47
Jordan	4.18	4.59	38.64	28.28	0.71	1.12	-3.84	-4.32
Kazakhstan	8.49	8.29	27.72	32.72	1.92	2.44	-1.00	-0.06
Kenya	2.94	3.05	35.13	23.13	0.53	0.61	-2.35	-2.86
Korea	3.45	3.37	2.94	3.09	0.79	0.80	0.52	0.72
Kuwait	5.81	5.77	22.75	47.37	0.25	0.65	0.04	0.09
Kyrgyzstan	6.40	6.61	76.79	45.86	1.72	1.47	-2.29	-2.14
Laos	4.91	5.18	85.42	29.74	1.99	1.63	-2.57	-3.53
Latvia	16.14	14.97	33.77	44.49	0.59	0.63	-0.30	-0.35
Lebanon	2.64	3.19	33.88	26.20	0.96	1.27	-6.50	-8.11
Lesotho	2.93	2.41	94.88	32.11	-0.05	-0.14	1.76	1.83
Lithuania	8.98	8.65	36.42	26.03	0.39	0.39	-0.79	-0.82
Luxembourg	14.99	14.42	16.38	9.36	0.33	0.36	-0.44	-0.35
Macedonia	7.68	7.50	59.94	65.82	0.32	0.12	-0.50	-0.45
Madagascar	7.12	6.85	49.72	39.47	0.55	0.34	-0.41	-0.50
Malawi	4.69	5.15	76.84	37.87	1.14	1.36	-6.71	-8.49
Malaysia	11.84	11.46	20.77	16.63	1.44	0.89	-0.34	-0.50
Mali	6.35	6.92	77.13	61.46	1.44	1.62	-9.93	-10.21
Malta	12.01	11.80	35.60	13.40	0.55	0.43	-1.08	-0.93
Mauritius	9.37	8.69	34.24	12.69	0.68	0.61	-0.49	-0.80
Mexico	13.72	13.05	11.30	12.50	0.37	0.18	-0.81	-0.83
Mongolia	16.55	14.55	27.96	35.39	6.69	3.78	-0.68	-0.87
Morocco	5.68	5.91	24.15	13.23	0.29	0.19	-0.72	-0.70
Mozambique	3.70	3.98	49.07	29.41	0.94	1.17	-5.09	-5.82
Myanmar	1.36	1.52	32.05	31.91	0.35	0.30	-0.63	-0.88
Namibia	9.40	9.51	45.69	20.15	1.05	1.23	-3.60	-3.90
Nepal	7.93	7.94	53.54	48.08	2.01	1.03	-1.21	-1.67
Netherlands	11.85	11.28	14.06	17.65	0.25	0.43	0.07	0.03
New Zealand	5.58	5.34	11.04	8.66	0.40	0.59	-0.40	-0.37
Nicaragua	6.83	6.83	70.00	65.42	0.52	0.44	-1.68	-2.06
Norway	5.84	5.88	12.68	9.83	0.31	0.47	-1.01	-1.06
Oman	9.23	8.90	24.45	9.88	1.04	0.92	-4.09	-4.94
Pakistan	2.61	2.56	18.70	19.99	0.52	0.31	0.33	0.39
Panama	11.09	10.41	48.12	42.62	0.61	0.26	-0.96	-0.95
Papua N.G.	12.38	11.81	50.69	55.29	0.82	0.68	-1.70	-1.29

Table 4 – Counterfactual Percent Changes in Real Income (cont.)

Country	10% Reduction		Symmetric		Rise of China		Rebalancing	
	Agg.	P.L.	Agg.	P.L.	Agg.	P.L.	Agg.	P.L.
Paraguay	3.06	3.42	37.69	45.46	0.76	1.22	-4.19	-5.14
Peru	2.86	2.95	15.69	19.81	0.24	0.62	-0.47	-0.35
Philippines	12.83	12.78	16.02	13.60	1.55	0.75	0.04	0.41
Poland	3.85	3.88	11.92	11.55	0.23	0.22	-0.93	-1.02
Portugal	5.41	5.45	9.83	7.04	0.37	0.29	-1.31	-1.46
Qatar	12.81	12.06	23.76	15.40	0.92	1.53	-2.13	-2.02
Romania	5.21	5.22	20.71	18.91	0.20	0.11	-0.29	-0.28
Russia	6.84	6.54	6.81	6.62	0.33	0.53	1.57	2.19
Rwanda	0.75	1.23	95.03	66.29	0.60	1.04	-4.06	-6.37
Saint Lucia	7.46	8.47	140.44	122.52	1.45	1.80	-10.59	-11.86
Samoa	8.82	9.84	103.00	28.14	1.20	0.92	-4.77	-4.82
Saudi Arabia	4.26	4.45	9.17	17.37	0.47	0.86	-1.92	-2.04
Senegal	5.00	5.17	55.07	30.95	0.71	0.82	-3.30	-3.85
Seychelles	8.53	8.04	80.05	8.05	0.72	0.74	-1.40	-1.66
Singapore	14.80	14.42	76.48	88.81	2.21	1.88	0.42	0.85
Slovakia	11.20	10.77	17.28	14.83	0.31	0.30	-0.04	-0.06
Slovenia	10.15	9.86	23.62	17.79	0.31	0.21	-0.43	-0.46
Solomon Is.	12.53	12.04	113.50	161.72	1.21	1.16	-2.77	-2.19
South Africa	2.90	2.82	6.83	7.09	0.13	0.14	0.32	0.45
Spain	3.80	3.73	3.66	3.10	0.23	0.26	-0.75	-0.83
Sri Lanka	14.17	13.14	13.55	12.55	2.41	0.83	-0.57	-1.39
St. Kitts	15.83	14.86	97.18	61.11	0.87	0.81	-1.60	-1.28
Sudan	1.80	2.27	45.35	25.16	0.69	1.09	-4.24	-5.76
Swaziland	4.81	5.06	54.93	17.57	0.63	0.90	-2.50	-3.12
Sweden	6.25	6.03	5.80	5.79	0.09	0.15	0.63	0.66
Syria	4.96	5.45	39.51	36.32	0.71	0.88	-3.62	-4.45
Tajikistan	4.71	4.65	78.46	77.95	0.98	1.71	0.61	1.73
Tanzania	3.21	3.58	49.35	36.28	0.81	1.09	-4.34	-5.44
Thailand	9.08	8.72	9.00	8.38	0.99	0.55	0.52	0.55
Timor-Leste	1.61	2.90	160.99	157.98	1.61	3.02	-15.95	-22.42
Togo	5.78	5.54	89.37	31.85	0.81	0.82	-3.66	-4.49
Tonga	2.94	4.65	167.96	146.71	1.77	3.05	-16.75	-21.58
Trin. and Tob.	6.44	6.21	54.35	59.49	0.23	0.29	0.03	0.14
Tunisia	7.33	7.45	26.12	15.14	0.41	0.19	-1.15	-1.20
Turkey	3.57	3.57	8.40	9.43	0.33	0.27	-1.39	-1.68
Turkmenistan	8.39	9.01	52.50	58.22	1.13	1.44	-5.81	-6.60
U.A.E.	7.39	7.04	9.31	7.51	0.82	1.19	-3.21	-3.67
U.K.	4.28	4.18	3.38	3.86	0.21	0.28	-0.55	-0.61
U.S.A	1.93	1.93	8.32	10.92	0.24	0.36	-0.62	-0.71
Uganda	1.85	2.26	57.25	42.45	0.92	1.25	-6.05	-7.62
Ukraine	6.24	5.95	20.41	17.84	0.10	0.14	0.67	0.96
Uruguay	5.31	5.29	26.59	19.38	0.37	0.53	-0.60	-0.53
Uzbekistan	3.95	4.18	43.58	47.55	0.73	0.78	-1.44	-1.56
Venezuela	2.43	2.48	17.39	20.33	0.15	0.22	-0.35	-0.39
Viet Nam	7.55	7.10	17.08	21.13	1.71	0.40	-1.12	-1.40
Yemen	2.16	2.96	44.36	51.69	0.98	1.61	-6.59	-8.95
Zambia	8.23	8.16	54.07	56.13	0.82	1.40	-2.25	-1.52
Zimbabwe	4.74	4.60	42.44	36.11	0.37	0.25	-0.66	-0.51

Table 5: Counterfactual Percent Changes in Real Income

Country	U.S.	Ger.	China	Country	U.S.	Ger.	China	Country	U.S.	Ger.	China
Albania	-18.04	-7.26	-8.81	Guatemala	-15.86	-5.24	-19.06	Paraguay	-16.92	-8.04	-12.10
Algeria	-12.80	-2.82	-22.09	Guinea	-13.09	-2.82	-15.40	Peru	-15.09	-5.17	-15.04
Angola	-10.88	-5.23	-17.08	Guinea-Bissau	-26.82	-13.75	-14.38	Philippines	4.51	-2.20	-19.88
Argentina	-12.10	-0.91	-20.52	Haiti	-22.68	-20.57	-8.54	Poland	-16.49	-5.22	-15.49
Armenia	-12.03	-8.00	-21.35	Honduras	-15.48	-6.76	-17.37	Portugal	-21.57	-9.74	-18.07
Australia	-11.16	-4.83	-19.52	Hungary	-8.21	-2.20	-6.54	Qatar	-8.35	-2.23	-11.73
Austria	-12.89	-4.68	-11.63	Iceland	-17.97	-12.14	-16.18	Romania	-19.40	-9.62	-10.43
Azerbaijan	-13.77	-12.73	-24.51	India	-17.95	-14.32	-27.34	Russia	-17.27	-10.21	-28.53
Bangladesh	-18.21	-7.93	-6.69	Indonesia	-8.75	-0.03	-20.97	Rwanda	-24.84	-18.25	-31.57
Belarus	-21.51	-9.88	-26.36	Iran	-29.35	-17.77	-36.43	Saint Lucia	-11.63	-2.44	-6.38
Belgium	-11.41	-8.67	-15.46	Iraq	-10.94	-4.00	-14.39	Samoa	-7.91	-11.35	-24.64
Benin	-22.19	-14.72	-15.43	Ireland	-6.09	-10.28	-9.52	Saudi Arabia	-9.45	-2.45	-15.97
Bhutan	-19.97	-5.58	-29.48	Israel	-12.80	-18.14	-24.74	Senegal	-18.74	-12.99	-18.99
Bolivia	-15.88	-11.38	-23.32	Italy	-17.53	-7.73	-15.87	Seychelles	-28.51	-22.78	-26.02
Bosnia	-16.79	-3.92	-12.99	Jamaica	-18.87	-12.61	-19.06	Singapore	-2.98	-6.17	-13.95
Brazil	-11.44	-2.58	-24.44	Japan	-7.01	-5.70	-20.44	Slovakia	-12.31	-0.83	-15.35
Brunei	-17.87	-13.57	-16.41	Jordan	-17.65	-8.10	-23.36	Slovenia	-16.17	-4.30	-17.04
Bulgaria	-17.93	-8.14	-16.11	Kazakhstan	-9.27	2.01	-19.58	Solomon Is.	-8.34	-0.24	-10.83
Burkina Faso	-24.54	-18.80	-25.92	Kenya	-18.01	-8.11	-24.45	South Africa	-9.05	-1.02	-16.57
Burundi	-19.46	-10.30	-29.97	Korea	-2.44	-9.42	-24.14	Spain	-17.30	-3.35	-20.27
C.A.R.	-15.61	-12.51	-36.32	Kuwait	-1.94	7.02	-4.63	Sri Lanka	-9.37	-0.89	-11.00
Cambodia	-20.39	-9.01	-7.13	Kyrgyzstan	-15.84	-5.60	-24.98	St. Kitts	-9.31	-0.27	-5.60
Canada	-9.16	1.72	-17.00	Laos	-30.86	-23.19	-19.70	Sudan	-22.06	-15.04	-25.06
Cape Verde	-17.80	-8.25	-15.80	Latvia	-13.02	-2.20	-12.63	Swaziland	-24.56	-16.68	-26.10
Chad	-32.31	-35.66	-48.89	Lebanon	-19.18	-10.53	-17.13	Sweden	-10.70	-4.38	-16.67
Chile	-11.76	-2.65	-13.70	Lesotho	-44.45	-45.17	-27.14	Syria	-17.40	-6.51	-21.87
China	-6.94	-2.98	-1.94	Lithuania	-17.69	-5.51	-16.99	Tajikistan	-7.40	-12.69	-12.59
Colombia	-18.08	-9.06	-22.94	Luxembourg	-11.37	-5.26	-15.91	Tanzania	-17.07	-7.48	-16.70
Comoros	-18.89	-11.28	-14.03	Macedonia	-17.49	-5.42	-19.40	Thailand	-1.65	0.00	-12.60
Cook Isl.	-12.48	-6.89	-10.41	Madagascar	-21.73	-11.40	-6.12	Timor-Leste	-14.78	-3.03	-21.35
Croatia	-20.77	-13.02	-18.08	Malawi	-19.42	-8.90	-20.41	Togo	-26.52	-15.99	-24.52
Cyprus	-13.90	-9.56	-20.55	Malaysia	3.19	-1.88	-18.55	Tonga	-16.04	-8.29	-20.59
Czech Rep.	-13.46	-3.95	-13.67	Mali	-16.97	-6.76	-21.33	Trin. and Tob.	-10.15	-1.07	-17.87
Cte d'Ivoire	-23.35	-18.45	-24.54	Malta	-0.21	-8.65	-18.81	Tunisia	-18.52	-12.93	-12.40
Denmark	-14.68	-8.93	-9.62	Mauritius	-23.64	-12.52	-10.95	Turkey	-16.67	-5.33	-19.46
Djibouti	-12.21	-2.56	-12.40	Mexico	-5.24	3.42	-6.38	Turkmenistan	-12.96	0.50	-16.43
Dom. Rep.	-19.39	-14.22	-5.66	Mongolia	-8.22	5.43	-14.45	U.A.E.	-11.45	-7.60	-9.06
Ecuador	-20.07	-10.34	-23.42	Morocco	-19.18	-14.43	-14.26	U.K.	-7.66	-4.85	-10.14
Egypt	-15.74	-12.89	-21.61	Mozambique	-20.08	-10.13	-21.26	U.S.A	0.00	-9.74	-17.82
El Salvador	-22.08	-12.34	-22.66	Myanmar	-28.67	-20.39	-19.87	Uganda	-19.07	-9.03	-21.43
Eritrea	-16.63	-6.09	-25.01	Namibia	-16.18	-6.41	-15.28	Ukraine	-18.88	-7.76	-18.71
Ethiopia	-11.00	-8.24	-27.95	Nepal	-18.29	-7.42	-21.18	Uruguay	-19.10	-7.93	-19.67
Finland	-11.48	-6.40	-15.99	Netherlands	-7.61	-7.33	-5.61	Uzbekistan	-10.90	0.84	-20.07
France	-10.53	-6.61	-15.36	New Zealand	-13.11	-5.52	-17.27	Venezuela	-15.86	-5.38	-20.44
Gabon	-8.11	-1.75	-16.32	Nicaragua	-16.78	-4.48	-13.55	Viet Nam	-12.31	-0.46	-12.32
Gambia	-19.07	-6.93	-11.82	Norway	-15.86	-12.44	-15.79	Yemen	-20.18	-11.31	-22.00
Germany	-8.64	-0.08	-12.76	Oman	-12.02	-2.54	-25.70	Zambia	-12.65	-1.02	-17.69
Ghana	-6.44	-4.64	-18.18	Pakistan	-19.61	-12.30	-28.70	Zimbabwe	-11.74	-2.32	-18.46
Greece	-20.81	-11.21	-14.81	Panama	-11.19	2.06	-10.63				
Grenada	-9.91	-0.65	-5.79	Papua N.G.	-9.47	-2.59	-15.90				

A Data

A.1 Bilateral Trade Data

The product-level bilateral trade data are taken from the Comtrade database provided by the United Nations Statistical Division. The data used are those reported by national statistical agencies using the 1996 revision of the Harmonized Commodity Description and Coding System (HS) for the year 2000. The choice of year was made to maximize the number of countries reporting in a common product classification. The data report trade flows valued in US dollars reported by both importers and exporters. I use trade flows reported by importers as these are generally considered to be of higher quality. The data cover 247 countries and 5,132 product categories.

A.2 Gravity Variables

The bilateral relationship variables – distance, shared border, common language, and colonial relationship – are from CEPII's dist_cepii database described in Mayer and Zignago (2011). I use their measures of population-weighted distances, whether countries share a common official language, and wether countries ever had a colonial link.

A.3 Manufacturing Output

Output data for the manufacturing sector is taken primarily from the United Nations Industrial Development Organization's (UNIDO) Industrial Statistics (INDSTAT) database. Data are available for manufacturing industries classified according to the International Standard Industrial Classification (ISIC) Revision 3, for 151 3- and 4-digit industries. Since data is only available for manufactured goods, I dropped all trade flows that did not respond to ISIC categories beginning with 15 through 37. Product-level trade flows were allocated to an ISIC industry using the concordance available from Comtrade. Where manufacturing output data from INDSTAT was unavailable, I imputed manufacturing output from data on manufacturing value added available from the World Bank's World Development Indicators (WDI). Manufacturing output data were available from INDSTAT for 77 countries, while output data could be imputed from WDI for an additional 71 countries. As a result, the final dataset is made up up 148 countries and 4,618 manufacturing product categories.[20]

B Proofs

B.1 Proposition 1

Proof of Proposition 1, Case 1. Because $\frac{Y_i^j}{Y^j} = \alpha^j$, it follows immediately that $\sum_j \alpha^j = 1$ and $\frac{Y_i^j}{Y_j} = \frac{Y_i}{Y}$. Conjecture that $\Pi_i^j = \lambda^j \Pi_i$, $\forall i, j$. Combined with (7) we have that

$$\begin{aligned} (P_n^j)^{-\theta} &= (\lambda^j)^\theta \sum_i \left(\frac{d_{ni}}{\Pi_i}\right)^{-\theta} \frac{Y_i}{Y} \\ &\equiv (\lambda^j)^\theta \tilde{P}_n^{-\theta}. \end{aligned}$$

Substituting into 4 (and using $\beta_n^j = \beta^j$) gives

$$\begin{aligned} (P_n)^{-\sigma} &= \sum_j \beta^j \left(\frac{\tilde{P}_n}{\lambda^j}\right)^{-\sigma} \\ &= (\tilde{P}_n)^{-\sigma} \sum_j \beta^j (\lambda^j)^\sigma. \end{aligned}$$

Substituting into 2 yields

$$X_n^j = \frac{\beta^j (\lambda^j)^\sigma}{\sum_j \beta^j (\lambda^j)^\sigma} X_n.$$

Summing over n and using the market clearing condition that $Y_i^j = \sum_n X_{ni}^j$ gives

$$Y^j = \frac{\beta^j (\lambda^j)^\sigma}{\sum_j \beta^j (\lambda^j)^\sigma} Y,$$

[20]Due to lack of industry-level data from WDI and complications associated with the industry-level INDSTAT data – i.e. a large number of categories are censored, and for the remaining categories, a large fraction are reported to have output that is less than exports indicating serious problems associated with misclassification or mis-concordance – only aggregate manufacturing output data is used in the paper.

which, given $Y^j = \alpha^j Y$, implies that $\lambda^j = \left(\frac{\alpha^j}{\beta^j}\right)^{\frac{1}{\sigma}}$ and $\tilde{P}_n = P_n$. Substituting the expressions for P_n^j and X_n^j into equation 8, we have that

$$\begin{aligned}(\Pi_i^j)^{-\theta} &= \left(\frac{\alpha^j}{\beta^j}\right)^{\frac{-\theta}{\sigma}} \sum_n \left(\frac{d_{ni}}{P_n}\right)^{-\theta} \frac{Y_n}{Y} \\ &= \left(\left(\frac{\alpha^j}{\beta^j}\right)^{\frac{1}{\sigma}} \Pi_i\right)^{-\theta}\end{aligned}$$

which verifies the conjecture. Substituting the expressions for P_n^j and Π_i^j into 10 and combining with 9 gives the result of Proposition 1, Case 1. $\square$

Proof of Proposition 1, Case 2. Given that $\frac{Y_i^j}{Y^j} \in \{0,1\}, \forall i,j$, 7 reduces to

$$P_n^j = \frac{d_{ni}}{\Pi_{i^*(j)}^j},$$

where $i^*(j)$ denotes the exporter for which $\frac{Y_i^j}{Y^j} = 1$. Substituting into 4 (and using $\beta_n^j = \beta^j$ gives that

$$(P_n)^{-\sigma} = \sum_j \beta^j \left(\frac{d_{ni}}{\Pi_{i^*}^j}\right)^{-\sigma}.$$

Substituting into 10 gives that

$$(\tilde{\Pi}_{ni})^{\theta} = \left(\frac{d_{ni}}{P_n}\right)^{\theta-\sigma} \frac{Y}{Y_i} \sum_{j \in \Omega_i^j} \beta^j (\Pi_i^j)^{-\sigma},$$

where $\Omega_i^j = \{j : i = i^*(j)\}$. Define

$$\hat{\Pi}_i \equiv \left(\frac{Y}{Y_i} \sum_{j \in \Omega_i^j} \beta^j (\Pi_i^j)^{-\sigma}\right)^{-\frac{1}{\sigma}}.$$

Then, it is clear from 9 that

$$X_{ni} = \frac{X_n Y_i}{Y} \left(\frac{d_{ni}}{P_n \hat{\Pi}_i}\right)^{-\sigma}, \tag{24}$$

and, using 4, it is straightforward to show that

$$P_n^{-\sigma} = \sum_i \left(\frac{d_{ni}}{\hat{\Pi}_i}\right)^{-\sigma} \frac{Y_i}{Y}.$$

Then, summing 24 over n, using the market clearing condition $Y_i = \sum_n X_{ni}$, and simplifying gives that

$$\hat{\Pi}_i^{-\sigma} = \sum_n \left(\frac{d_{ni}}{P_n}\right)^{-\sigma} \frac{X_n}{Y}.$$

$\square$

Proof of Proposition 1, Case 3. Given that $\sigma = \theta$ and that $\beta_n^j = \beta^j$, 10 simplifies to

$$\tilde{\Pi}_{ni} = \left(\frac{Y}{Y_i}\sum_j \beta^j (\Pi_i^j)^{-\sigma}\right)^{-\frac{1}{\sigma}} \equiv \hat{\Pi}_i.$$

The rest of the proof follows exactly as for Case 2.

$\square$

Proof of Proposition 1, Case 4. Given that $d_{ni} = 1, \forall n, i$, it follow immediately that

$$(P_n^j)^{-\theta} = \sum_i \left(\frac{1}{\Pi_i^j}\right)^{-\theta} \frac{Y_i^j}{Y^j} = (P^j)^{-\theta},$$

and that

$$(\Pi_i^j)^{-\theta} = \sum_n \left(\frac{1}{P_n^j}\right)^{-\theta} \frac{X_n^j}{Y^j} = (\Pi^j)^{-\theta},$$

which implies a system of two equations that is satisfied at $P^j = (\Pi^j)^{-1}$. Combined with 4 and given that $\beta_n^j = \beta^j$, this implies that

$$P_n^{-\sigma} = \sum_j \beta^j (\Pi^j)^{\sigma} = P.$$

Substituting for P_n^j and P_n in 2 and imposing the marking clearing condition, $\sum_n X_n^j = Y^j$, gives that

$$\frac{\beta^j}{(\Pi^j)^{\sigma}} P_n^{-\sigma} = \frac{Y^j}{Y}.$$

Substituting this into 10 and simplifying gives that

$$\tilde{\Pi}_{ni} = P^{\theta},$$

which implies that 9 reduces to

$$X_{ni} = \frac{X_n Y_i}{Y}$$

□

B.2 Reformulation of Product-Level Trade Flows

Proof of Proposition 2. Summing 6 over all $i \neq n$ and rearranging gives that

$$\frac{X_n^j}{(P_n^j)^{-\theta}} = \frac{X_{nf}^j}{(P_{nf}^j)^{-\theta}},$$

where

$$(P_{nf}^j)^{-\theta} = \sum_{i \neq n} \left(\frac{d_{ni}}{\Pi_i^j}\right)^{-\theta} \frac{Y_i^j}{Y^j}, \tag{25}$$

and $X_{nf}^j = \sum_{i \neq n} X_{ni}^j$. Substituting this expression back into 6 gives that

$$X_{ni}^j = \frac{X_{nf}^j Y_i^j}{Y^j} \left(\frac{d_{ni}}{P_{nf}^j \Pi_i^j}\right)^{-\theta}. \tag{26}$$

Summing 26 over all $n \neq i$ and rearranging gives that

$$\frac{Y_i^j}{Y^j}(\Pi_i^j)^{\theta} = \frac{Y_{if}^j}{Y_f^j}(\Pi_{if}^j)^{\theta} \tag{27}$$

where

$$(\Pi_{if}^j)^{-\theta} = \sum_{n \neq i} \left(\frac{d_{ni}}{P_{nf}^j}\right)^{-\theta} \frac{X_{nf}^j}{Y_f^j}$$

$Y_{if}^j = \sum_{n \neq i} X_{ni}^j$ and $Y_f^j = \sum_j Y_{if}^j$. Substituting this express back into 26 yields

$$X_{ni}^j = \frac{X_{nf}^j Y_{if}^j}{Y_f^j} \left(\frac{d_{ni}}{P_{nf}^j \Pi_{if}^j}\right)^{-\theta}.$$

Substituting 27 into 25 completes the specification given in Proposition 2.

□

C Models Generating Product-Level Gravity

In this appendix, I show that generalizations of two well-known trade models – the Ricardian model of Eaton and Kortum (2002) and the monopolistic model of Chaney (2008) – generate the product-level gravity framework laid out in section 2. This amounts to showing that both are consistent with the reduced-form demand system given by (1) - (4).

In both models, there are a continuum of varieties, indexed by k, in each product category. A representative consumer in each country, n, maximizes the following nested-CES utility function:

$$U_n = \left(\sum_j (\hat{\beta}_n^j)^{\frac{1}{1+\gamma}} \left(\int_{\Omega_n^j} q_n^j(k)^{\frac{\gamma}{1+\gamma}} dk \right)^{\frac{1+\gamma}{\gamma} \frac{\delta}{1+\delta}} \right)^{\frac{1+\delta}{\delta}}$$

subject to

$$\sum_j \int_{\Omega_n^j} x_n^j(k) dk \leq X_n,$$

where Ω_n^j is the set of varieties of product j available in country n, $x_n^j(k) = p_n^j(k) q_n^j(k)$, and $p_n^j(k)$ is the price of variety k of product j in country n, and $\gamma > \delta > 0$. Expenditure by country n on variety k of good j is given by

$$x_n^j(k) = \left(\frac{p_n^j(k)}{\hat{P}_n^j} \right)^{-\gamma} \hat{\beta}_n^j \left(\frac{\hat{P}_n^j}{\hat{P}_n} \right)^{-\delta} X_n,$$

where

$$\hat{P}_n^j = \left(\int_{\Omega_n^j} \left(p_n^j(k) \right)^{-\gamma} dk \right)^{-\frac{1}{\gamma}},$$

and

$$\hat{P}_n = \left(\sum_j \hat{\beta}_n^j (\hat{P}_n^j)^{-\delta} \right)^{-\frac{1}{\delta}}.$$

Each source country has a set of potential producers, each of which has a level of productivity, $z_i^j(k)$, drawn from the productivity distribution, $G_i^j(z)$.

The marginal cost of delivering variety k of product j to destination n from source i is

$$c_{ni}^j(k) = \frac{w_i t_{ni}}{z_i^j(k)},$$

where w_i is the cost of a bundle of inputs in i, $t_{ni} \geq 1$ is the "iceberg" trade cost of delivering goods from i to n. I normalize $t_{ii} = 1$ for all i. The fixed cost associated with delivering a positive amount of a product from i to n is $f_{ni} \geq 0$.

C.1 Ricardian Model

The Ricardian version of this framework requires the following additional restrictions: (1) there is a fixed unit measure of varieties of each product, $k \in [0, 1]$, and each country can produce every variety; (2) $f_{ni} = 0$; (3) there is perfect competition among producers of each variety; (4) the productivity distribution is type II extreme value (Fréchet) over $[0, \infty)$,

$$G_i^j(z) = e^{-T_i^j z^{-\theta}},$$

where $T_i^j > 0$ governs expected productivity for a given country and product category – i.e. comparative advantage across product categories – and $\theta \in (\gamma, \infty)$ governs the shape of the distribution – i.e. comparative advantage within product categories – with a smaller value of θ implying greater variance. This framework nests that of Eaton and Kortum (2002), differing only in that the shift parameter of the productivity distribution varies across products, as in Costinot et al. (2012) and French (2009).

Condition (2) implies that all varieties are available to all destination countries, and condition (3) implies that the price of each variety is equal to the cost of delivering it there, i.e. $p_{ni}^j(k) = c_{ni}^j(k)$. Each country will purchase only the variety of each product with the lowest price, so $p_n^j(k) = \min\{p_{ni}^j(k)\}_{i=1}^N$. Following Eaton and Kortum (2002), it is straightforward to show that p_n^j is distributed according to

$$F_n^j(p) = \Pr(p_n^j(k) < p) = 1 - e^{-\Phi_n^j p^\theta} \tag{28}$$

and the the probability that country i is the low-cost provider of a variety to n is

$$\pi_{ni}^j = \frac{T_i^j (w_i t_{ni})^{-\theta}}{\Phi_n^j}, \tag{29}$$

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

To show that this framework is consistent with (1) - (4), I define

$$\hat{p}_{ni}^j = \left(\int_{\Omega_{ni}^j} \left(p_{ni}^j(k) \right)^{-\gamma} dk \right)^{-\frac{1}{\gamma}}, \tag{30}$$

where $\Omega_{ni}^j \subset \Omega_n^j$ denotes the set of varieties of j that n purchases from i. First, note that

$$\sum_i \left(\hat{p}_{ni}^j \right)^{-\gamma} = \left(\hat{P}_n^j \right)^{-\gamma}.$$

Next, combining (30) with (28) and (29),

$$\begin{aligned} \left(\hat{p}_{ni}^j \right)^{-\gamma} &= \int_0^\infty p^{-\gamma} \pi_{ni}^j dF_n^j(p) \\ &= T_i^j (w_i t_{ni})^{-\theta} \int_0^\infty p^{\theta-\gamma-1} \theta 1 - e^{-\Phi_n^j p^\theta} dp \\ &= \hat{\gamma} T_i^j (w_i t_{ni})^{-\theta} \left(\Phi_n^j \right)^{\frac{\gamma-\theta}{\theta}}, \end{aligned}$$

where $\hat{\gamma}$ is a function of parameters.[21] This implies that $\left(\hat{P}_n^j \right)^{-\gamma} = \hat{\gamma} \left(\Phi_n^j \right)^{\frac{\gamma}{\theta}}$ and, therefore, that

$$\frac{X_{ni}^j}{X_n^j} = \left(\frac{\hat{p}_{ni}^j}{\hat{P}_n^j} \right)^{-\gamma} = \frac{T_i^j (w_i t_{ni})^{-\theta}}{\Phi_n^j}.$$

Similarly, summing $(\hat{P}_n^j)^{-\delta})$ over j, we have that $\hat{P}_n^{-\delta} = \hat{\gamma}^{\frac{\delta}{\theta}} \Phi_n^{\frac{\delta}{\theta}}$ and, thus, that

$$\frac{X_n^j}{X_n} = \hat{\beta}_n^j \left(\frac{\hat{P}_n^j}{\hat{P}_n} \right)^{-\delta} = \hat{\beta}_n^j \left(\frac{\Phi_n^j}{\Phi_n} \right)^{\frac{\delta}{\theta}}.$$

[21] $\hat{\gamma} = \Gamma\left(\frac{\theta-\gamma}{\gamma} \right)$, where $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$ is the gamma function. The result is obtained by substituting $\Phi_n^j p^\theta$ for t.

where $\Phi_n^{\frac{\delta}{\theta}} = \sum_j \hat{\beta}_n^j (\Phi_n^j)^{\frac{\delta}{\theta}}$.

Now, I define the following:

$$\begin{aligned}
p_i^j &\equiv (T_i^j)^{-\frac{1}{\theta}} w_i \\
P_n^j &\equiv \left(\Phi_n^j\right)^{-\frac{1}{\theta}} = \left(\sum_i (p_i^j t_{ni})^{-\theta}\right)^{-\frac{1}{\theta}} \\
P_n &\equiv \Phi_n^{-\frac{1}{\theta}} = \left(\sum_i \hat{\beta}_n^j \left(P_n^j\right)^{-\delta}\right)^{-\frac{1}{\delta}}.
\end{aligned} \tag{31}$$

It is, then, straightforward to show that expenditure by n on varieties of j from i is

$$X_{ni}^j = \left(\frac{p_i^j t_{ni}}{P_n^j}\right)^{-\theta} \hat{\beta}_n^j \left(\frac{P_n^j}{P_n}\right)^{-\delta} X_n.$$

Thus, letting $\sigma = \delta$, $d_{ni} = t_{ni}$, and $\beta_n^j = \hat{\beta}_n^j$, it is clear that the model is consistent with (1) - (4).

C.2 Monopolistic Competition

The monopolistic competition version of this framework requires the following additional restrictions: (1) each producer produces a unique variety of a product; (2) there is a measure, T_i^j, of producers in source country i that can produce a variety of product j; and (3) the productivity distribution is Pareto over $[1, \infty)$,

$$G_i^j(z) = 1 - z^{-\theta},$$

where $\theta \in (\gamma, \infty)$. If we set $\gamma = \delta$, this framework reduces to a one-sector monopolistic competition model, similar to that studied in Helpman et al. (2008). If we let $\delta \to 0$, then it becomes a multi-sector model in which expenditure shares across sectors are exogenously fixed (as in Chaney (2008)).

Condition 1 implies that each producer is a monopolist in the market for its unique variety. The CES demand structure, then, implies that each producer will charge a constant markup, $\bar{m} = \frac{1+\gamma}{\gamma}$, over the marginal cost of delivering its variety to country n, so

$$p_{ni}^j(k) = \bar{m}\frac{w_i t_{ni}}{z_i^j(k)}.$$

A producer of a variety of product j from i with a productivity of z selling in n will earn a profit, net of the fixed cost, of

$$\pi_{ni}^j(z) = \frac{1}{1+\gamma}\left(\frac{w_i t_{ni}}{z}\frac{\bar{m}}{\hat{P}_n^j}\right)^{-\gamma} X_n^j - f_{ni},$$

which is positive if z is greater than

$$\bar{z}_{ni}^j = \bar{m}\left(\frac{1}{1+\gamma}\right)^{-\frac{1}{\gamma}} \frac{w_i t_{ni}}{\hat{P}_n^j}\left(\frac{X_n^j}{f_{ni}}\right)^{-\frac{1}{\gamma}}.$$

Defining $\hat{p}_{ni}^j$ as in (30), and using conditions 2 and 3,

$$\begin{aligned}\left(\hat{p}_{ni}^j\right)^{-\gamma} &= T_i^j \int_{\bar{z}_{ni}^j}^{\infty}\left(\bar{m}\frac{w_i t_{ni}}{z}\right)^{-\gamma}\theta z^{-\theta-1}dz \\ &= \hat{\eta} T_i^j (w_i t_{ni})^{-\theta} f_{ni}^{-\frac{\theta-\gamma}{\gamma}} (X_n^j)^{\frac{\theta-\gamma}{\gamma}} (\hat{P}_n^j)^{\theta-\gamma},\end{aligned}$$

where $\hat{\eta}$ is a function of parameters.[22] Summing over i and rearranging gives that

$$\left(\hat{P}_n^j\right)^{-\gamma} = \hat{\eta}(X_n^j)^{\frac{\theta-\gamma}{\gamma}}(\hat{P}_n^j)^{\theta-\gamma}\Phi_n^j, \tag{32}$$

which implies that

$$\frac{X_{ni}^j}{X_n^j} = \left(\frac{\hat{p}_{ni}^j}{\hat{P}_n^j}\right)^{-\gamma} = \frac{T_i^j (w_i t_{ni})^{-\theta} f_{ni}^{-\frac{\theta-\gamma}{\gamma}}}{\Phi_n^j}.$$

where $\Phi_n^j = \sum_i T_i^j (w_i t_{ni})^{-\theta} f_{ni}^{-\frac{\theta-\gamma}{\gamma}}$. Recalling that $X_n^j = \hat{\beta}_n^j \left(\frac{\hat{P}_n^j}{\hat{P}_n}\right)^{-\delta} X_n$, rearranging (32) yields

$$\left(\hat{P}_n^j\right)^{-\delta} = (\hat{\beta}_n^j)^{\frac{\delta-\alpha}{\delta}}(\Phi_n^j)^{\frac{\alpha}{\theta}} X_n^{\frac{\delta-\alpha}{\delta}} (P_n)^{\delta-\alpha}$$

[22] $\hat{\eta} = \bar{m}^{-\theta}\frac{\theta}{\theta-\gamma}\left(\frac{1}{1+\gamma}\right)^{\frac{\theta-\gamma}{\gamma}}$

where $\frac{1}{\alpha} = \frac{1}{\delta} + \frac{1}{\theta} - \frac{1}{\gamma}$ and $\alpha \in (0, \delta)$. It then follows that $\hat{P}_n^{-\delta} = X_n^{\frac{\alpha-\delta}{\delta}} (P_n)^{\alpha-\delta} \Phi_n^{\frac{\alpha}{\theta}}$, and

$$\frac{X_{ni}^j}{X_n^j} = \hat{\beta}_n^j \left(\frac{\hat{P}_n^j}{\hat{P}_n} \right)^{-\delta} = (\hat{\beta}_n^j)^{2-\frac{\alpha}{\delta}} \left(\frac{\Phi_n^j}{\Phi_n} \right)^{\frac{\alpha}{\theta}},$$

where $\Phi_n^{\frac{\alpha}{\theta}} = \sum_j (\hat{\beta}_n^j)^{2-\frac{\alpha}{\delta}} (\Phi_n^j)^{\frac{\alpha}{\theta}}$.

Now, I define the following:

$$\begin{aligned}
\beta_n^j &\equiv (\hat{\beta}_n^j)^{2-\frac{\alpha}{\delta}} \\
p_i^j &\equiv (T_i^j)^{-\frac{1}{\theta}} w_i \\
P_n^j &\equiv \left(\Phi_n^j \right)^{-\frac{1}{\theta}} = \left(\sum_i (p_i^j t_{ni} f_{ni}^{\frac{1}{\theta}-\frac{1}{\gamma}})^{-\theta} \right)^{-\frac{1}{\theta}} \\
P_n &\equiv \Phi_n^{-\frac{1}{\theta}} = \left(\sum_i \beta_n^j \left(P_n^j \right)^{-\alpha} \right)^{-\frac{1}{\alpha}}.
\end{aligned}$$

It is, then, straightforward to show that expenditure by n on varieties of j from i is

$$X_{ni}^j = \left(\frac{p_i^j t_{ni} f_{ni}^{\frac{1}{\theta}-\frac{1}{\gamma}}}{P_n^j} \right)^{-\theta} \beta_n^j \left(\frac{P_n^j}{P_n} \right)^{-\alpha} X_n.$$

Thus, letting $\sigma = \alpha$ and $d_{ni} = t_{ni} f_{ni}^{\frac{1}{\theta}-\frac{1}{\gamma}}$, it is clear that the model is consistent with the system (1) - (4).