Heterogeneity and the Distance Puzzle*

Elizaveta Archanskaia† Guillaume Daudin‡

January 2012

Preliminary version - Comments welcome

Abstract

This paper shows that reduced heterogeneity of exporter-specific goods can provide a direct explanation of the distance puzzle. Using COMTRADE 4-digit bilateral trade data we find that the distance coefficient has increased by 8% from 1962 to 2009. Theoretical foundations of the gravity equation indicate that the distance coefficient is the product of the elasticity of trade costs to distance and a measure of heterogeneity, e.g. the substitution elasticity between exporter-specific goods in the Armington framework. This parameter has increased by at least 29% from 1962 to 2009. The evolution of the distance coefficient is thus compatible with a 19% *reduction* in the elasticity of trade costs to distance.

Keywords: gravity equation, distance puzzle, trade elasticity, trade costs

JEL codes: F15, N70

*This paper has greatly benefited from helpful discussions with Lorenzo Caliendo, Thomas Chaney, Lionel Fontagné, Samuel Kortum, Jacques Le Cacheux, Fernando Parro, and suggestions by seminar participants at SciencesPo, ADRES, AFSE, ETSG, the International Trade Working Group at the University of Chicago, and the international workshop on the Economics of Global Interactions (Bari).

†SciencesPo/OFCE

‡Corrsponding author. Lille-I/EQUIPPE and Sciences Po/OFCE. Email:gdaudin@mac.com

I Introduction

The estimated effect of distance in gravity equations has remained stable, or even slightly increased, in the past 50 years.[1] At first approach, this may seem like a puzzle as the common opinion is that technological developments in transports, e.g. the airplane and the container, had contributed to the "death of distance".[2]

Recent work on the distance puzzle has gone in two directions. Following Santos Silva and Tenreyro, one strand of the literature has sought to correct the estimation method using non linear estimators.[3] The canonical log-linear estimation does not generally provide consistent coefficient estimates if the trade equation in levels is subject to heteroskedasticity. Furthermore, the log-linear estimation strategy suffers from sample selection bias because it cannot take into account nil trade flows. The growth of trade has been both intensive, in the sense that the volume of trade of established trade relations has increased, and extensive, in the sense that new trade relations have been established.[4] Taking zeros into account might change the evolution of the distance coefficient if new trade relations have in priority been established between distant partners.

Using a non-linear estimator, Bosquet and Boulhol (2009) find that the distance elasticity of trade has been stable within the .6-.75 range between 1948 and 2006. Coe et al. (2007) find that it has decreased from 0.5 to 0.3 in 1975-2000.

The other strand of the literature has argued that there is no puzzle, as there are theoretical and empirical explanations of a non decreasing distance elasticity of trade.[5]

Firstly, a composition effect might explain a non decreasing distance coefficient. It might be that the composition of traded goods has changed toward goods which are either less transportable or which consumption is more sensible to trade costs. Theoretically, Duranton and Storper (2008) show how falling transport costs can induce firms into trading goods with higher transaction costs, leading to an increasing distance sensitivity of trade. Empirically, Berthelon and Freund (2008) test the impact of the changing composition of world trade between 1985

1 Disdier and Head (2008) find that distance impedes trade by 37% more since 1990 than it did from 1870 to 1969 in a meta-analysis based on 1467 estimations of the distance coefficient in gravity equations spanning the period 1870-2001. Berthelon and Freund (2008) find that the distance coefficient has increased by 10% in absolute value over the period 1985-2005. See also Bosquet and Boulhol (2009); Coe et al. (2007); Combes et al. (2006); Brun et al. (2005); Buch et al. (2004).

2 See Cairncross (1997); Levinson (2006); Friedman (2007).

3 Santos-Silva and Tenreyro (2006); Coe et al. (2007); Bosquet and Boulhol (2009). See §II.2 for a discussion.

4 Helpman et al. (2008); Baldwin and Harrigan (2007).

5 Bosquet and Boulhol (2009); Duranton and Storper (2008); Buch et al. (2004).

and 2005 on the distance coefficient, but find that it has had a negligible effect.[6] Using a log-linear estimation strategy, Brun et al. (2005) and Márquez-Ramos et al. (2007) find decreasing distance elasticities once the sample is restricted to developed countries. This suggests another explanation: the increasing importance of certain components of trade costs, such as trading on time, for most traded goods.

Secondly, it is not certain that transport costs have declined relative to the price of traded goods.[7] However, it might be the case that other distance-related components of trade costs, such as delays, have a growing importance.[8]

Thirdly, the value of the distance coefficient in theoretically derived gravity models of trade is not directly related to the level of trade costs. The level of trade costs influences the openness of each country, and this is captured by country-specific fixed effects. The coefficient is the product of two elasticities: the elasticity of trade flows to distance-related trade costs, and the elasticity of trade costs to distance. Whether transport costs have declined or increased, the crucial question for the evolution of the coefficient is whether non-distance dependent transport costs, such as loading costs at ports, have declined relatively to distance-dependent transport costs, such as fuel costs.[9]

The discussion on the distance puzzle in this second strand of the literature has thus far been mainly concerned with the evolution of the elasticity of trade costs to distance. This is forgetting half of the picture. One should look as well at the elasticity of trade flows to trade costs. Following Arkolakis et al. (2009), we refer to this parameter as the 'trade elasticity'. The structural interpretation of the trade elasticity depends on the micro foundations used to derive the gravity equation: in the canonical models this paper builds on, it corresponds alternatively to the dispersion in productive efficiency across sectors, to Armington elasticities of substitution between country-composite goods, or to the intrasectoral dispersion in firm productivity. In all cases, the trade elasticity is inversely linked to the measure of heterogeneity in some dimension.

For example, since the 1960s, a number of countries have demonstrated their ability to produce a set of goods bearing the same brands as the goods produced in developed countries. This may have resulted in more uniformity in country-specific varieties. In the Armington framework this could have resulted in an increased elasticity of substitution between country-specific

[6] Berthelon and Freund find that 39% of products and 54% of trade has become more sensitive to distance over this period against 2.8% of trade becoming less sensitive to distance.
[7] See Hummels (1999); Daudin (2003, 2005); Hummels (2007).
[8] Hummels (2001).
[9] Feyrer (2009).

composite goods, and hence increased elasticity of trade flows to trade costs, providing a direct explanation of the distance puzzle.[10] In the Ricardian framework, it could be argued that countries' technological capabilities have become more homogeneous across sectors overtime which would result in reducing the strength of comparative advantage in determining trade flows, leading to an increased sensitivity of trade flows to trade costs.[11] In the framework of monopolistic competition with heterogeneous firms increased sensitivity of trade flows to trade costs would correspond to reduced efficiency dispersion of firms within any given sector, e.g. a relatively bigger share of firms situated near the export threshold.[12]

This paper's original contribution is to assess the role of the evolution of the *trade elasticity* in the distance puzzle. This paper finds that the parameter which corresponds to the aggregate trade elasticity in the framework of the Armington model has increased by at least 29% from 1962 to 2009.[13] Two conclusions follow. *First*, the non-decreasing distance elasticity of trade is fully explained by the 29% increase in the elasticity of trade flows to trade costs. In the Armington model, this corresponds to increased substitutability of countries' export bundles. *Second*, the distance puzzle ceases to exist insofar as the elasticity of trade costs to distance is found to have decreased by at least 19% in 1962-2009. Future work will strive to estimate the evolution of the trade elasticity parameter in the Ricardian framework to check whether this explanation of the distance puzzle is robust to alternative micro-foundations of the gravity equation.[14]

II The distance puzzle in our data

As shown by Disdier and Head (2008), estimates of the distance elasticity are sensitive to the set of data used. Previous studies have quantified the impact of subsets of possible explanatory factors using two different datasets. Berthelon and Freund (2008) work with a COMTRADE 4-digit dataset for 1985-2005, averaging the data over 5 year intervals, to study the composition effect. They do not consider sample or FTAs effects. Bosquet and Boulhol (2009) work with aggregate trade data from DOTS for 1945-2006 and annual estimates to study the impact of the

[10] The elasticity of substitution between composite country-specific goods could play a role in explaining the distance puzzle according to the Anderson and van Wincoop (2003) derivation of the gravity model which uses the demand framework pioneered by Armington (1969). See §III.3 for a discussion.

[11] See Eaton and Kortum (2002). See §III.1 for a discussion.

[12] See Chaney (2008). See §III.2 for a discussion.

[13] See Appendix A and B for the explanation of why the data we use do not allow estimating the trade elasticity parameter in the Ricardian or the Melitz-Chaney frameworks.

[14] Appendix B explains why estimating the evolution of the aggregate trade elasticity in the Ricardian framework is sufficient to caracterize the evolution of the trade elasticity in the Melitz-Chaney framework.

estimation strategy and of the selection into FTAs, but they lump all FTAs into a single variable, and do not consider the dissolution of FTAs due to emergence of new countries.

This section evaluates the ability of this traditional approach to explain the distance puzzle on a single set of data. We quantify individual and combined effect of factors identified by previous studies as explanatory of movements in the distance coefficient. These factors are:[15]

- the *composition effect* defined as the evolution in the goods' composition of world trade in 1962-2009;

- the *sample effect* which covers two types of developments: the formation of new trading relationships between previously existing countries; and the formation or disappearance of countries through dissolution of previously existing political entities;

- the *selection into FTAs effect* which corresponds to the formation of new FTAs between previously existing countries.

1 Data

We use the COMTRADE world bilateral trade dataset spanning the years 1962-2009[16]. We use the 4-digit SITC Rev.1 classification of goods (600-700 goods), as this provides the longest coverage of bilateral trade flows, restricting the sample to trade in goods which are attributed to specific 4-digit categories. Our sample covers between 88% and 99% of reported trade in COMTRADE. However, it covers only 70% of world merchandise trade according to the WTO in 1962.[17] This increases to 80% in the 1960s, stays stable until the late 1980s, and then increases to approximately 90% in the 1990s and 2000s.

As trade data is of better quality for imports, the estimation is conducted only on import flows. Data on bilateral distance, bilateral trade costs' controls such as adjacency, common language, colonial linkages, and data on belonging or having once belonged to the same country is taken from the CEPII.[18] The database on GATT/WTO and FTA membership was constructed

[15] Results on the impact of the estimation strategy are not reported because they are qualitatively similar to Bosquet and Boulhol (2009).

[16] The 228 trading partners present in this dataset are listed in App.G

[17] See http://stat.wto.org/Home/WSDBHome.aspx, accessed in May 2011.

[18] This database, constructed by Mayer and Zignago, is available online at www.cepii.fr. For East and West Germany, USSR, and Czechoslovakia, absent from the database, we constructed bilateral distance and bilateral costs' controls.

by the authors.[19]

2 Methodology

We follow the main theoretical derivations of the gravity trade model, e.g. Anderson and van Wincoop, Eaton and Kortum, and Chaney, noting that they result in a qualitatively similar estimation equation for aggregate bilateral trade flows.[20]

In Anderson and van Wincoop (1), bilateral exports X_{ij} are a function of the nominal income of each trading partner $Y_{i,j}$, of world income Y, of bilateral trade costs τ_{ij}, and of multilateral trade resistance terms Π_i, P_j.[21] The parameter σ is the Armington elasticity of substitution which measures the degree of substitutability of goods of different national origin.[22]

$$X_{ij} = \left(\frac{Y_i Y_j}{Y}\right)\left(\frac{\tau_{ij}}{\Pi_i P_j}\right)^{-(\sigma-1)} \tag{1}$$

In Eaton and Kortum (2), bilateral exports are a function of total sales of the exporter i and of total expenditure of the importer j, as well as of the total world market as perceived by exporter i. Bilateral exports are also a function of bilateral trade barriers of exporter i with importer j, and of the price index of the importer. The parameter θ captures the degree of heterogeneity in productive efficiency within countries across goods, e.g. the strength of comparative advantage.[23]

$$X_{ij} = \frac{\left(\frac{\tau_{ij}}{p_j}\right)^{-\theta} Y_i Y_j}{\sum_{n=1}^{N}\left(\frac{\tau_{in}}{p_n}\right)^{-\theta} Y_n} \tag{2}$$

In Chaney (3), bilateral exports are a function of the share of expenditure μ on goods which trade is costly,[24] the nominal income of trading partners, world income, fixed and variable bilateral trade costs (f_{ij} and τ_{ij}, respectively), workers' productivity in the exporting country

[19] Data for membership of GATT/WTO was taken from the WTO website http://www.wto.org/english/thewto_e/gattmem_e.htm. Data for common membership of an FTA was constructed on the basis of Crawford and Fiorentino (2005), Fontagné and Zignago (2007), and the RTA Information System on the WTO website http://rtais.wto.org/UI/PublicMaintainRTAHome.aspx. See App. F for the list of included FTAs and the years they appear in our data.

[20] See footnote 20 in Eaton and Kortum (2002) for a discussion of the equivalence of the resulting equations in the Ricardian framework, in Armington models such as Anderson and van Wincoop (2003), and in monopolistic competition models such as Krugman (1980), extended by Chaney (2008) to the setting of heterogeneous firms.

[21] Π_i, P_j are respectively inward and outward MR terms.

[22] See Anderson and van Wincoop (2003). Appendix C recalls the distinctive features of this model.

[23] See Eaton and Kortum (2002). Appendix A recalls the distinctive features of this model.

[24] The remaining expenditure is allocated to the numeraire sector producing a homogeneous freely tradable good.

w_i, and a measure of importer's remoteness θ_j.[25] The parameter γ is the shape parameter of the Pareto distribution. It measures the degree of firm productivity heterogeneity in the costly-trade sector.[26] The degree of substitutability between firm-level varieties is measured by σ_f.

$$X_{ij} = \mu \frac{Y_i Y_j}{Y} \left(\frac{w_i \tau_{ij}}{\theta_j} \right)^{-\gamma} f_{ij}^{-\left(\frac{\gamma}{\sigma_f - 1} - 1\right)} \tag{3}$$

Working with bilateral imports data, assuming that fixed bilateral trade costs are not a function of distance while certain components of variable trade costs can be modelled as a function of distance: $\tau_{ij} = dist_{ij}^{\rho}$, the three frameworks result in (4) to be estimated in cross section:

$$X_{ij} = \exp\left(\alpha_0 - \alpha_1 \ln dist_{ij} + \beta_1 Z_1 + \beta_2 Z_2 + fe_{exp} + fe_{emp}\right) \varepsilon_{ij} \tag{4}$$

X_{ij} is the value of goods from country i consumed in country j, e.g. imports at cif prices; $dist_{ij}$ is bilateral distance, Z_1 comprises bilateral trade costs' controls such as adjacency, common language, colonial linkages, and data on belonging or having once belonged to the same country; Z_2 comprises bilateral trade cost controls linked to trade policy such as common membership of GATT/WTO and common membership of an FTA; fe_{exp} and fe_{imp} are respectively exporter and importer fixed effects, and ε_{ij} is a multiplicative error term.

The coefficient of interest α_1 has a different structural interpretation in each framework: $\alpha_1^{ARMINGTON} = \rho(\sigma - 1)$; $\alpha_1^{RICARDO} = \rho\theta$; $\alpha_1^{HET-FIRMS} = \rho\gamma$; $\alpha_1^{HOMOG-FIRMS} = \rho\left(\sigma_f - 1\right)$. If instead certain components of fixed costs are assumed to be a function of distance, the distance coefficient is no longer a product of two elasticities. Specifically, Chaney (2008) allows certain components of fixed costs to be distance related. Assuming $\tau_{ij} = dist_{ij}^{\rho_1}$ and $f_{ij} = dist_{ij}^{\rho_2}$, gives: $\alpha_1^{HET-FIRMS-FULL} = \rho_1\gamma + \rho_2\left(\frac{\gamma}{\sigma_f - 1} - 1\right)$.

The estimation procedure can accomodate fixed costs being a function of distance. However, in subsequent parts of the paper, we write the distance coefficient as a product of two elasticities, e.g. we assume that $\rho_2 = 0$, leaving $\alpha_1^{HET-FIRMS-FULL} = \rho_1\gamma$. This is more flexible than assuming that fixed trade costs are either source- or destination-specific, and thus accounted for by fixed effects. Fixed trade costs could be bilateral but would then be accounted for by bilateral proximity controls such as common language. The assumption is then that fixed trade costs are

25 This remoteness measure is reminiscent of MR terms in Anderson and van Wincoop (2003), but it also accounts for firm heterogeneity in productivity and fixed bilateral trade costs. See Chaney (2008). Appendix B goes over the main features of this model.

26 Productivity is distributed over $[1, +\infty)$ as follows: $P\left(\widetilde{\phi}_k < \phi\right) = \mathcal{G}_k(\phi) = 1 - \phi^{-\gamma_k}$ where ϕ is unit labour productivity. Firm-specific productivity determines the marginal cost of production ($c_i = w_i/\phi$) where the wage w_i is pinned down by country-specific labor productivity in the numeraire sector [see Appendix B].

not a function of distance. For consistency, we make this hypothesis throughout the paper.

Equation (4) could be estimated using a log-linear specification. But Santos Silva and Tenreyro have shown that slope estimates from the log linear estimation will generally be biased since the trade equation in levels is subject to heteroskedasticity. This is because the expected value of the logarithm of a random variable depends on the higher moments of the distribution. Thus, if the variance of the multiplicative error term depends on the regressors, the expected value of the logarithm of the error term in the log linear specification will also depend on the regressors, violating the condition for consistency of OLS. The magnitude of the bias depends on the structure of the variance of the residuals, which is unknown. The point made by Santos Silva and Tenreyro is that the structure of trade data is such that the additive error term in the log-linearized model will in general be heteroscedastic, and its variance will in general depend on the regressors, but there is no argument to justify the assumption that the conditional variance of raw trade flows follows a Gamma distribution, i.e. is proportional to the square of the conditional mean, which is the only case in which OLS would give consistent slope estimates.[27]

Santos Silva and Tenreyro advocate the use of the Poisson pseudo maximum likelihood (PPML) estimator arguing that the mean and variance function this estimator assumes is closest to the underlying conditional variance function in raw trade data.[28] Specifically, given that the conditional variance of trade flows increases with the conditional mean, the PPML which assumes that the conditional variance is proportional to the conditional mean is likely to be more efficient than the NLS estimator which assumes that the conditional variance is independent of the regressors. The Gamma PML estimator is a valid alternative to PPML, but contrary to PPML, it is found to be sensitive to certain forms of measurement error such as the rounding of the dependent variable. Arguably, this type of measurement error is present in trade data.[29]

These non-linear strategies have the additional advantage of allowing the inclusion in the data of zero trade observations.[30] However, their implied functional forms do not take into account the fact that zero trade observations are an important share of all trade observations. One way of explicitly accounting for the prevalence of zeros in the data is to couple the PPML

[27] See Manning and Mullahy (1999); Santos-Silva and Tenreyro (2006). OLS gives consistent coefficient estimates in the case of $V[y_i|x] = \mu(x_i\beta)^2$.

[28] The general structure of the conditional variance function in the generalized linear model is: $Var[y_i|x] = \lambda_0\mu(x_i\beta)^{\lambda_1}$. The NLS corresponds to $\lambda_1 = 0$, the PPML to $\lambda_1 = 1$, and GPML to $\lambda_1 = 2$.

[29] Santos-Silva and Tenreyro (2006); Bosquet and Boulhol (2009) get qualitatively similar results in terms of GPML sensitivity to measurement error and of NLS inefficiency.

[30] Felbermayr and Kohler (2006) argue that distance coefficient estimates in the 1960s were biased downwards in absolute value because of a greater presence of zeros, and that this bias was subsequently reduced thanks to the extensive margin of trade.

estimator with a logit estimation of the chances of being part of the zero category by doing a zero-inflated Poisson regression (ZIP).[31]

In terms of the underlying model, it is clear that the three canonical frameworks presented above do not allow for zero trade observations. However, Helpman et al. have shown that placing an upper bound on the Pareto distribution of productivity allows generalizing the Anderson and van Wincoop framework to get both zero trade and asymmetric trade flows, with a gravity equation augmented by a term which accounts for the fraction of exporting firms.[32] Thus, while amending the model to generate zero trade flows underscores that the standard estimation procedure of the gravity equation suffers not only from selection but also from asymmetry bias which stems from not adequately controlling for the pair-specific fraction of exporting firms, it strengthens the case of using a two-part model such as the ZIP in case of uncertainty over the selection variable.[33]

Given uncertainty over the selection variable, we opt for the two-part model because it improves the fit to the data while being consistent with the augmented gravity equation.[34] This model estimates the effects of the explanatory variable on both the extensive margin (through the logit model) and the intensive margin (through the Poisson model restricted to predicted non-zeros). Both models use the same set of explanatory variables. This procedure has the additional advantage of yielding the effect of distance on both the extensive and the intensive margins.

The baseline regression includes bilateral distance, bilateral costs' controls, and country fixed effects.[35] We find that α_1, the absolute value of the distance coefficient, has increased over 1962-2009 (see Fig.1). There is indeed a *distance puzzle*.

The estimation procedure allows circumscribing the distance puzzle. We find that the coefficient of distance in the logit regression has decreased, e.g. the formation of a trading re-

[31] See Greene (2003), p. 750. The name comes from Lambert (1992). It is also named 'With zeros' model by Mullahy (1986) or 'Zero-Altered Poisson' by Greene (1994). This approach is the one advocated in Burger et al. (2009), and is similar to the one advocated in Helpman et al. (2008). However, not everybody believes that using the ZIP procedure on trade data is valid: see the page of J.M.C. Santos Silva at http://privatewww.essex.ac.uk/~jmcss/LGW.html.

[32] See Helpman et al. (2008). Eaton et al. (2011) have shown that considering a finite number of firms generates zero aggregate trade flows in the Eaton and Kortum framework.

[33] See Martin and Pham (2008). Mullahy (1986) shows that two-stage and two-part models give qualitatively similar results when there is clarity on the first-stage selection variable. Further, it is shown that the two-part approach performs better when there is uncertainty on the choice of the selection variable than the two-stage approach.

[34] The Vuong test indicates that the number of zero trade observations is excessive and that a ZIP specification is preferable to a PPML regression.

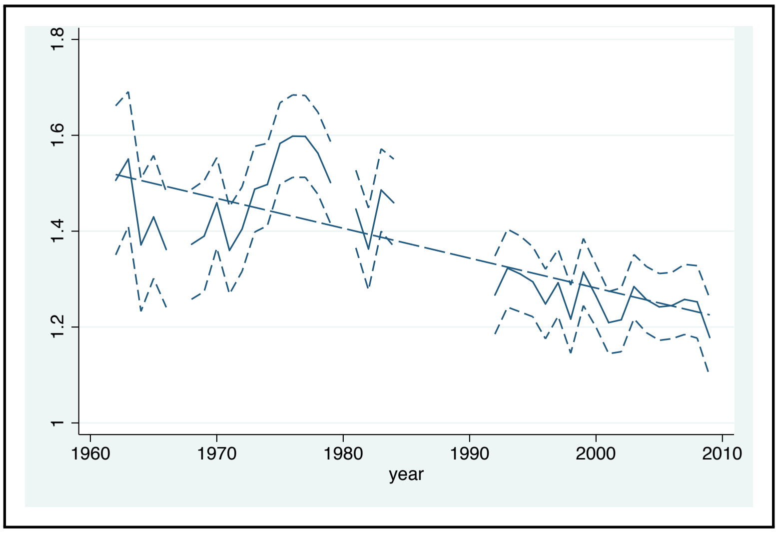
[35] $M_{ij} = \exp(\alpha_0 - \alpha_1 \ln dist_{ij} + \beta_1 Z_1 + fe_{exp} + fe_{emp})\varepsilon_{ij}$.

Figure 1: Distance coefficient in the baseline PPML regression

lationship is less sensitive to distance (see Fig. 2), while the second stage of the ZIP regression gives qualitatively similar results to the baseline PPML regression. Thus, the distance puzzle is specific to the intensive margin: conditional on trading, *trade volumes* have become more sensitive to distance in 1962-2009.

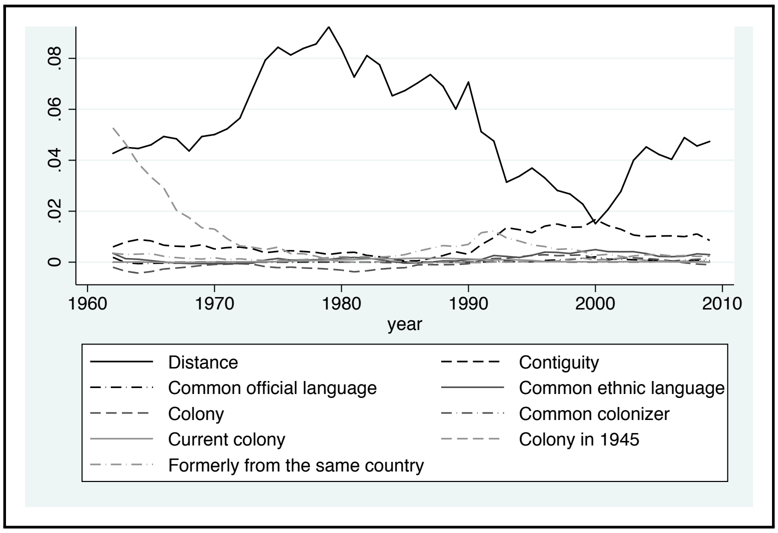
The amount of trade variance explained by distance does not have any trend over the whole

Figure 2: Distance coefficient in the logit baseline ZIP regression



period. The partial coefficient of determination, or the marginal contribution of distance to the coefficient of determination, has steadily declined from the late 1970s to 2000, yet it is in the late 2000s at the same level as in the 1960s (see Fig. 3).

Figure 3: Coefficient of partial determination in the PPML baseline equation



3 Robustness of the distance puzzle

To verify the robustness of our result on the distance puzzle, we estimate three variants of the baseline regression. First, entry and exit of countries in the sample could drive the increasing sensitivity of trade flows to distance. To test this, we restrict the sample to trading partners which have non-zero trade both ways in every year over 1962-2009. This leaves 32 countries corresponding to 786 stable two-way pairs.[36] Fig.4 shows that the sample effect does not reduce the distance puzzle.

Second, the changing composition of trade in terms of goods could explain the distance puzzle. To compute the impact of the composition effect, trade shares of each good are fixed to 1962 shares of the good in total trade. As shown by Fig.5, the composition effect deepens the distance puzzle.[37]

Third, we check whether controlling for countries' selection into FTAs and GATT/WTO membership reduces the distance puzzle. Separate controls are included for each FTA as notified to the WTO.[38]

[36] These are: Argentina, Belgium-Luxembourg, Brazil, Canada, Chile, Colombia, Denmark, France, Germany, Great Britain, Greece, Hong Kong, Iceland, Israel, Italy, Japan, Malaysia, Mexico, Netherlands, Philippines, Portugal, Paraguay, Singapor, South Korea, Spain, Sweden, Switzerland, Thailand, Tunisia, Turkey, Venezuela, USA. China, India, Sub-Saharan African countries do not have a reciprocal trading relationship in all years with any single country, they are excluded. Similarly, Central and Eastern European countries are excluded.

[37] Robustness checks have been conducted using 1984 and 2009 weights. Results are qualitatively similar.

[38] See Appendix F for the full list of included FTAs. Restricting this list to FTAs which have led to effective across the board reductions in bilateral trade costs leads to qualitatively similar results. This limited FTA list includes the European Community (followed by the European Union), the USA-Canada FTA (followed by NAFTA), the Comecon, EFTA, ASEAN, Mercosur, as well as GATT-WTO membership.

Figure 4: Distance coefficient in sample of stable pairs (PPML)

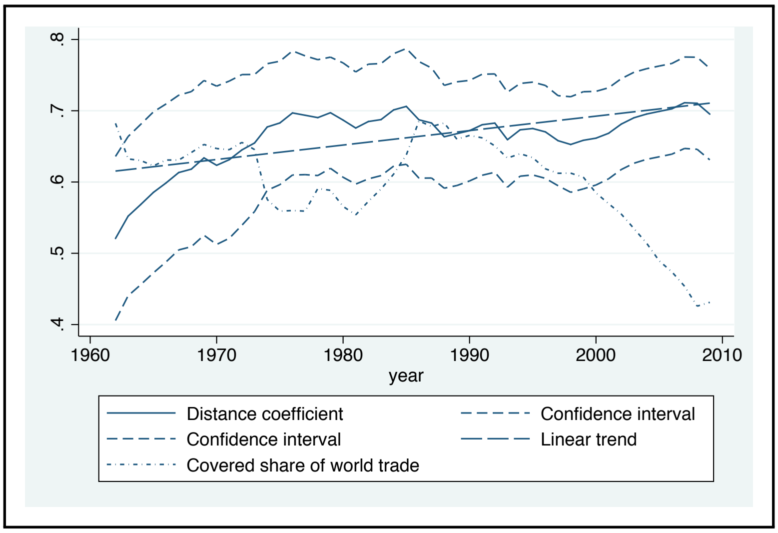
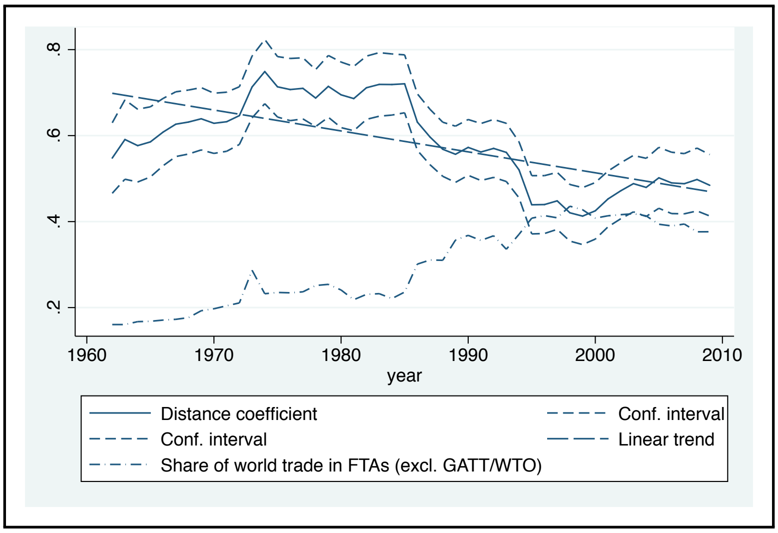


Figure 5: Distance coefficient controlling for composition (1962 weights)

As shown in Fig. 6, controlling for countries' participation into free trade agreements *solves* the distance puzzle. However, this effect is mechanical: the distribution of trade over distance is largely stable in 1962-2009 while over the same period, an increasing share of nearby countries start participating in FTAs. This is shown in Fig. 7. In-FTA trade has more than doubled as a share of world trade, increasing from less than 20% in the early 1960s to approximately 40% in the 2000s. Over the same period the share of intra-FTA trade among nearby (less than 2000 km) countries trade has increased from 38% in 1962 to 79% in 2009.[39] The inclusion of

[39] Trade at less than 2000 km corresponds to the first decile of the distance distribution of trade in 1962-2009.

Figure 6: Distance coefficient controlling for FTAs (full sample)



controls for FTA membership thus amounts to adding a growing number of proximity controls in the regression. This reduces the sensitivity of trade flows to distance. A further argument for caution in interpreting results with FTA controls is the probable endogeneity of FTAs for which we do not correct in our estimation. In this respect, Bosquet and Boulhol (2009) show that the distance coefficient evolves in the same way when they do not control for FTAs (fig. 9 in their paper) and when then control for FTAs correcting for endogeneity (fig. 11 in their paper).

Figure 7: Share of intra-FTA trade among nearby countries (2000km or less)

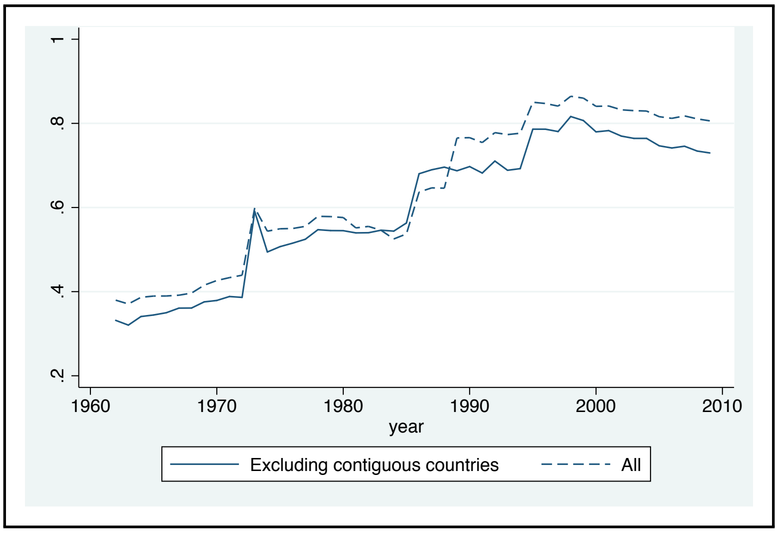


Table 1 summarizes our findings on the distance puzzle. It also explores the effect of combining sample, composition, and FTAs effects. Controlling for sample, composition, or the combined sample-composition effects does not change the magnitude of the distance puzzle whereas controlling for FTAs does. As argued, the explanation of the distance puzzle through FTAs is not satisfactory because this amounts to adding a growing number of proximity controls in the regression and because FTAs are endogenous to trade intensity. We conclude that the distance puzzle is robust, and that the traditional approach is little informative of the driving forces behind the increased sensitivity of trade flows to distance. The main contribution of this paper is the alternative approach to identifying the determinants of increased sensitivity of trade flows to distance to which we turn next.

Table 1: Evolution of the distance coefficient depending on sample, composition, and FTAs effects

	% change relatively to baseline		**Tot.change coef. 62-09 (geom. trend)**	
	PPML	ZIP	PPML	ZIP
Baseline			1.07	1.08
Sample effect	7%	5%	1.14	1.14
Reweight effect	7%	6%	1.14	1.15
FTA effect	-54%	-52%	0.49	0.52
Reweight sample effect	7%	5%	1.14	1.14
Reweight fta effect	-29%	-29%	0.75	0.77
Sample fta effect	-59%	-59%	0.44	0.44
Sample, reweight, FTA	-54%	-55%	0.49	0.49

III Interpreting the distance coefficient

The distance coefficient, or the elasticity of trade to distance, is the product of two elasticities: the *elasticity of trade costs to distance* and the *elasticity of trade flows to trade costs*. The three main microfoundations of the gravity model of trade give structurally different interpretations to the elasticity of trade flows to distance.

1 The Ricardian framework

1.1 The model

In the Eaton and Kortum framework, varieties of a given good are perfect substitutes in demand, and there is perfect competition in supply. Factors of production can be shifted across the set of goods without impacting factor prices. Consumers maximize a CES utility function defined at the inter-sectoral level. Consumers thus shop around the world for the cheapest supplier of each sectoral good.

A combination of ingredients determines the cheapest supplier. Two of these are country-specific: the fundamental cost component c_i which corresponds to the cost of the bundle of inputs in the exporting country, and the level of the overall technological ability of the exporting country T_i which corresponds to its accumulated stock of know-how. Another ingredient is country-pair specific: the bilateral multiplicative trade costs τ_{ij}. The final ingredient is the source of intrinsic heterogeneity in the model, and it is due to the assumption that countries are not equally efficient in production across the continuum of goods. Countries draw sectoral efficiency parameters from a Fréchet distribution common to all countries, and the Fréchet distribution parameter θ captures the degree of dispersion in productive efficiency across sectors within countries.[40] By the law of large numbers, the fraction of expenditure of country j on goods from country i (X_{ij}/X_j) is equal to the probability π_{ij} that i's cif price in country j is lowest among the set of all possible suppliers to j. This probability is given by (5):

$$\pi_{ij} = \Pr\left[p_{ij}(k) \leq \min\left\{p_{sj}(k)\right\} s \neq i\right] = \int_0^{\infty} \prod_{s \neq i} \left[1 - G_{sp}(p)\right] dG_{ij}(p) = \frac{T_i(c_i\tau_{ij})^{-\theta}}{\Phi_j} \quad (5)$$

Where $\Phi_j = \sum_{n=1}^{N} T_n(c_n\tau_{ij})^{-\theta}$ is the overall price distribution parameter in country j, $[1 - G_{sp}(p)]$ is the probability that the cif price of exporter s to j is higher than p, and $G_{ij}(p)$ is the probability that the cif price in i to j is lower than p.[41]

From (5), the elasticity of trade flows to bilateral trade costs is given by θ. The intuition is that this parameter embodies the strength that comparative advantage exerts on the direction and volume of trade flows. A higher value of this parameter generates less heterogeneity in domestic producers' efficiency across the set of goods. Comparative advantage exerts a weaker force against trade resistance imposed by trade barriers: the elasticity of trade to trade costs is higher in absolute value.

1.2 What has been the evolution of θ through time?

In the original Eaton and Kortum framework, just as in the multisector version of the model by Costinot et al., θ is invariant by assumption.[42] But in the data, θ could have evolved either way since the 1960s. Many source countries evolved towards producing a larger set of goods

[40] A lower θ corresponds to a higher degree of dispersion in productive efficiency across sectors.

[41] See Eaton and Kortum (2002, 2010) and Appendix A.

[42] Within each sector, efficiency draws are taken from a Pareto distribution with parameter θ, common to all sectors. The intersectoral efficiency distribution which keeps the best efficiency draw in each sector results Fréchet with parameter θ. See Eaton and Kortum (2002); Costinot et al. (2010) and Appendix A for details.

for a greater number of destinations. It could be argued that the decrease in trade barriers led to an expansion in the set of traded goods which could be mirrored by a shift in the *observed* variability of productive efficiency in exports.[43] If observed variability were to increase, the aggregate trade elasticity would decrease, deepening the distance puzzle. However, it could also be argued that the last 50 years had been characterized by *intersectoral convergence* in productive efficiency which would have weakened the strength of comparative advantage leading to increased sensitivity of trade flows to trade barriers. This would correspond to an *increasing* θ and reduce the distance puzzle.

Dynamic Ricardian comparative advantage models predict an increasing strength of comparative advantage through trade - i.e. a decreasing θ - only in the initial models of learning-by-doing.[44] More complex modelling of innovation and production in the recent literature has allowed formulating more nuanced predictions in terms of countries' specialization and the welfare consequences of trade, but has not generated any clearcut predictions as to the evolution of the strength of comparative advantage overtime.[45] It is an empirical question to check whether θ has evolved overtime, and how.

Three theoretically grounded methodologies have been used to estimate aggregate trade elasticities in the Ricardian framework.[46] To the best of our knowledge, the only paper to provide evidence on the *evolution* of intersectoral productivity dispersion in 1960-2010 using a Ricardian framework is Levchenko and Zhang.[47] They find evidence of within-country convergence in sectoral knowledge stocks, e.g. reduced dispersion in intersectoral productivity, even though there is great variability in the evolution of the strength of comparative advantage across countries. Levchenko and Zhang assume an invariant θ, but their results would be consistent with an *increasing* θ in 1960-2010. This would contribute to explaining the distance puzzle.

2 Distance sensitivity of trade in the Melitz-Chaney framework

2.1 The model

Monopolistic competition with representative firms yields determinants of bilateral exports very similar to those in the Armington framework, except that the Armington elasticity of substitu-

[43] The argument is similar in spirit to Felbermayr and Kohler (2006) who argue that the bias in the estimate of the distance coefficient could have been reduced overtime through the extensive margin of trade.
[44] See Krugman (1987); Lucas (1988); Boldrin and Scheinkman (1988); Young (1991); Murat and Pigliaru (1998).
[45] Davis (1995); Eaton and Kortum (2006); Rodríguez-Clare (2007, 2010).
[46] Eaton and Kortum (2002); Simonovska and Waugh (2011); Costinot et al. (2010); Caliendo and Parro (2011).
[47] Levchenko and Zhang (2011).

tion is replaced by the substitution elasticity between firm varieties as the framework shifts the heterogeneity dimension of the model to the level of the firm.[48] Hence there is no need to discuss it further.

Monopolistic competition with heterogeneous firms modifies this set-up in two ways. First, it is assumed that firms draw their efficiency from an underlying productivity distribution, common to all countries, but potentially sector-specific: a Pareto distribution is usually assumed. Second, to acquire exporter status, a firm must cover bilateral fixed costs of exporting, assumed common to all firms. Bilateral exports are given by (6):

$$X_{ij} = \mu \frac{Y_i Y_j}{Y} \left(\frac{c_i \tau_{ij}}{\theta_j} \right)^{-\gamma} f_{ij}^{-\left(\frac{\gamma}{\sigma-1}-1\right)} \tag{6}$$

where μ is the share of the tradeable good in consumption, c_i is the wage in country i, τ_{ij} is the bilateral trade cost, θ_j is a measure of aggregate remoteness of country j, f_{ij} is the bilateral fixed cost, and γ is the Pareto distribution parameter which captures the degree of efficiency dispersion across firms.[49]

In the investigation of the distance puzzle, we focus on the evolution of the elasticity of trade to variable trade costs, assuming that bilateral fixed costs are not a function of distance. The elasticity of trade to variable trade costs no longer depends on consumer preferences in the heterogeneous firms' framework.[50] Rather, it is firms' efficiency dispersion which determines the aggregate trade elasticity: the lower the degree of efficiency dispersion in domestic markets, the higher the trade elasticity.[51]

The intuition for this result is given in Chaney (2008). The parameter γ which captures the degree of firms' efficiency dispersion is increasing in the fraction of small size low-productivity firms. If firms' efficiency dispersion is relatively low, the efficiency cut-off above which firms are able to acquire export status is in proximity of a substantial mass of firms. A reduction in trade barriers decreases the cut-off productivity needed to acquire exporter status, and this triggers entry of relatively many firms into exporting: lower efficiency dispersion among firms corresponds to an amplification of the aggregate trade elasticity through the extensive margin.

[48] Dixit and Stiglitz (1977); Krugman (1980).

[49] Melitz (2003); Chaney (2008). Appendix B recalls the main features of the model.

[50] See Chaney (2008): the elasticity of substitution has an inversely proportional effect on the extensive and intensive margins of trade. The elasticity of trade flows to variable trade costs is thus invariant to perceived substitutability of firm-level varieties.

[51] The proof that the elasticity of trade flows to variable trade costs is common across countries and equal to γ is given in the technical appendix to Chaney (2008). The proof uses the small country assumption.

2.2 What has been the evolution of γ through time?

In the original Melitz model generalized by Chaney to asymmetric countries, there is no mechanism to bring about a change in the aggregate trade elasticity overtime other than a shock on firms' efficiency dispersion. This is because a reduction in variable trade costs shifts productivity cut-offs for entry in domestic and foreign markets, but has no impact on firms' efficiency dispersion by assumption of the Pareto distribution, invariant to truncation. Movements in the aggregate trade elasticity would be due to changes in *measured* efficiency dispersion. Expansion in the range of traded goods and sector-specific shocks to firms' efficiency dispersion, particularly in the low-γ sectors, could shift the parameter estimated on aggregate data.[52]

The impact of intrasectoral changes on the parameter estimated on aggregate data is not immediate. This is because two things happen: first, the parameter estimated in gravity equations results from aggregation across sectors potentially characterized by different efficiency dispersion parameters. Second, trade integration over 1962-2009 has led to an expansion in the range of traded goods which could have increased the cross-sectoral dispersion of γ parameters. The extent of cross-sectoral heterogeneity may impact the parameter estimated on aggregate trade data.[53] Finally, given that the sectoral gravity relation is a power law, what matters for the aggregate parameter are the *lowest* γ sectors.[54] If trade integration leads to entry of lower γ sectors or if trade integration leads to a decrease in observed γ of the lowest γ sectors, then the aggregate parameter would decrease, deepening the distance puzzle. Similarly, the aggregate parameter increases if the lowest γ sectors see a reduction in efficiency dispersion, reducing the distance puzzle.

To give an example, suppose manufactured goods are characterized by the lowest γ, i.e. firm efficiency dispersion is strongest in this sector. Then what matters is not the increasing weight of this sector in world trade, but whether this sector is characterized by decreasing productivity dispersion. If γ-manuf has increased overtime while remaining lowest in cross-section, the trade elasticity estimated on aggregate data would increase, explaining the distance puzzle.

[52] An additional mechanism for changes in the estimated trade elasticity in the framework of n asymmetric countries with destination-specific fixed costs of exporting is provided by Di Giovanni et al. (2011). As bigger firms have a higher probability of entering more markets, the power law exponent estimated on the exports' distribution is lower than the fundamental heterogeneity parameter. If trade openness increases, with more markets a firm can potentially enter, the distribution of exports becomes more fat-tailed. In terms of our sample, this would correspond to a decreasing γ overtime, deepening the distance puzzle.

[53] Imbs et al. (2002).

[54] The inheritance mechanism for power laws states that the property of being distributed according to a power law is conserved under addition: a combination of n random variables distributed according to a power law is dominated by the one with the smallest exponent. See Gabaix (2008).

A model with producer heterogeneity and productivity dynamics is needed to investigate under which conditions active firm investment in innovation or technology spillovers would generate decreasing productivity dispersion. One possible mechanism is the reduction in imitation costs overtime. Endogenizing the Pareto exponent, Luttmer (2007) finds that if imitation becomes less costly, the distribution of operating firms has a thinner tail. This corresponds to a higher γ. In a dynamic model with endogenous innovation but no spillovers from incumbents to entrants, Atkeson and Burstein (2010) find that a reduction in variable trade costs leads to enhanced innovation incentives for exporters under the assumption of persistent exporter status. The distribution of operating firms has a relatively fatter tail. This corresponds to a lower γ.[55] Both models work with a one-sector economy while endogenous innovation and spillovers could be sector-specific. Again, sector-specific dynamics would only matter for the distance puzzle if they have an impact on the parameter caracterizing firm efficiency dispersion in the aggregate economy.

We conclude that endogenizing the investment decision and allowing for spillovers may drive a wedge between fundamental and observed efficiency dispersion, but the combined impact of these two mechanisms it not clearcut. To the best of our knowledge, there is no direct empirical evidence on the evolution of sectoral or aggregate γ overtime.

3 The Armington framework

3.1 The model

In the Anderson and van Wincoop framework, there is no intersectoral or intrasectoral heterogeneity in productive efficiency. The production process in each country and sector is constant returns to scale. The heterogeneity dimension comes from the assumption that goods produced by different countries are not homogeneous. Each country is specialized in production of country-specific varieties which on aggregate give a country-specific composite good. Supply of these composite goods is assumed perfectly inelastic. This replicates the hypothesis formulated by Armington (1969). The elasticity of substitution between country-specific varieties is therefore called an 'Armington elasticity of substitution'.

Consumers have identical homothetic preferences in all countries, approximated by a CES utility function. They perceive products of different national origin as intrinsically imperfect

55 See also Aw et al. (2009).

substitutes and choose to consume all of the available country-specific varieties.[56]

Under the assumption of symmetric bilateral trade costs and given market clearing, bilateral exports are given by (7):

$$X_{ij} = v_i Y_j \left(\frac{c_i \tau_{ij}}{P_j}\right)^{-(\sigma-1)} \tag{7}$$

where c_i is the cost of producing domestic varieties in country i, τ_{ij} is the bilateral trade cost, v_i is the preference parameter for country i products,[57] P_j is the aggregate price index in country j, and σ is the Armington elasticity of substitution.

From (7), consumer preferences determine the sensitivity of trade flows to trade costs. In the Anderson and van Wincoop specification of the gravity equation, the level of the Armington elasticity of substitution determines the potential of substituting away from distant to nearer trade partners' products.

3.2 The evolution of substitutability overtime

There is no mechanism in the Armington framework which would result in an endogenous change in trade flows' sensitivity to trade costs overtime. Any evidence on the evolution of Armington elasticities since the 1960s would point to a shock on consumer preferences or to a change in *measured* substitutability. We briefly review the available empirical evidence. Following Sato, it is common to distinguish 'upper-tier' and 'lower-tier' Armington elasticities.[58] The 'upper-tier' elasticity pertains to substitutability of domestic products and an *aggregate* import good while the 'lower-tier' elasticity measures substitutability *between* importers of a given good. It is the lower tier Armington elasticity of substitution which is relevant in the context of the gravity equation.[59]

In the investigation of the distance puzzle, we focus on the evolution of the heterogeneity parameter measured on aggregate trade data because, as shown by Imbs and Méjean (2009), trade responsiveness measured on aggregate data cannot be mimicked by a theoretically grounded weighted average of sectoral elasticities of substitution. This amounts to constraining sectoral

[56] See Krugman (1980): consumers value variety for itself because an increase in the number of varieties mechanically reduces the aggregate price index. If there were no trade costs, expenditure would be equally spread among all available varieties.

[57] In Anderson and van Wincoop (2003) $v_i = \beta_i^{1-\sigma}$ because the preference parameter enters the utility function with exponent $(1-\sigma)/\sigma$ while in Eaton and Kortum (2002) $v_i = \alpha_i^{\sigma-1}$ because the preference parameter enters the utility function with exponent $(\sigma-1)/\sigma$.

[58] Sato (1967); Reinert and Shiells (1991); Saito (2004).

[59] Feenstra (1994); Broda et al. (2006); Broda and Weinstein (2006); Welsch (2006).

substitution elasticities to equality. In the Anderson and van Wincoop framework, this corresponds to the parameter which measures the substitutability of composite country-specific goods.

Evidence on the evolution of lower-tier Armington elasticities of substitution measured on aggregate data is close to none. For France, Welsch (2006) provides estimates of aggregate lower-tier Armington elasticities since the 1970s. He finds that among exporters to the French market the elasticity peaked in the 1970s, and progressively decreased in the 1990s. Broda and Weinstein (2006) study the evolution of sectoral Armington lower-tier elasticities between 1972-1988 and 1990-2001 for American imports. They find that they have decreased for all types of goods, and at all levels of product disaggregation, e.g. at the 10-digit, 5-digit, and 3-digit levels. These results indicate that the parameter estimated on aggregate data would also have decreased, deepening the distance puzzle. But to the best of our knowledge, no paper has as yet provided evidence on the evolution of Armington elasticities for aggregate trade.

4 Summing up: heterogeneity and the gravity equation

All theoretical frameworks used to derive the gravity equation are characterized by a common feature: the elasticity of trade flows to trade costs, the 'trade elasticity', is decreasing in the degree of heterogeneity observed in a single dimension, and it is this dimension which is framework-specific. In the Ricardian framework, heterogeneity is intra-country and inter-sector: it measures the degree of dispersion in production efficiency within countries across goods, with the dispersion parameter assumed common across countries and sectors. In the Armington framework, the dimension of heterogeneity is the flipside of the Ricardian framework. Heterogeneity is inter-country and intra-sector, and corresponds to the parameter which measures the degree of perceived substitutability of country-specific varieties of each good. In the monopolistic competition framework with firm heterogeneity, the dimension of heterogeneity is intra-country and intra-sector. It is captured by the parameter which measures the degree of dispersion in firm productivity within a given sector, and this parameter is assumed common to all countries.

Three main factors could drive changes in the heterogeneity parameter measured on aggregate trade data. First, the expansion in the range of traded goods could drive changes in *measured* heterogeneity, without any changes in the degree of underlying heterogeneity. This explanation is valid across frameworks. Second, aggregation specific issues linked to estimating a single parameter across sectors potentially characterized by varying degrees of fundamental heterogeneity could evolve overtime, either due to intra-sectoral changes, or to changes in the range of traded goods. This explanation is relevant for frameworks allowing for sector specific parameters, e.g. the Armington and the heterogeneous firms' framework. Third, strategies adopted by economic agents in order to adapt to a changing economic environment, such as a reduction in trade barriers, could lead to endogenous changes in heterogeneity This explanation is particularly relevant in firm heterogeneity models which allow for endogenous innovation.

Finally, we note that there is relatively little empirical evidence on the evolution of both sector-specific and aggregate trade elasticites overtime. We attempt to bridge this gap in the following section.

IV Trade elasticity and the distance puzzle

The distance puzzle defined as the non-decreasing distance elasticity of trade flows in 1962-2009 is documented in section II. To understand the mechanics of this puzzle, we disentangle the contributing role of the two parameters which product gives the distance elasticity by estimating the elasticity of trade flows to trade costs and deducing the evolution of the remaining parameter, the elasticity of trade costs to distance. The characteristics of the data, e.g. trade flows and unit values, allow estimating σ. This parameter measures the perceived substitutability of country-composite goods in the three frameworks analyzed in section III. But it corresponds to the aggregate trade elasticity only in the Armington framework.

1 Theoretical justification

1.1 Which trade elasticity?

All frameworks used for deriving the gravity equation give an expression for aggregate bilateral trade as a function of total expenditure of the destination and of the source-specific contribution to the destination-specific price index across all sources including oneself as in (8):

$$X_{ij} = Y_j \frac{\nu_i \left(c_i \tau_{ij}\right)^{-\zeta}}{\sum_{s=1}^{N} \nu_s \left(c_s \tau_{sj}\right)^{-\zeta}} \tag{8}$$

where X_{ij} is the value of imports from i to j, c_i is the cost of the bundle of inputs in i, τ_{ij} is the bilateral trade cost, ν_i is the preference parameter for country i products,[60], Y_j is expenditure in the destination, $s = 1,...,N$ is the set of countries, and ζ is the trade elasticity. The trade elasticity parameter we are after corresponds to θ in the Ricardian framework, γ in the Melitz-Chaney framework, and $(\sigma-1)$ in the Armington framework.

In each framework consumer preferences are assumed well represented by a CES utility function. Aggregate bilateral trade can thus be written as a function of the relative price of the composite good shipped by the source relatively to the price of the composite good the destination gets from all sources, as in (9):

$$X_{ij} \quad = \quad Y_j \left(\frac{P_{ij}}{P_j}\right)^{-(\sigma-1)} \tag{9}$$

where P_{ij} is the cif price of the composite good shipped from i to j and P_j is the price index in the destination. The exponent $(\sigma-1)$ in (9) is the trade elasticity ζ in the Armington framework.

The data contains information on trade flows and unit values for each destination's bilateral imports at the SITC 4-digit disaggregation level. In 1.2 we show that this information is sufficient to estimate the aggregate trade elasticity parameter in the Armington framework. The first step is to apply to this sectoral data a consistent aggregation procedure to get relative prices of country composite goods. The second step is to estimate the substitutability parameter for total bilateral trade using these relative prices.

Our data does not allow estimating the trade elasticity parameter in the Ricardian or the Melitz-Chaney framework because it does not contain information on domestic prices in the destination.[61] The intuition for this result is the following: the threshold which determines the fraction of firms which enter any market is destination-specific. Therefore, the price distribution in each destination across all sources is needed to estimate the shape parameter of the productivity distribution.[62] In the Armington model, producer heterogeneity is not modelled, and the source-specific cost component gives directly the price of the exported good. The trade elasticity parameter can be estimated using source-specific price distributions.

[60] Alternatively, T_i, the source-specific stock of ideas in the Ricardian framework, or the mass of entrants from source i in the monopolistic competition framework. See footnote 20 in Eaton and Kortum (2002).

[61] Under the assumption of a common efficiency distribution in all sectors, $\gamma = \theta$.

[62] See Appendix A.3 and Appendix B for details.

1.2 The market share equation

This subsection uses the assumption that consumer preferences are well represented by a two-tier CES utility function to ~~to~~ construct relative prices of country composite goods using a consistent aggregation procedure for expenditure share and unit value data, reported at the SITC 4-digit disaggregation level in the COMTRADE database.

1. Define aggregate imports from source i as the sum of imports from each sector k where a sector corresponds to a SITC 4-digit category: $X_{ij} = \sum_k X_{ij}(k)$.[63]

2. Given CES utility at the intersectoral level, sectoral demand in j in sector k is given by:

$$Y_j(k) = \left(\frac{P_j(k)}{P_j}\right)^{-(\sigma-1)} Y_j$$

3. Assume each country exports a specific national variety. Preferences within each sector k between national varieties are assumed well represented by a CES utility function. Intrasectoral demand for varieties exported by i in j in sector k is:

$$X_{ij}(k) = \left(\frac{P_{ij}(k)}{P_j(k)}\right)^{-(\sigma_k-1)} Y_j(k)$$

   Replacing $Y_j(k)$ by its value gives:

$$X_{ij}(k) = \left(\frac{P_{ij}(k)}{P_j(k)}\right)^{1-\sigma_k} \left(\frac{P_j(k)}{P_j}\right)^{1-\sigma} Y_j \tag{10}$$

4. Under the assumption that $\sigma_k = \sigma'$ in all sectors and $\sigma' = \sigma$:

$$X_{ij}(k) = \left(\frac{P_{ij}(k)}{P_j}\right)^{1-\sigma} Y_j$$

5. Summing over all SITC 4-digit sectors:

$$\begin{aligned} \frac{1}{Y_j}\sum_{k=1}^{K} X_{ij}(k) &= \frac{1}{P_j}\sum_{k=1}^{K} \left(P_{ij}(k)\right)^{1-\sigma} \\ \frac{X_{ij}}{Y_j} &= \left(\frac{P_{ij}}{P_j}\right)^{1-\sigma} \end{aligned}$$

6. Going back to the expression for $X_{ij}(k)/Y_j$, use (10) to write:

$$\frac{X_{ij}(k)}{Y_j(k)}\frac{Y_j(k)}{Y_j} = \left(\frac{P_{ij}(k)}{P_j(k)}\right)^{1-\sigma} \left(\frac{P_j(k)}{P_j}\right)^{1-\sigma}$$

[63] When several quantity units are observed, the sector is defined at the product*qty-unit level.

Replacing $\frac{P_j(k)}{P_j}$ by its value and defining $\frac{Y_j(k)}{Y_j} = \omega_j(k)$, we get:

$$\frac{X_{ij}(k)}{Y_j(k)}\frac{Y_j(k)}{Y_j} \quad = \quad \left[\frac{P_{ij}(k)}{P_j(k)}\left(\omega_j(k)\right)^{1/1-\sigma}\right]^{1-\sigma}$$

7. Summing over all SITC 4-digit sectors:

$$\sum_{k=1}^{K}\frac{X_{ij}(k)}{Y_j(k)}\frac{Y_j(k)}{Y_j} \quad = \quad \sum_{k=1}^{K}\omega_j(k)\left[\frac{P_{ij}(k)}{P_j(k)}\right]^{1-\sigma}$$

8. Multiplying and dividing the right hand side of the expression by $\omega_j(k)^{1-\sigma}$ and taking logs:

$$\ln\left[\frac{X_{ij}}{Y_j}\right] \quad = \quad \ln\left\{\sum_{k=1}^{K}\omega_j(k)^{\sigma}\left[\omega_j(k)\frac{P_{ij}(k)}{P_j(k)}\right]^{1-\sigma}\right\} \tag{11}$$

9. Working with the right hand side of (11), and using the approximation that for a large number of sectors k, $\ln\sum_k X_{k,ij} \approx \sum_k \ln X_{k,ij}$:

$$\ln\left\{\sum_{k=1}^{K}\omega_j(k)^{\sigma}\left[\omega_j(k)\frac{P_{ij}(k)}{P_j(k)}\right]^{1-\sigma}\right\} \approx \sum_{k=1}^{K}\ln\left[\omega_j(k)^{\sigma}\right]+\sum_{k=1}^{K}\ln\left[\left(\omega_j(k)\frac{P_{ij}(k)}{P_j(k)}\right)^{1-\sigma}\right]$$

The first term disappears because:

$$\sigma\sum_k \ln\omega_{k,j} \quad \approx \quad \sigma\ln\left(\sum_k Y_{k,j}/Y_j\right) = \zeta\ln 1 = 0$$

Using the same approximation as previously for the second term:

$$(1-\sigma)\sum_{k=1}^{K}\ln\left[\omega_j(k)\frac{P_{ij}(k)}{P_j(k)}\right] \quad \approx \quad (1-\sigma)\ln\left[\sum_{k=1}^{K}\omega_j(k)\frac{P_{ij}(k)}{P_j(k)}\right]$$

10. The market share equation for aggregate bilateral trade as a function of the weighted average of sectoral relative prices of i in j is:

$$\ln\left[\frac{X_{ij}}{Y_j}\right] \quad \approx \quad -(\sigma-1)\ln\left[\sum_{k=1}^{K}\omega_j(k)\frac{P_{ij}(k)}{P_j(k)}\right] \tag{12}$$

11. Exponentiating gives the equation to be estimated:

$$X_{ij}/Y_j \quad = \quad \exp\left[\lambda_0-(\sigma-1)\ln\left(\sum_k \omega_k\frac{P_{ij}(k)}{P_j(k)}\right)+fe_{exp}+fe_{imp}\right]\eta_{ij} \tag{13}$$

where fe_{exp} and fe_{imp} are source and destination fixed effects, η_{ij} is a multiplicative error term, and λ_0 is a constant. Source fixed effects control for the world preference for products of this origin. Destination fixed effects control for unobserved domestic prices. The PPML estimator is used because of the heteroskedasticity of the market share equation in levels (See II.2.)

Using the properties of the CES aggregator and the assumption $\sigma' = \sigma$, the market share equation for aggregate bilateral trade can be written as a weighted average of sectoral relative prices of each source in the destination. The relative price of the composite good exported by the source is constructed in a consistent way. Indeed; each sectoral relative price is weighted according to the share of the sector in total expenditure of the destination, and identical weights are applied on both sides of the equation.[64]

Bringing the equation to the data gives the following: the market share is constructed as the ratio of total bilateral trade observed for the pair to total observed imports of the destination across all sources. Destination-specific weights $\omega_j(k)$ are constructed as the expenditure share on sector k imports in the destination to destination expenditure on imports across all sectors. Destination-specific sectoral prices $P_j(k)$ are constructed as a weighted average of observed unit values for each source in sector k where weights are given by the market share of each source in this sector in the destination: $P_j(k) = \sum w_{ij}(k) P_{ij}(k)$ with $w_{ij}(k) = X_{ij}(k)/Y_j(k)$. Sources for which some trade but no unit value is observed in the sector are excluded from the computation of the sectoral price. Source-specific sectoral prices $P_{ij}(k)$ are defined as the unit values of SITC 4-digit categories where these sectoral unit values are taken as a proxy for consistently aggregated cif prices of varieties exported within each sector k. The relative price of the composite good exported by the source is constructed as the weighted average of observed relative prices for this source. The next section examines the implications of these choices in accounting for lacking unit values and zero trade flows.

2 Dealing with missing unit values and zero trade flows

To estimate the market share equation on the COMTRADE dataset we need to tackle the question of missing information on trade flows and unit values (uv).

A first difficulty arises when the trade flow is observed but information on quantities is missing, and it is therefore not possible to compute the unit value. On average, *lacking uv* corresponds to 14% of total recorded trade in 1962-2009, with a gradual increase in coverage from

[64] To see this, rewrite aggregate imports: $\frac{X_{ij}}{Y_j} = \sum_k \frac{X_{ij}(k)}{Y_j(k)} \frac{Y_j(k)}{Y_{,j}}$ where $\frac{Y_j(k)}{Y_{,j}} = \omega_j(k)$.

83% to 90% between 1962-2000, and a subsequent decrease back to 82% in 2001-2009.[65] We assume that information on quantities is missing due to imperfections in the data collection procedure, and that the bilateral trade flow is observed with a similar degree of precision whether or not quantity has been recorded. To deal with the missing value, we assume that unobserved relative prices in these sectors are equal to the relative price of the composite good exported by the source which is constructed across those sectors where unit values have been recorded.[66]

A second difficulty arises when both quantity and trade data are missing. Zero trade flows (ztf) are a prevalent feature of the data while under model assumptions some trade should be observed in every sector k between all pairs ij. [67] We make the assumption that this information is missing because the underlying trade flow is positive but so small that it does not pass the threshold applied by the data collecting authorities. Hence, these trade flows, if recorded, would not substantially modify the distribution of observed market shares in the destination, precisely because they are an order of magnitude smaller than observed trade, i.e. the left hand side of (13) is adequately captured in the data.

This is not the case of the right-hand side of (13) because the constructed relative price of the composite good systematically underestimates the true underlying relative price if there are ztf. This is because in constructing this relative price we assume that unobserved relative prices in ztf sectors are equal to the weighted average price across sectors in which trade flows are observed while, under model assumptions, these very small and statistically unobserved trade values must correspond to a higher cif price than the maximum observed price in the destination across all sources and sectors.[68] Due to the prevalence of ztf we do not have the full set of information on the realized price distribution of the source in the destination, and the constructed relative price is an underestimation of the true relative price.

The assumption we make on unobserved prices would not be problematic if the underesti-

[65] Unit value coverage corresponds to a mean of 86% of total trade in 1962-2009, with a standard deviation of 2.4 percentage points. Annual uv coverage varies between 82-91%, with an increase from 83-85% in 1962-1973 to 86-87% in 1974-1983, and 87-90% in 1984-2000. There is a low of 82% in 1969 and 1984, and another low of 83% in 2000. In 2001-2006 uv coverage is 85-87%, and about 82% in 2007-2009.

[66] An alternative procedure would consist in imputing the relative price observed for another source which has a similar market share in this sector and destination. Results are not sensitive to using this alternative procedure.

[67] The share of observed SITC 4-digit flows relatively to the total number of potential SITC 4-digit bilateral trade flows increases from 10% to 14% in 1962-2009. Another way to show the prevalence of ztf in the data is to construct annual square samples of countries which trade with every other country in the sample in that year. If such a square is restricted to the sample of SITC 4-digit categories which each source exports to and imports from every other source, the sample corresponds to a marginal amount of world trade in any year.

[68] A necessary condition for the price to be observed is for the flow to pass the recording threshold. The price corresponding to the unobserved flow is necessarily higher than the maximum price across sectoral trade flows to the destination which have been recorded because from the intrasectoral demand equation for the CES utility function given $\sigma \geq 1$: $X_{ij}(\underline{k})/\min\{X_{sj}(k)\} < 1$ iff $P_{ij} > \max\{P_{sj}(k)\}$, $\forall s = 1,...,N$ and $\forall k = 1,...,K$.

mation factor were constant across exporters.[69] This scalar would cancel out across sources, and the estimated substitutability parameter would correspond to the true substitutability parameter. This is obviously not the case. Table 2 shows that the underestimation factor is not constant across exporters: the share of ztf is strongly decreasing in market share. The relative price of the composite good is underestimated by more for small exporters.[70]

Table 2: **Proportion of zero trade flows as a function of market share**

depvar: Share of ZTF	(1)	(2)	(3)	(4)
ms	-0.0401***	-0.2446***	-0.0427***	-0.2573***
	(0.0001)	(0.0134)	(0.0001)	(0.013)
year	-0.0029***	-0.0020***	-0.0033***	-0.0024***
	(0.0000)	(0.0001)	(0.0000)	(0.000)
$ms * year$		0.0001***		0.0001***
		(0.0000)		(0.000)
constant	5.3474***	3.5852***	6.0976***	4.2515***
	(0.0335)	(0.1372)	(0.0366)	(0.134)
Destination FE	NO	NO	YES	YES
Observations	657001	657001	657001	657001

Notes: The share of ZTF is computed at the SITC 4-digit level. The estimation is conducted in PPML in order to include observations where ztf=0. The log of the market share is used in the estimation. Destination fixed effects are included in (3) and (4). Robust standard errors are in parentheses. *** p<0.01.

For a given observed distribution of market shares, the underlying dispersion in relative prices of the composite good is greater than the observed dispersion in relative prices. The estimated parameter $\widetilde{\sigma}$ *overestimates* the underlying true substitutability parameter σ.

Regression results presented in table 2 include an interaction term for the market share and year which coefficient is significant and positively signed. This means that the relationship be-

[69] An alternative method consists in imputing unobserved relative prices with some arbitrary price above the maximum observed in the destination. As ztf constitute 85-90% of all 4-digit trade flows, this method is problematic because results are driven by imputed rather than effectively observed prices.

[70] See Appendix A.3 for the proof that unobserved prices are necessarily higher than the maximum observed price and that lower market shares are necessarily associated with a higher proportion of ztf, e.g. the relative price of the composite good is necessarily underestimated by more for small exporters.

tween market share and the proportion of ztf becomes less tight overtime: the reduction in the share of ztf proceeds at quicker pace in 1962-2009 for small exporters.[71]

Table 3 presents the predicted share of ztf for 4 types of exporters in 1962 and 2009. For a very small exporter with .02% market share, the initial share of ztf is predicted to be .95, and it is reduced to .83 by 2009, i.e. a 12 percentage point decrease. Consider a relatively big exporter, with a 10% market share: its share of ztf is reduced from .72 to .65, a 7 percentage point decrease. The assumption that the relative price of the composite good is underestimated by a constant factor for all exporters becomes relatively more plausible overtime as the gap between the share of ztf for big and small exporters is reduced. The estimated substitutability parameter $\widetilde{\sigma}$ consistently overestimates the underlying σ; but by less overtime.

Table 3: **Predicted share of ztf for exporters with different market share, 4-digit level**

year	ms=0.02%	ms=1%	ms=10%	ms=28.7%
1962	0.95	0.80	0.72	0.69
2009	0.83	0.71	0.65	0.62

Columns (1) and (4) correspond to the mean and to 2 st. deviations above the mean in the distribution of log market share. Columns (2) and (3) correspond to the mean and to 2 st. deviations above the mean in the distribution of market share.

We conclude that the hypothesis we make on unobserved sectoral prices due to ztf does not impede interpreting the evolution of the true underlying trade elasticity if it is found that the estimated parameter $\widetilde{\sigma}$ increases in absolute value overtime. Indeed, this evolution would provide a lower bound on the increase in the absolute value of the underlying trade elasticity because we have shown that the true substitutability parameter is overestimated by less overtime.

3 Results

In this section we present the results on the evolution of $\widetilde{\sigma}$ obtained when (13) is estimated on annual crossections of the COMTRADE dataset.

We proceed as follows. First, the relative price of each source in the destination is con-

[71] Taking column (4), the full year effect for a given market share is given by $(-0.0024+0.0001*ln(ms))$. The second term, given by the interaction effect, is increasing in the market share. The full year effect is stronger for smaller exporters.

structed at the highest disaggregation level: for each product and quantity unit in which the source is active. Second, we proceed level by level for aggregation: the relative price of the composite sectoral good of the source is first constructed at the 3-digit level using the weighted average relative price observed at the 4-digit level, with destination-specific weights for each variety of the 4-digit good the source is active in. Then given relative prices constructed at the 3-digit level, destination-specific weights are used to aggregate these up to 2-digit levels, and so on until the relative price for the composite good is constructed using relative prices at the 1-digit level. Measures of market share are unchanged by this hierarchical aggregation procedure because at each step the market share observed for the source at that aggregation level remains unchanged under the assumption that unobserved trade flows at that aggregation level are arbitrarily small. Figure 8 presents the results: the trade elasticity is found to increase by about 29% in 1962-2009.[72]

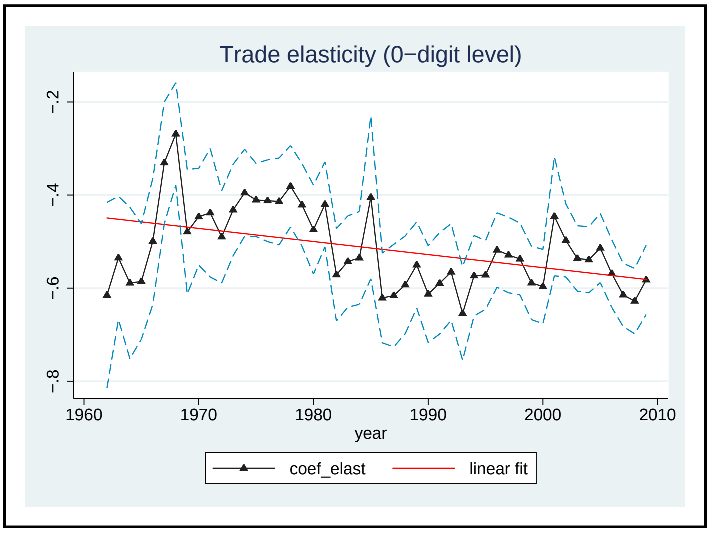


Figure 8: Estimated $(1-\widetilde{\sigma})$, hierarchical construction of the relative price for the composite good

4 Instrumenting: motivation and results

The results presented in the previous section are subject to caution for at least three sets of reasons. First, we do not have information on domestic prices or domestic absorption. Second, unit values are plagued by measurement error. Third, if supply schedules are not horizontal, contrary to the assumption made in the estimation procedure presented in the previous section, the

[72] Appendix (D) discusses the robustness of this result.

demand elasticity parameter estimated in the market share equation will suffer from attenuation bias due to not controlling for potentially positive and finite supply elasticities. This attenuation bias would not be problematic for analyzing the evolution of the substitutability parameter if only the level of the parameter were affected. The problem arises because, as shown by Feenstra (1994), this attenuation bias impacts both the level and the evolution of the parameter.[73] To gauge the robustness of our results, we reestimate equation (13) after instrumenting the relative price of the composite good.

4.1 Procedure

The objective is to find an instrument which adequately captures exporter-specific shocks to the price of the composite good which are not demand-driven, such as exogenous shocks to the price of inputs. One possible instrument would be the Producer Price Index (PPI) since it captures the evolution of prices faced by producers on the inputs' side. Unfortunately, we do not have PPI data for most countries and years in our sample. We therefore settle for an alternative exporter-specific price level indicator: the GDP price level in international prices as reported in the Penn World Tables for 189 countries in 1950-2009.[74]

The instrumenting procedure is the following. First, we compute the evolution of the relative prices for country-specific composite goods constructed hierarchically and aggregated up to composite good prices at the level of aggregate trade. Second, for each destination, we compute the mean evolution of gdp price levels of its sources, weighted by their market shares in this destination. Third, we predict the growth rate of each exporter's relative price in the destination on the evolution of this exporter's gdp prices relatively to all other exporters to the same destination. Fourth, we construct the predicted relative price of the composite good in year $(t+1)$ as the product of the observed relative price in t and the predicted growth rate. Finally, we estimate (13) using the constructed relative price instead of the observed relative price.

Two caveats of this procedure should be mentioned. First, the instrument is weak: while the point estimate is always signed as expected, e.g. increasing domestic prices predict increasing relative cif prices of exports, this estimate is not significant. We suspect that this is primarily due to the fact that our instrument is a very imprecise indicator of shocks to input prices

[73] See Feenstra (1994) for a detailed presentation of the nature of the problem, and for an alternative methodology to the one followed in this paper for dealing with the problem.

[74] See Heston et al. (2011). Our sample of countries does not coincide perfectly with the countries in the Penn World Tables. However, as it is the smallest exporters which drop out, the sample adjustment in terms of world trade coverage is minor.

which combined with incomplete pass-through of input price shocks to exports' prices results in instrument weakness. This could also be due to measurement error. Second, it could be argued that allowing for just one lag inadequately captures the temporal relationship between shocks to inputs' prices and their pass-through to the price of exported output. Indeed, if prices are relatively persistent, then the instrumenting procedure would amount to little more than replacing observed prices in $t+1$ with lagged observed prices in t. We therefore also estimate (13) using relative prices constructed with predicted price evolution in $t+s$ where $s = 1,...,10$: $\overline{P_{ij,t+s}} = P_{ij,t} * \prod_{s=1}^{S} \overline{g_{i,s}}$ where $\overline{g_{i,s}}$ is the predicted growth rate between s and $s-1$. Appendix (E) shows that increasing the number of lags reinforces the result obtained with one lag.

4.2 Results with instrumented prices

Results obtained with one lag ($s = 1$) are shown in Fig.(9). The substitutability parameter has increased by 12% in 1962-2009 while its level increases by 8-23% relatively to the estimate obtained with non-instrumented prices.

This result is robust to increasing the number of periods in which the evolution of exports'

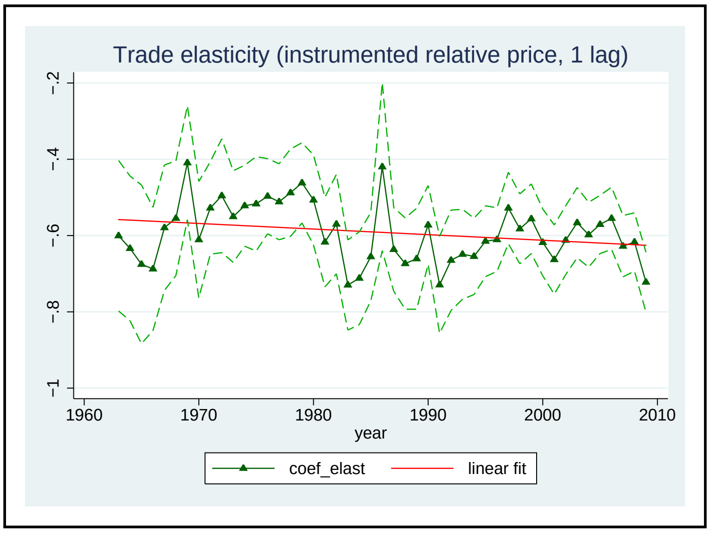


Figure 9: Estimated $(1-\widetilde{\sigma})$, instrumented relative price of composite good

prices is predicted with the evolution of domestic prices (see Appendix (E)). Thus, for $s = 10$, the elasticity increases by 21% in 1972-2009, and the absolute value of the parameter increases by 25-26% relatively to the estimate obtained with non-instrumented prices. The evolution of the parameter becomes steeper as we increase the number of lags. Therefore, it is likely that our estimate provides a lower bound on the increase in the true substitutability parameter.

5 Evolution of the trade elasticity in the Armington framework

What is the economic interpretation of an increasing substitutability parameter measured on aggregate data?

First, the degree of perceived similarity of country-composite goods may have increased. Since the 1960s, a growing number of countries started producing a set of goods similar to that of developed countries. This process has increased the number of available varieties and, potentially, their degree of similarity.[75]

Second, composition effects may lead to changes in the parameter estimated on aggregate data. If the reduction in trade barriers has led to the expansion in the range of traded goods, trade in previously non-traded sectors could modify measured substitutability of country-composite goods. Non-uniform reductions in sectoral trade costs would also modify the composition of world trade, leading to a change in the substitutability parameter measured on aggregate data.

Ideally, we would like to separate out the impact of composition and sector-specific effects to quantify the net effect of increased perceived similarity of country-composite goods. This is however impossible because, as shown by Imbs and Méjean (2009), the parameter estimated on aggregate data cannot be mimicked by a weighted average of sectoral parameters. The bottom line is that an increase in measured substitutability for country-composite goods is consistent with complex competition dynamics in price and quality documented by Amiti and Khandelwal (2012) as well as with increased vertical specialization of countries within sectors documented by Fontagné et al. (2008).

6 The evolution of the elasticity of trade costs to distance: is there a distance puzzle?

This section has provided empirical evidence on the evolution of the aggregate substitutability parameter for world trade in 1962-2009. This substitutability parameter corresponds to the aggregate trade elasticity in the Armington framework. We find that this parameter has increased by 29% between 1962 and 2009 in the benchmark estimation, and by 21% when relative prices are instrumented. Both estimates are likely to be providing lower bounds on the increase in the true substitutability parameter. Section II has shown that the distance elasticity of trade has increased by 8% over the same period. Combining these two results to evaluate the magnitude of the distance puzzle redefined as a *non-decreasing elasticity of trade costs to distance*,

[75] Schott (2004) documents increased similarity in the set of exported goods of US trade partners while Broda et al. (2006) document the increase in the number of imported varieties since the 1970s.

we conclude that there is no distance puzzle in the framework of the Armington model in as much as the elasticity of trade costs to distance has decreased by 19% in 1962-2009. The non-decreasing distance elasticity of trade is fully explained by *increased perceived substitutability* of country-specific composite goods.

V Conclusion

The estimated effect of distance in gravity equations has remained stable, or even slightly increased, in the past 50 years, despite substantial innovation in transportation and communication: this is the 'distance puzzle'. Using COMTRADE 4-digit bilateral trade data in 1962-2009, this paper finds that the evolution of the elasticity of trade flows to trade costs, referred to as the 'trade elasticity', provides a direct explanation of the non-decreasing distance elasticity of trade. Increased sensitivity of trade flows to relative prices has more than compensated the reduction in the elasticity of trade costs to distance.

The paper proceeds in three steps. First, it shows that the distance puzzle is a feature of our data by estimating yearly cross-section gravity equations. In our baseline estimation the distance coefficient has increased by 8% from 1962 to 2009. This result qualitatively holds when we correct for changes in the sample of trading partners and the composition of world trade. Taking into account FTAs seems to solve the distance puzzle, but this is an artefact of their growing importance: introducing FTAs dummies amounts to adding a time-growing number of proximity controls in the estimation.

Second, the paper underlines that the main theoretical foundations of the gravity equation all show that the distance coefficient is the product of the elasticity of trade costs to distance and a measure of heterogeneity. In the Ricardian framework, heterogeneity is intra-country and inter-sector. The trade elasticity corresponds to the degree of dispersion in productivity within countries across goods. In monopolistic competition with firm heterogeneity, the dispersion parameter is intra-country and intra-sector. The trade elasticity corresponds to the degree of dispersion in intrasectoral firm productivity. In the Armington framework, heterogeneity is inter-country and intra-sector. The trade elasticity corresponds to the degree of perceived substitutability of country-specific varieties of each good.

Third, the paper estimates the evolution of the 'trade elasticity' in the Armington framework (i.e. the substitution elasticity between country-specific aggregate goods). It uses 4-digit unit values as proxies for prices. However, they are not available when no trade in a product is recorded between two countries. Unobserved unit values are assumed equal to the observed weighted mean relative unit value of the source: this causes an underestimation of the price of the composite good. As unobserved relative unit values are less prevalent for large than for small exporters, this assumption leads to an overestimation of the substitutability parameter. As the number of unobserved unit values declines faster for small exporters, the overestimation bias is reduced overtime. As the overestimation bias is not fully eliminated by the end of the sample, our estimate provides a lower bound on the increase in the absolute value of the underlying trade elasticity. Given that the estimated elasticity increases by 29% in our estimations, the evolution of the distance coefficient is compatible with a decrease of the elasticity of trade costs to distance of at least 19%. This result is confirmed when relative prices are instrumented.

References

Amiti, M. and A. Khandelwal (2012). Import competition and quality upgrading. *Review of Economics and Statistics (forthcoming).*

Anderson, J. E. and E. van Wincoop (2003). Gravity with gravitas: A solution to the border puzzle. *American Economic Review 93*(1), 170–192.

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2009). New trade models, same old gains? *NBER Working Paper Series, 15628, (forthcoming American Economic Review).*

Armington, P. S. (1969). A theory of demand for products distinguished by place of production. *IMF Staff Papers 16*(1), 159–178.

Atkeson, A. and A. T. Burstein (2010). Innovation, firm dynamics, and international trade. *Journal of Political Economy 118*(3).

Aw, B. Y., R. M. J., and D. Y. Xu (2009). R&d investment, exporting, and productivity dynamics. *NBER Working Paper Series, 14670, (forthcoming American Economic Review).*

Baldwin, R. and J. Harrigan (2007). Zeros, quality and space: trade theory and trade evidence. *NBER Working Paper Series* (13214).

Berthelon, M. and C. Freund (2008). On the conservation of distance in international trade. *Journal of International Economics 75*(2), 310–320.

Boldrin, M. and J. A. Scheinkman (1988). Learning by doing, international trade and growth: A note. In P. W. Anderson, K. J. Arrow, and D. Pines (Eds.), *The Economy as an Evolving Complex System*. Reading, M.A.: Addison-Wesley Publishing Company.

Bosquet, C. and H. Boulhol (2009). Gravity, log of gravity and the distance puzzle. *GREQAM Working Paper 2009*(12).

Broda, C., J. Greenfield, and D. Weinstein (2006). From groundnuts to globalization: A structural estimate of trade and growth. *NBER Working Paper Series* (12512).

Broda, C. and D. E. Weinstein (2006). Globalization and the gains from variety. *Quarterly Journal of Economics 121*(2), 541–585.

Brun, J.-F., C. Carrère, P. Guillaumont, and J. d. Melo (2005). Has distance died? evidence from a panel gravity model. *World Bank Economic Review 19*, 99–120.

Buch, C. M., J. Kleinert, and F. Toubal (2004). The distance puzzle: on the interpretation of the distance coefficient in gravity equations. *Economics Letters 83*, 293–298.

Burger, M. J., F. G. van Oort, and G.-J. Linders (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis 4*(2), 167–190.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution will change our Lives*. London: Orion Business Books.

Caliendo, L. and F. Parro (2011). Estimates of the trade and welfare effects of nafta. *mimeo, University of Chicago*.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–1721.

Chatterjee, S. and E. Rossi-Hansberg (2012). Spinoffs and the market for ideas. *International Economic Review, forthcoming*.

Coe, D., A. Subramanian, and N. Tamirisa (2007). The missing globalization puzzle: Evidence of the declining importance of distance. *IMF Staff Papers 54*(1), 34–58.

Combes, P.-P., T. Mayer, and J.-F. Thisse (2006). *Economie Géographique*. Paris: Economica.

Costinot, A., D. Donaldson, and I. Komunjer (2010). What goods do countries trade? a quantitative exploration of ricardo's ideas.

Crawford, J.-A. and R. V. Fiorentino (2005). The changing landscape of regional trade agreements. *WTO Discussion Papers* (8).

Crozet, M. and P. Koenig (2010). Structural gravity equations with intensive and extensive margins. *Canadian Journal of Economics 43*(1).

Daudin, G. (2003). La logistique de la mondialisation. *Revue de l'OFCE* (87), 411–435.

Daudin, G. (2005). Les transactions de la mondialisation. *Revue de l'OFCE* (92), 223–262.

Davis, D. R. (1995). Intra-industry trade: A heckscher-ohlin-ricardo approach. *Journal of International Economics 39*(3-4), 201–226.

Di Giovanni, J., A. Levchenko, and R. Ranciére (2011). Power laws in firm size and openness to trade: Measurement and implications. *Journal of International Economics 85*(1), 42–52.

Disdier, A.-C. and K. Head (2008). The puzzling persistence of the distance effect in bilateral trade. *The Review of Economics and Statistics 90*(1), 37–48.

Dixit, A. K. and J. E. Stiglitz (1977). Monopolistic competition and optimum product diversity. *American Economic Review 67*, 297–308.

Duranton, G. and M. Storper (2008). Rising trade costs. agglomeration and trade with endogenous transaction costs. *Canadian Journal of Economics 41*(1), 292–319.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Eaton, J. and S. Kortum (2006). Innovation, diffusion, and trade. *NBER Working Paper Series 12385*.

Eaton, J. and S. Kortum (2010). *Technology in the Global Economy: A Framework for Quantitative Analysis*. Unpublished manuscript.

Eaton, J., S. Kortum, and S. Sotelo (2011). International trade: Linking micro and macro. *University of Chicago, mimeo*.

Feenstra, R. (1994). New product varieties and the measurement of international prices. *The American Economic Review 84*(1), 157–177.

Felbermayr, G. J. and W. Kohler (2006). Exploring the intensive and extensive margins of trade. *Review of World Economics 142*(4), 642–674.

Feyrer, J. (2009). Trade and income: Exploiting time series in geography. *NBER Working Paper Series* (14910).

Fontagné, L., G. Gaulier, and S. Zignago (2008). North-south competition in quality. *Economic Policy 23*(53), 51–91.

Fontagné, L. and S. Zignago (2007). A re-evaluation of the impact of regional trade agreements on trade patterns. *Economic Internationale 109*, 31–51.

Friedman, T. L. (2007). *The World is Flat: A Brief History of the Twenty-first Century, 2nd edition*. Farrar Straus & Giroux.

Gabaix, X. (2008). Power laws in economics and finance. *Annual Review of Economics 1*, 255–293.

Greene, W. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models. *Stern School of Business Department of Economics Working Paper* (94-10).

Greene, W. (2003). *Econometric Analysis* (5th ed.). Upper Saddle River, N.J.: Prentice Hall.

Helpman, E., M. Melitz, and Y. Rubinstein (2008). Estimating trade flows: trading partners and trading volumes. *Quarterly Journal of Economics 123*(2), 441–487.

Heston, A., R. Summers, and B. Aten (2011, May). Penn world table version 7.0. *Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania*.

Hummels, D. (1999). Have international transportation costs declined?

Hummels, D. (2001). Towards a geography of trade costs.

Hummels, D. (2007). Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives 21*(3), 131–154.

Imbs, J. and I. Méjean (2009). Elasticity optimism.

Imbs, J., H. Mumtaz, M. O. Ravn, and H. Rey (2002). Ppp strikes back: Aggregation and the real exchange rate.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review 70*, 950–959.

Krugman, P. (1987). The narrow moving band, the dutch disease, and the competitive consequences of mrs. thatcher: Notes on trade in the presence of dynamic scale economies. *Journal of Development Economics 27*, 41–55.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*(1), 1–14.

Levchenko, A. A. and J. Zhang (2011). The evolution of comparative advantage: Measurement and welfare implications. *NBER Working Paper Series 16806*.

Levinson, M. (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton: Princeton University Press.

Lucas, R. E. J. (1988). On the mechanics of economic development. *Journal of Monetary Economics 22*(1), 3–42.

Luttmer, E. G. (2007). Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics 122*(3), 1103–1144.

Manning, W. G. and J. Mullahy (1999). Estimating log models: To transform or not to transform?

Márquez-Ramos, L., I. Martinez-Zarzoso, and C. Suárez-Burguet (2007). The role of distance in gravity regressions: Is there really a missing globalization puzzle? *B.E. Journal of Economic Analysis and Policy 7*(1).

Martin, W. and C. S. Pham (2008). Estimating the gravity model when zero trade flows are frequent. *World Bank Working Paper, unpublished*.

Melitz, M. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Mullahy, J. (1986). Specification and testing of some modified count data models. *Journal of Econometrics 33*(3), 341–365.

Murat, M. and F. Pigliaru (1998). International trade and uneven growth: a model with intersectoral spillovers of knowledge. *The Journal of International Trade and Economic Development: An International and Comparative Review 7*(2), 221–236.

Reinert, K. A. and C. R. Shiells (1991). Trade substitution elasticities for analysis of a north american free trade area. estimations of non-nested and lower-tier nested Armington elasticities for the US.

Rodríguez-Clare, A. (2007). Trade, diffusion, and the gains from openness. *Unpublished manuscript*.

Rodríguez-Clare, A. (2010). Offshoring in a ricardian world. *American Economic Journal: Macroeconomics 2*(2), 227–258.

Saito, M. (2004). Armington elasticities in intermediate inputs' trade: a problem in using multilateral data. *The Canadian Journal of Economics 37*(4), 1097–1117.

Santos-Silva, J. M. C. and S. Tenreyro (2006). The log of gravity. *The Review of Economics and Statistics 88*(4), 641–658.

Sato, K. (1967). A two-level constant-elasticity-of-substitution production function. *Review of Economic Studies 34*, 201–218.

Schott, P. (2004). Across product vs. within-product specialization in international trade. *The Quarterly Journal of Economics 119*(2), 647–678.

Simonovska, I. and M. Waugh (2011). The elasticity of trade: Estimates and evidence.

Welsch, H. (2006). Armington elasticities and induced intra-industry specialization: The case of france, 1970-1997. *Economic Modelling 23*, 556–567. Armington elasticities' estimates for France '70-'90s.

Young, A. (1991). Learning by doing and the dynamic effects of international trade. *Quarterly Journal of Economics 106*(2), 369–405.

A The Eaton and Kortum framework

This appendix goes over the main features of the Ricardian microfoundation of the gravity structure of trade. First, the Eaton and Kortum (2002) model is presented in which there is a continuum of sectors and perfect competition, and the computation of the price index is recalled. It is shown that an equation similar to (13) can be derived. It is shown that it is not possible to estimate the heterogeneity parameter driving the distance elasticity of trade in the Ricardian framework, e.g. the degree of efficiency variability across goods, in perfect competition with data on observed cif prices of bilateral exports. Second, following Eaton and Kortum (2010), it is recalled how the model accomodates monopolistic competition with an upper bound on the number of available intrasectoral varieties, and the computation of the price index is recalled. It is shown that equation (13) can be derived in the Ricardian framework with monopolistic competition but that it does not allow estimating the efficiency heterogeneity parameter. Only the substitutability parameter σ can be estimated using data on observed cif prices. This appendix concludes that to estimate the efficiency heterogeneity parameter in the Ricardian framework with price data it is necessary to work with *destination-specific* price indices.[76]

1. ***The Ricardian model in perfect competition (Eaton and Kortum (2002, 2010))***.

   (a) In each country i, there is a continuum of sectors k. Output in each sector is homogeneous, within and across producing countries.

   (b) Output can be produced using one of the production techniques for sector k available in country i. Production techniques vary in efficiency z. Perfect competition corresponds to a world of freely available technology, e.g. only the best technique is used for production in each sector. Further, it is assumed that each source accumulates its own stock of technology from which only its firms can make draws: it is through trade that countries acquire access to foreign technology stocks.

   (c) The fundamental building block of the model is the description of technology improvement overtime within each sector k. Efficiency is drawn from a Pareto distribution with parameter θ and efficiency lower bound at $\underline{z}$:

$$\Pr\left[Z > z\right] = \left(z/\underline{z}\right)^{-\theta}$$

[76] Such methodologies have been implemented by Eaton and Kortum (2002); Simonovska and Waugh (2011) for a point in time. We do not have sufficient price data to implement this procedure for 1962-2009.

Techniques' arrival follows a Poisson process with parameter $\beta R(t)$, denoting research productivity and effort, respectively. Normalizing $\beta \underline{z}^{\theta} = 1$, and defining the stock of techniques available at t by $T(t) = \int_{-\infty}^{t} R(v)\,\mathrm{d}v$, the number of techniques for producing output in sector k with efficiency $Z > z$ is distributed Poisson with parameter $\lambda = T(t)z^{-\theta}$.

(d) Given a Poisson process for ideas' arrival in any sector k, the probability of no technique with efficiency $Z > z$ arriving in a unit interval is given by the Poisson density for $X = 0$:

$$\Pr[Z \leq z] \;=\; \Pr[X = 0] = \frac{\left(T(t)z^{-\theta}\right)^{0} \exp\left\{-T(t)z^{-\theta}\right\}}{0!} = \exp\left\{-T(t)z^{-\theta}\right\}$$

The probability that a technique of efficiency $Z > z$ occurs is then distributed Fréchet with parameter $T(t)z^{-\theta}$:

$$\Pr[Z > z] \;=\; \Pr[X > 0] = 1 - \Pr[X = 0] = 1 - \exp\left\{-T(t)z^{-\theta}\right\}$$

(e) In perfect competition, only the best idea is operational within each country-sector, and its distribution is Fréchet. As the technology improvement process takes place independently within each sector k of which there is a continuum, the distribution of the best-of ideas across the continuum of sectors in each country is also Fréchet with parameter $T_i(t)z^{-\theta}$:

$$\Pr[Z_i > z] = 1 - \exp\left[-T_i(t)z^{-\theta}\right]$$

The location parameter of the distribution $T_i(t)$ describes the country-specific stock of techniques (absolute advantage), while the variability parameter θ common to all countries measures the strength of comparative advantage (high θ means low efficiency variability).

(f) The cost of the bundle of inputs in country i, c_i, is the same across sectors. Sectors do not differ in input shares, inputs are mobile, production is CRS. Factors of production can thus be shifted across sectors without bidding up factor prices.

(g) The unit cost of producing k in i is the realization of a random variable $W_i = c_i/Z_i$ where efficiency of country i in sector k is the realization of the random variable Z_i, with independent draws for each sector from the Fréchet distribution. With iceberg transport costs τ_{ij}, delivered price in destination j is the realization of a random variable $P_{ij} = c_i \tau_{ij}/Z_i$.

(h) Using (d) and (f), the distribution of prices from i to j for a given T_i is:

$$G_{ij}(p) = \Pr\left[P_{ij} \le p\right] = \Pr\left[Z_i \ge \frac{c_i \tau_{ij}}{p}\right] = 1 - \exp\left[-T_i\left(c_i \tau_{ij}\right)^{-\theta} p^{\theta}\right]$$

(i) Consumers have CES preferences over the continuum of sectoral goods:

$$U = \left[\int_0^1 Q(k)^{\frac{\sigma-1}{\sigma}}\,\mathrm{d}k\right]^{\frac{\sigma}{\sigma-1}}$$

(j) Product homogeneity entails that each destination $j = \{1, ..., N\}$ buys sectoral output from the lowest cost supplier. Price realization in j is:

$$P_j(k) = \min\{P_{sj}(k); s = 1, ..., N\}$$

(k) The distribution of prices in destination j is then:

$$\begin{aligned} G_j(p) &= \Pr\left[P_j \le p\right] = 1 - \prod_{s=1}^{N}\left[\Pr\left(P_{sj} > p\right)\right] = 1 - \prod_{s=1}^{N}\left[1 - G_{sj}(p)\right] \\ &= 1 - \exp\left[-\sum_{s=1}^{N} T_s\left(c_s \tau_{sj}\right)^{-\theta} p^{\theta}\right] \end{aligned}$$

(l) The probability π_{ij} that source i is the supplier to j in any sector k is given by the probability that i has the lowest delivered price for sector k output among all possible sources:

$$\pi_{ij} = \Pr\left[P_{ij}(k) \le \min\left\{P_{sj}(k); s \ne i\right\}\right] = \int_0^{\infty} \prod_{s \ne i}\left(1 - G_{sj}(p)\right)\mathrm{d}G_{ij}(p)$$

where the first term corresponds to the joint probability across all other sources that their price is at least as high as p, and the second term is the density of the price distribution in i. This density is given by:

$$dG_{ij}(p) = T_i\left(c_i \tau_{ij}\right)^{-\theta} \theta p^{\theta-1} \exp\left\{-T_i\left(c_i \tau_{ij}\right)^{-\theta} p^{\theta}\right\} dp$$

The joint probability across all other sources is given by:

$$\prod_{s \ne i}\left(1 - G_{sj}(p)\right) = \exp\left[-\sum_{s \ne i} T_s\left(c_s \tau_{sj}\right)^{-\theta} p^{\theta}\right]$$

Replacing in the expression for π_{ij}, and rearranging:

$$\begin{aligned}
\pi_{ij} &= T_i \left(c_i \tau_{ij}\right)^{-\theta} \int\limits_0^\infty \exp\left[-\sum_{s=1}^N T_s \left(c_s \tau_{sj}\right)^{-\theta} p^\theta\right] \theta p^{\theta-1}\, \mathrm{d}p \\
&= \frac{T_i \left(c_i \tau_{ij}\right)^{-\theta}}{-\sum_{s=1}^N T_s \left(c_s \tau_{sj}\right)^{-\theta}} \int\limits_0^\infty \exp\left[-\sum_{i=s}^N T_s \left(c_s \tau_{sj}\right)^{-\theta} p^\theta\right] \left[-\sum_{s=1}^N T_s \left(c_s \tau_{sj}\right)^{-\theta}\right] \theta p^{\theta-1}\, \mathrm{d}p
\end{aligned}$$

(m) Defining the destination-specific price distribution parameter as $\Phi_j = \sum T_s \left(c_s \tau_{sj}\right)^{-\theta}$, and integrating over the price distribution:

$$\pi_{ij} \quad = \quad -\frac{T_i \left(c_i \tau_{ij}\right)^{-\theta}}{\Phi_j} \left[\exp\left\{-\Phi_j p^\theta\right\}\right]_0^\infty = \frac{T_i \left(c_i \tau_{ij}\right)^{-\theta}}{\Phi_j} \tag{14}$$

By the law of large numbers, this probability is also the fraction of expenditure in j on imports from i: $\pi_{ij} = X_{ij}/Y_j$. This is the left-hand side in the estimated equation (13).

2. ***Price index computation in perfect competition and θ estimation with our data***

To estimate θ given our data characteristics, the right-hand side of the equation has to be expressed in terms of observed relative prices of the source and destination countries. It is shown that θ cannot be estimated with our baseline procedure because it requires observing the *potential* source i price index which cannot be observed because of trade selection. To estimate the heterogeneity parameter in the Ricardian framework, it is thus necessary to devise an estimation methodology which works with destination-specific price indices.[77]

(a) For CES preferences, the ideal price index across the continuum of goods which j consumes is:

$$I_j(p)^{1-\sigma} = \int\limits_0^\infty p^{1-\sigma}\, \mathrm{d}G_j(p)$$

(b) In perfect competition, only the least cost goods delivered to destination j are effectively consumed. Thus, the destination specific price index is given by the $(1-\sigma)$

[77] To estimate the evolution of θ overtime, the procedure devised by Eaton and Kortum (2002) and modified by Simonovska and Waugh (2011) should be used, but it cannot be applied to our data because it has no information on domestic prices.

moment of the least cost distribution in destination j:

$$I_j(p)^{1-\sigma} = E\left[\left(W^{(1)}\right)^{1-\sigma}\right]$$

The properties of the distribution of the ordered costs help compute the relevant moment of the costs' distribution.

(c) Given truncation invariance of the Pareto, the conditional efficiency distribution for Z is:

$$\Pr\left[Z \geq z' | Z \geq z\right] = \left(z'/z\right)^{-\theta}$$

Using the definition $W = c/Z$, the conditional costs' distribution for $(w' < w)$ is:

$$\Pr\left[c/W \geq c/w' | c/W \geq c/w\right] = \Pr\left[W \leq w' | W \leq w\right] = \left(w'/w\right)^{\theta}$$

(d) Replacing z by its value in the Poisson distribution of techniques' efficiency, the number of techniques for production of output in a given sector k is distributed Poisson with parameter $T(t)c^{-\theta}w^{\theta}$ (using $z = c/w$). The probability of no technique allowing production with cost less than w arriving in a unit interval is given by $\Pr[X=0] = \frac{\left(T(t)(c/w)^{-\theta}\right)^0 \exp\left\{-T(t)(c/w)^{-\theta}\right\}}{0!} = \exp\left\{-T(t)(c/w)^{-\theta}\right\}$. Then the probability of a lower cost draw arriving is given by $1 - \Pr[X=0]$. The distribution of the lowest cost is then Weibull with parameter $T(t)c^{-\theta}w^{\theta}$:

$$F(w) = \Pr[W \leq w] = 1 - \exp\left\{-T(t)c^{-\theta}w^{\theta}\right\}$$

(e) Similarly, the second-least cost probability distribution is given by the probability that at least 2 techniques had arrived which cost of production was lower than w:

$$F_2(w) = \Pr[W^{(2)} \leq w] = 1 - \sum_{v=0}^{1} \frac{\left[T(t)c^{-\theta}w^{\theta}\right]^v \exp\left\{-T(t)c^{-\theta}w^{\theta}\right\}}{v!}$$

This is just $1 - \Pr[X=0] - \Pr[X=1]$. Indeed, for the second-lowest cost to be lower than w, we need to subtract the probability that no cost draw was lower than w as well as the probability that only one draw were lower than w.

(f) More generally, ordered costs $W^{(\alpha)}$, where α denotes the rank of the cost in the pool of available techniques, are random variables which have a gamma distribution given by:[78]

$$F_\alpha(w) \;=\; \Pr[W^{(\alpha)} \leq w] = 1 - \sum_{v=0}^{\alpha-1} \frac{\left[T(t)c^{-\theta}w^{\theta}\right]^{v} \exp\left\{-T(t)c^{-\theta}w^{\theta}\right\}}{v!}$$

(g) Defining $\Phi = T(t)c^{-\theta}$ and $\lambda = \theta w^{\theta-1}\Phi$, the ordered costs' pdf is given by:

$$F'_\alpha(w) \;=\; \lambda \exp\left\{-\Phi w^{\theta}\right\} - \exp\left\{-\Phi w^{\theta}\right\} \left[\sum_{v=1}^{\alpha-1} \frac{\left(\Phi w^{\theta}\right)^{v-1} v\lambda}{v!} - \frac{\lambda\left(\Phi w^{\theta}\right)^{v}}{v!}\right]$$

where the summation in squared brackets simplifies to $\left[\lambda - \frac{\lambda\left(\Phi w^{\theta}\right)^{\alpha-1}}{(\alpha-1)!}\right]$. The pdf is:

$$f_\alpha(w) \;=\; \frac{\lambda\left(\Phi w^{\theta}\right)^{\alpha-1} \exp\left\{-\Phi w^{\theta}\right\}}{(\alpha-1)!} = \frac{\theta \Phi^{\alpha} w^{\theta\alpha-1} \exp\left\{-\Phi w^{\theta}\right\}}{\Gamma(\alpha)}$$

(h) In perfect competition, the pdf for prices and costs for surviving producers are both given by the least cost pdf. For $\alpha = 1$:

$$f_1(w) \;=\; \theta \Phi w^{\theta-1} \exp\left\{-\Phi w^{\theta}\right\}$$

(i) The $(1-\sigma)$ moment of the least cost distribution is:

$$E\left[\left(W^{(1)}\right)^{1-\sigma}\right] \;=\; \int_0^\infty w^{1-\sigma} f_1(w)\,\mathrm{d}w$$

(j) Computing this moment for j[79]:

$$E\left[\left(W_j^{(1)}\right)^{1-\sigma}\right] \;=\; \int_0^\infty w^{1-\sigma} \theta \Phi_j w^{\theta-1} \exp\left\{-\Phi_j w^{\theta}\right\} \mathrm{d}w$$

Define $v = \Phi_j w^{\theta}$, therefore $dv = \Phi_j \theta w^{\theta-1} dw$ and $(v/\Phi)^{(1-\sigma)/\theta} = w^{1-\sigma}$. Changing the variable of integration and rearranging:

$$E\left[\left(W_j^{(1)}\right)^{1-\sigma}\right] \;=\; \Phi_j^{-(1-\sigma)/\theta} \int_0^\infty v^{(1-\sigma)/\theta} \exp\left\{-v\right\} \mathrm{d}v$$

By the definition of the gamma function with parameter: $\gamma = 1 + \frac{1-\sigma}{\theta}$, the integral

78 The gamma distribution describes the waiting time in a Poisson process until the α change is observed. And since in this case draws are made from a Pareto distribution, the parameter θ accounts for draws' variability.

79 Lemma 2 in Eaton and Kortum (2010).

is equal to $\Gamma\left[\frac{\theta+1-\sigma}{\theta}\right]$.

(k) The price index in j which is well defined under parameter restrictions $1 \leq \sigma < \theta+1$ is:

$$I_j^{1-\sigma} = \Phi_j^{-(1-\sigma)/\theta}\Gamma\left[\frac{\theta+1-\sigma}{\theta}\right]$$

Solving for Φ_j:

$$\Phi_j = I_j^{-\theta}\left(\Gamma\left[\frac{\theta+1-\sigma}{\theta}\right]\right)^{\frac{\theta}{1-\sigma}}$$

(l) Since the distribution of potential costs is invariant to trade costs, the $(1-\sigma)$ moment for the least-costs distribution over *potential* costs for goods produced in i and delivered to j is:

$$I_{ij}^{1-\sigma} = \Phi_i^{-(1-\sigma)/\theta}\Gamma\left[\frac{\theta+1-\sigma}{\theta}\right]$$

where $\Phi_i = T_i(c_i\tau_{ij})^{-\theta}$. Solving for Φ_i:

$$\Phi_i = I_{ij}^{-\theta}\left(\Gamma\left[\frac{\theta+1-\sigma}{\theta}\right]\right)^{\frac{\theta}{1-\sigma}}$$

(m) Replacing Φ_i and Φ_j by their values in (14), we get an expression for the trade share in terms of relative prices:

$$\frac{X_{ij}}{Y_j} = \left(\frac{I_{ij}}{I_j}\right)^{-\theta} \tag{15}$$

(n) The problem is that while the distribution of prices in destination j is observed, the potential distribution of costs from i to j is not. This is because all not-lowest cost producers are eliminated through trade selection. As shown by Eaton and Kortum (2002), *observed* imports from i to j adjusted for the trade share inherit the price distribution in the destination:

$$\begin{aligned}
\pi_{ij}G_j(p) &= \int_0^p \exp\left[-\sum_{s=1}^{N} T_s\left(c_s\tau_{sj}\right)^{-\theta}p^{\theta}\right]\theta T_i\left(c_i\tau_{ij}\right)^{-\theta}p^{\theta-1}\,\mathrm{d}p \\
&= \frac{T_i\left(c_i\tau_{ij}\right)^{-\theta}}{-\Phi_j}\int_0^p \exp\left[-\Phi_j p^{\theta}\right]\left[-\Phi_j\right]\theta p^{\theta-1}\,\mathrm{d}p \\
&= \pi_{ij}\left[1-\exp\left\{-\Phi_j p^{\theta}\right\}\right]
\end{aligned}$$

To estimate θ using information on observed prices, it is therefore necessary to devise an estimation procedure which uses destination price indices.

(o) Three remarks are in order before we move on to monopolistic competition.

- The correct weights for computing source and destination price indices are destination-specific since for products exported to j, i inherits j's pdf. The destination-specific price index should be computed over all goods in j, with j-weights.
- Price imputation for goods that i does not export to j using prices observed in j cannot be justified since it is only in the precise range that i effectively exports to j that i has j's price distribution. Nothing can be inferred about non-observed prices since out of the exported product range i has a price distribution which differs from j's. We cannot infer I_{ij} using prices observed in j, nor can we infer I_{ij} using prices observed in i.
- The data we use does not allow estimating the efficiency heterogeneity parameter in the Ricardian framework because observed prices of exports from i to j do not correspond to the least-cost distribution in source i across the goods' continuum given i technology parameters. Instead they mimick the least cost distribution in j given technology parameters in all sources and each source bilateral costs with j.

3. ***Price index computation with monopolistic competition in intrasectoral varieties***

If perfect competition (pc) were the true market structure, we would observe one price per delivered product in each destination. But in the data, there are several sources with heterogeneous prices for a given product. We therefore switch to a market structure with monopolistic competition (mc) which assumes that each firm's production technique is its private property and each firm's output is a specific variety within some sector k. Again, we are interested in computing the right hand side of (14) in terms of relative prices of source and destination countries.

(a) The number of varieties within each sector is infinite but countable. Each variety is indexed by its α in the ordered costs' distribution among production techniques available in sector k. Consumption choices at the intrasectoral level are described by a CES aggregator:

$$Q(k) = \left[\sum_{\alpha=1}^{\infty} Q^{(\alpha)}(k)^{\frac{\sigma'-1}{\sigma'}} \right]^{\frac{\sigma'}{\sigma'-1}}$$

with $\sigma' \geq \sigma > 1$ which are respectively the intra- and intersectoral substitution elasticities, and σ' finite.

(b) The share of expenditure on some intrasectoral variety is then a function of the relative price of the variety to the sectoral price index:

$$\frac{X^{(\alpha)}(k)}{X(k)} = \left[\frac{P^{(\alpha)}(k)}{P(k)} \right]^{1-\sigma'}$$

where the sectoral price index is:

$$P(k) = \left[\sum_{\alpha=1}^{\infty} P^{(\alpha)}(k)^{1-\sigma'} \right]^{\frac{1}{1-\sigma'}}$$

Since intersectoral consumption choices are still represented by a CES aggregator, the expenditure on a sectoral good is given by:

$$\frac{X(k)}{X} = \left[\frac{P(k)}{P} \right]^{1-\sigma}$$

where the overall price index is:

$$P = \left[\int_0^1 P(k)^{1-\sigma} \mathrm{d}k \right]^{\frac{1}{1-\sigma}}$$

(c) Eaton and Kortum (2010) show that under appropriate parameter restrictions, two set-ups are possible for monopolistic competition in order to get a well-defined price index.[80] First, the set-up in which there are no overhead costs and therefore all varieties are available : parameter restrictions are $1 < \sigma < \theta + 1 < \sigma'$ (in pc: $\sigma' \to \infty$). Second, the set-up in which there are overhead costs which restrict available varieties to the subset $\alpha = 1, \ldots, A(k)$ where $A(k) = \max\left\{\alpha : W^{(\alpha)}(k) \leq \overline{w}\right\}$. Parameter restrictions are $1 < \sigma < \theta + 1$ and $\sigma' = \sigma$ with an upper bound on costs $\overline{w}$.[81] In the rest of this appendix we work with the second set-up because the bounded number

[80] See Theorem 2 in Eaton and Kortum (2010).

[81] The true parameter restriction in Eaton and Kortum (2010) is $\sigma' \geq \sigma$ but in practice they work with $\sigma' = \sigma$.

of available varieties fits better the features of our data. It is shown that in this set-up the efficiency heterogeneity parameter θ cannot be estimated with our data while the substitutability parameter σ can be estimated.

BOUNDED NUMBER OF AVAILABLE VARIETIES

(a) The second set-up of monopolistic competition in Eaton and Kortum (2010) considers overhead costs which restrict the number of varieties available in each sector to the subset $\alpha = 1, ..., A(k)$ where $A(k) = \max\left\{\alpha : W^{(\alpha)}(k) \leq \overline{w}\right\}$. Further, $\sigma' = \sigma$ by assumption. In this set-up it is indifferent whether the price index is computed as the $(1-\sigma)$ moment of expected sectoral prices or as the $(1-\sigma)$ moment of goods' prices across the goods' continuum. An important characteristic of this model is that it is directly comparable to Chaney (2008) where bilateral fixed costs of entry into foreign markets are incorporated into the model, as well as to Helpman et al. (2008) who make the assumption of an upper bound on firm productivity to generate zero trade flows.[82]

(b) The aggregate price index is computed in two steps: first as a function of Φ and of the cost cut-off $\overline{w}$, then by deriving the expression for $\overline{w}$ as a function of Φ, the ratio of market size X to fixed entry costs E defined in terms of labor, and model parameters σ and θ.

(c) The sectoral price index is now given by the restricted set of varieties:

$$P(k)^{1-\sigma} = \sum_{\alpha=1}^{A(k)} \left(P^{(\alpha)}(k)\right)^{1-\sigma}$$

Forming the expectation across all sectors and using the result from profit maximization that price is a fixed mark-up over costs (see 3e. below):

$$\begin{aligned} E\left[P^{1-\sigma}(k)\right] &= E\left[\sum_{\alpha=1}^{A(k)} P^{(\alpha)}(k)^{1-\sigma}\right] \\ E\left[\sum_{\alpha=1}^{A(k)} \left(MW^{(\alpha)}(k)\right)^{1-\sigma}\right] &= M^{1-\sigma} E\left[\sum_{\alpha=1}^{A(k)} W^{(\alpha)}(k)^{1-\sigma}\right] \end{aligned}$$

(d) The conditional distribution of costs is given by $Pr[W^{\alpha} \leq w | W^{\alpha} \leq \overline{w}] = \left(\frac{w}{\overline{w}}\right)^{\theta}$, with density $\theta w^{\theta-1} \overline{w}^{-\theta}$, and the number of costs allowing production with unit cost less

[82] see Appendix (B) and (C).

than $\overline{w}$ is given by $\Phi\overline{w}^{\theta}$.[83] This allows rewriting the expectation formed over the sum of ordered costs in the range $[1,...,A(k)]$ as the product of the expected number of cost draws below $\overline{w}$ and the expected cost of each such draw:

$$\begin{aligned} E\left[P(k)^{1-\sigma}\right] &= M^{1-\sigma}\Phi\overline{w}^{\theta}\int_0^{\overline{w}} w^{1-\sigma}\theta w^{\theta-1}\overline{w}^{-\theta}\,\mathrm{d}w \\ &= M^{1-\sigma}\Phi\theta\int_0^{\overline{w}} w^{\theta-\sigma}\,\mathrm{d}w \\ &= M^{1-\sigma}\Phi\frac{\theta}{\theta-\sigma+1}\overline{w}^{\theta-\sigma+1} \end{aligned} \tag{16}$$

(e) The zero profit condition is used to express $\overline{w}$ as a function of Φ and X/E. Variable profit of some variety is $\left\{\Pi^{v}(w)=(p-w)X^{(\alpha)}(k)/p\right\}$. Profit maximization gives $p=Mw$ where $\{M=\sigma/(\sigma-1)\}$. Replacing in the expression for variable profit:

$$\Pi^{v}(w) = \frac{(M-1)w}{Mw}X^{(\alpha)}(k)=\frac{X^{(\alpha)}(k)}{\sigma}$$

Using the demand equation for the variety $\left\{X^{(\alpha)}(k)=\left(\frac{P^{(\alpha)}(k)}{P}\right)^{(1-\sigma)}X\right\}$ and expressing price in terms of the marginal cost:

$$\Pi^{v}(w) = \frac{X}{\sigma}\left(\frac{Mw}{P}\right)^{(1-\sigma)}$$

For entry to be profitable, the fixed entry cost has to be covered by variable profit: $\left\{E\leq\frac{X}{\sigma}\left(\frac{Mw}{P}\right)^{(1-\sigma)}\right\}$. Rearranging to solve for the cut-off cost $\overline{w}$ gives:

$$\overline{w} = \left(\frac{E\sigma}{X}\right)^{1/(1-\sigma)}\frac{P}{M}$$

Replacing $\overline{w}$ by its value in the expression for the price index:

$$P^{1-\sigma} = M^{1-\sigma}\frac{\Phi\theta}{\theta-(\sigma-1)}\left(\frac{E\sigma}{X}\right)^{\frac{\theta-(\sigma-1)}{1-\sigma}}P^{\theta-(\sigma-1)}M^{-(1-\sigma)-\theta}$$

Rearranging and simplifying:

$$P = \Phi^{-1/\theta}M\left(\frac{\theta}{\theta-(\sigma-1)}\right)^{-1/\theta}\left(\frac{X}{E\sigma}\right)^{\frac{-[\theta-(\sigma-1)]}{(\sigma-1)\theta}} \tag{17}$$

[83] For an individual good, the number of costs which deliver cost less than some upper bound w is distributed Poisson with parameter Φw^{θ}. Across the goods' continuum Φw^{θ} is the measure of goods which can be produced with cost less than w.

(f) In mc, just as in pc, the destination-specific price index $P_j = \left[\int_0^1 P_j(k)^{1-\sigma} \mathrm{d}k \right]^{1/(1-\sigma)}$ which is well defined under parameter restrictions $1 \leq \sigma = \sigma' < \theta + 1$ and upper bound on available varieties $\overline{w}$ can be used to estimate the underlying heterogeneity parameter θ. Using (17), the destination parameter $\Phi_j = \sum_{s=1}^{N} T_s(c_s \tau_{sj})^{-\theta}$ can be written as a function of the ideal price index across the surviving firms from all sources present in this destination:

$$\Phi_j \quad = \quad P_j^{-\theta} M^{\theta} \left(\frac{\theta - (\sigma - 1)}{\theta} \right) \left(\frac{E_j \sigma}{X_j} \right)^{\frac{\theta - (\sigma - 1)}{(\sigma - 1)}}$$

(g) Fixed entry costs are defined in terms of destination market labor: $E_j = c_j L_j$ where c_j is the labor cost in j and L_j is the number of mobilized labor units. This definition of entry costs means that we do not first ask whether the firm is present in the domestic market i and as a second step whether it is able to export to j, but rather we ask whether the firm from i can get in any market j which includes the domestic i market given its cost draw from the i-specific productivity distribution **and** the fixed entry cost in j which has to be paid by any firm wishing to operate in j. The measure of active sellers in j is defined across all firms active in j: $H_j = \Phi_j \overline{w_j}^{\theta}$. This is just the number of expected cost draws below the cost cut-off. Using 3e:

$$H_j = \Phi_j \left[\left(\frac{\theta - (\sigma - 1)}{\theta} \frac{X}{E\sigma} \right)^{1/\theta} \Phi^{-1/\theta} \right]^{\theta} \quad = \quad \frac{\theta - (\sigma - 1)}{\theta \sigma} \frac{X}{E}$$

(h) Under the assumption that there is no upper bound on productivity draws and that entry costs are market-specific and common to all firms, a subset of producers from each source i survives in all sectors k in each destination j. Trade-driven selection does not impede observing *source-specific* costs' distribution across the goods' continuum. This source-specific distribution is observed in the sectoral bilateral trade data. The price index for goods delivered from i to j is defined by:

$$P_{ij} \quad = \quad \left[\int_0^1 P_{ij}(k)^{1-\sigma} \mathrm{d}k \right]^{1/(1-\sigma)}$$

Using (16) this price index is:

$$P_{ij} \;=\; M\left(\frac{\Phi_{ij}\theta}{\theta-\sigma+1}\right)^{1/1-\sigma} \overline{w_j}^{\frac{\theta-\sigma+1}{(1-\sigma)}} \tag{18}$$

where $\Phi_{ij} = T_i(c_i\tau_{ij})^{-\theta}$ is the source-specific Poisson parameter for cost draws inclusive of bilateral trade costs while the cost cut-off is destination-specific. What matters for a firm from i to get into market j is its ability to overcome the fixed entry cost in j given its cost draw from the i-specific efficiency distribution.[84]

(i) Using (16), the landed price of i exports relatively to the overall price index in j is:

$$\frac{P_{ij}}{P_j} \;=\; \frac{M\left(\frac{\Phi_{ij}\theta}{\theta-\sigma+1}\right)^{1/1-\sigma} \overline{w_j}^{\frac{\theta-\sigma+1}{(1-\sigma)}}}{M\left(\frac{\Phi_{j}\theta}{\theta-\sigma+1}\right)^{1/1-\sigma} \overline{w_j}^{\frac{\theta-\sigma+1}{(1-\sigma)}}}$$

Simplifying and solving for the Φ-ratio:

$$\frac{\Phi_{ij}}{\Phi_j} = \left(\frac{P_{ij}}{P_j}\right)^{(1-\sigma)}$$

Replacing the Φ-ratio in the right hand side of the market share equation (14), we get back the CES demand equation:

$$\frac{X_{ij}}{Y_j} = \left(\frac{P_{ij}}{P_j}\right)^{(1-\sigma)} \tag{19}$$

Thus, observed landed prices of exports can be used to estimate the substitutability parameter σ, but not the underlying heterogeneity parameter θ.

(j) Several remarks are in order.

- In monopolistic competition with an upper bound on costs, data characteristics only allow estimating the substitutability parameter σ, just as in perfect competition. In perfect competition, the cost distribution in i could not be deduced from observed prices of i products sold in j because the surviving lowest-cost firms replicated the price distribution in j for the first-best cost across all possible sources exporting to j. In monopolistic competition this price distribution is observed, but the cost cut-off is defined in terms of the destination-specific

[84] In the next appendix, it will be shown that allowing for explicit fixed costs of exporting on top of entry costs in the domestic market will induce a selection margin among firms able to survive in i. But also in that model the fixed costs of exporting are paid in destination-specific labor units. This additional selection margin will not change the conclusion that from observed prices of exports from i to j only the substitutability parameter σ can be estimated.

Poisson parameter for cost draws. This means that the Φ parameter which carries the $-1/\theta$ exponent is the destination-specific parameter which cancels in the ratio P_{ij}/P_j. This leaves a market share equation which underlines the tight link between the relative Φ's of the source and destination and their realized price distributions: i's market share in j is equal to the relative realized price of its goods in j to the power $(1-\sigma)$ or alternatively to the ratio Φ_{ij}/Φ_j.[85] To estimate θ using observed prices it is necessary to devise an estimation procedure which works with destination-specific price indices.

- Destination-specific sectoral weights should be used in any type of aggregation across observed sectoral relative prices since market shares and composite good prices for any source are determined by the overall destination-specific price distribution.

- Under model assumptions some trade would be observed in every sector k between all pairs ij. There being no zero trade flows, the question of whether prices should be imputed would not arise, and both trade shares and composite good relative prices would be computed over *observed* trade values and product prices.[86]

- Zero trade flows are a prevalent feature of the data. This means that while we observe sectoral market shares of the source in all sectors, we do not have the full set of information on the realized price distribution of the source across the continuum of sectors. This has to be explicitly taken into account in the estimation procedure. The model provides two important keys for understanding the relationship between observed price indices and the underlying realized price distributions across the continuum of goods.

  First, unobserved prices are necessarily above the maximum price observed in the destination in any sector. This is because the cost cut-off $\overline{w}$ defines the maximum landed price which makes entry profitable in the destination *across the continuum of goods*. Second, unobserved prices are source-specific because

[85] In perfect competition the CES aggregator at the intersectoral level could be used to estimate σ using the fact that i's market share in j is given by the relative price of products in which i is lowest cost supplier to the price index over lowest cost draws across the goods' continuum in j.

[86] In the data zero trade flows are observed. To get this prediction, it is necessary to modify the model so that it generates zero trade flows. One way to do this is to assume an upper bound on productivity draws. We look at the implications of this modification for the estimation methodology in App.(C).

they are drawn from the source efficiency distribution: the relevant parameter is $\Phi_{ij}\overline{w_j}^{\theta}$ which gives the expected number of draws in i below the cost cut-off in the destination. These two features allow characterizing the relationship between the price index constructed from observed prices for any source i exporting to j and the underlying true price index across the goods' continuum. This characterization combined with the implications of the variance in the number of zeros across exporters to a given destination leads to the conclusion that the true underlying substitutability parameter is lower than the parameter estimated using observed prices.
Proceed as follows.

i. Use the expression for the price index in (18) to show that an increase in the cost cut-off leads to a decrease in the ideal price index due to the variety effect: as the cut-off increases, the number of zero trade flows decreases.

$$\begin{aligned}\frac{\partial P_{ij}}{\partial \overline{w}} &= M\left(\frac{\Phi_{ij}\theta}{\theta-\sigma+1}\right)^{1/1-\sigma}\frac{\theta-\sigma+1}{1-\sigma}\overline{w_j}^{\frac{\theta}{(1-\sigma)}} \\ &= \frac{1}{1-\sigma}M(\Phi_{ij}\theta)^{1/(1-\sigma)}(\theta-\sigma+1)^{-\sigma/(1-\sigma)}\overline{w_j}^{\theta/(1-\sigma)} \quad (20)\end{aligned}$$

where $\Phi_{ij} = T_i(c_i\tau_{ij})^{-\theta}$ is the source-specific Poisson parameter for cost draws inclusive of bilateral trade costs while the cost cut-off $\overline{w_j}$ is destination-specific. The first component is negative given $\sigma > 1$ while all other components are positive. The ideal price index decreases when the cost cut-off increases which means that as the cut-off increases, more firms from i are able to enter j market. Because we work with the goods' continuum, each additional firm corresponds to the elimination of a sectoral zero trade flow.

ii. Use the expression for the average price of goods from i to show that eliminating the variety effect we verify that the price index corrected for the variety effect increases in the cost cut-off. Equation (16) gives the expression for the expected price of goods exported by i across the goods' continuum. The measure of active sellers from i in j for a given cost cut-off is given by

$H_{ij} = \Phi_{ij}\overline{w_j}^{\theta}$. The expected average price is:

$$\begin{aligned}
\frac{1}{H_{ij}}E\left[P(k)_{ij}^{1-\sigma}\right] &= \frac{\Phi_{ij}\overline{w_j}^{\theta-\sigma+1}}{\Phi_{ij}\overline{w_j}^{\theta}}M^{1-\sigma}\left(\frac{\theta}{\theta-\sigma+1}\right) \\
&= \left(M\overline{w_j}\right)^{1-\sigma}\left(\frac{\theta}{\theta-\sigma+1}\right)
\end{aligned}$$

The price index corrected for the variety effect $P_{ij}^{COR} = \left\{H_{ij}^{-1}E\left[P(k)_{ij}^{1-\sigma}\right]\right\}^{1/(1-\sigma)}$ is increasing in $\overline{w_j}$. Marginal entrants from i which reduce the number of sectoral zeros are high cost relatively to firms from i already present in the market.

$$P_{ij}^{COR} = M\left(\frac{\theta}{\theta-\sigma+1}\right)^{1/(1-\sigma)}\overline{w_j}$$

iii. Consider the total differential of the price index P_{ij} assuming σ, θ invariant:

$$dP_{ij} = \frac{\partial P_{ij}}{\partial \Phi_{ij}}d\Phi_{ij} + \frac{\partial P_{ij}}{\partial \overline{w_j}}d\overline{w_j}$$

Consider an increase in the cost cut-off with Φ_{ij} constant. For any exporter i to j:

$$dP_{ij} = \left\{\frac{1}{1-\sigma}M(\Phi_{ij}\theta)^{1/(1-\sigma)}(\theta-\sigma+1)^{-\sigma/(1-\sigma)}\overline{w_j}^{\theta/(1-\sigma)}\right\}d\overline{w_j}$$

iv. Consider two exporters i and s to some destination j. For a given small change in the cost cut-off, the relative change in their price indices is given by:

$$\frac{dP_{ij}}{dP_{sj}} = \frac{\partial P_{ij}/\partial \overline{w_j}}{\partial P_{sj}/\partial \overline{w_j}} = \left(\frac{\Phi_{ij}}{\Phi_{sj}}\right)^{1/1-\sigma}$$

Given $\sigma > 1$, $\frac{dP_{ij}}{dP_{sj}} < 1$ if and only if $\frac{\Phi_{ij}}{\Phi_{sj}} > 1$, eg if $\Phi_{ij} > \Phi_{sj}$. We conclude that for a given increase in the cost cut-off, the number of zero trade flows is reduced quicker for the exporter with a higher Φ[87] **and** his price index in j increases by less than the price index of the source with a smaller Φ.

v. Consider the relative market share of i and s in j.

$$\frac{X_{ij}}{X_{sj}} = \frac{\Phi_{ij}}{\Phi_{sj}}$$

[87] This comes from the definition of the mass of entrants $H_{ij} = \Phi_{ij}\overline{w}^{\theta}$. Differentiating with respect to the cost cut-off, $dH_{ij}/dH_{sj} = \Phi_{ij}/\Phi_{sj}$.

It is straightforward that the source with the higher Φ has higher market share in the destination.

vi. Combining all of the previous results:

- the relatively high Φ source will have a higher market share and a lower price index in destination
- for a given cost cut-off, the relatively high Φ source will have a smaller share of sectoral zero trade flows
- for a given increase in the cost cut-off, the relatively high Φ source will have a higher number of additional entrants (eg a stronger reduction in sectoral zero trade flows) while its price index will increase by less.

It follows that for a given pair of exporters with $X_{ij}/X_{sj} > 1$, the observed relative price index P_{ij}/P_{sj} is higher than the true underlying realized price distribution in the two sources across the continuum of sectors: $P_{ij}/P_{sj} >> P^R_{ij}/P^R_{sj}$. The observed price index is underestimated by more for small exporters. Generalizing to N exporters: for a given variance in market shares, the variance in unobserved true realized price distributions is higher than the variance in observed prices because of the strong negative correlation in the number of lacking prices and market shares. Another way to see that the price index is underestimated by more for small exporters is to note that unobserved prices in any sector for any source are higher than the maximum observed price in the destination (implication 1 of the model). As the small exporter has a greater share of unobserved prices, its true underlying price index is necessarily higher than the true underlying price index of a big exporter.[88]

vii. The substitutability parameter estimated on observed price indices and market shares ($\widetilde{\sigma}$) is *overestimated* relatively to the true underlying substitutability parameter (σ). This is because the variation in the number of unobserved prices across exporters entails that a given percentage change

[88] We verify that in the data smaller market share is associated with a greater number of sectoral zero trade flows, as predicted by the model. We find that overtime, the strength of the negative correlation between the number of zeros and market shares is progressively reduced. In terms of the model, this means that the variance in Φ's is reduced overtime. This is consistent with the finding by Levchenko and Zhang (2011) that there is a reduction in the strength of comparative advantage across countries over the period 1962-2009.

in market share is linked to a higher underlying (relatively to observed) percentage change in prices.

$$
\begin{aligned}
(\widetilde{\sigma}-1) &= \ln\left[\frac{X_{ij}/X_{sj}}{P_{sj}/P_{ij}}\right] \\
&\geq \\
(\sigma-1) &= \ln\left[\frac{X_{ij}/X_{sj}}{P^{R}_{sj}/P^{R}_{ij}}\right] \qquad (21)
\end{aligned}
$$

B The gravity equation in the Melitz-Chaney framework

This appendix goes over the main features of the Melitz-Chaney microfoundation of the gravity structure of trade. First, the Chaney (2008) model is presented and the computation of the price index is recalled. It is shown that it is not possible to estimate the degree of efficiency variability across firms, using data on observed cif prices of bilateral exports. Only the substitutability parameter σ can be estimated on our data. The appendix then recalls that if bilateral fixed costs of trade are assumed to be distance-dependent, the distance elasticity estimated in gravity equations in the context of the Chaney framework would be inversely dependent on σ. Our finding that the substitutability parameter has increased in 1962-2009 would only deepen the distance puzzle. If however it is assumed that fixed costs are not distance-dependent, then the trade elasticity in the Chaney framework would correspond to the efficiency heterogeneity parameter γ. Given that to estimate γ for aggregate trade it is necessary to make the assumption of a common efficiency variability parameter across sectors so that $\gamma_k = \gamma$, this appendix concludes that given the structural proximity of this model to Eaton and Kortum (2010), γ can be estimated using the methodology suggested in the first appendix to estimate θ.[89] This methodology is based on destination-specific price indices.

1. Each source i produces a homogeneous good indexed $k = 0$ with CRS technology. This good is freely traded. This pins down relative wages: the price of this good is normalized to 1, and the wage in country i, c_i, is the number of units of the numeraire good produced with one unit of labor. Thus, labor efficiency may vary across countries.

2. There is a finite number of sectors $k = 1, ..., K$ which each consist of a continuum of differentiated varieties. A firm, characterized by efficiency $z_{k,\alpha}$ drawn from a productivity distribution assumed common across countries but sector-specific, produces some variety $\alpha(k)$, paying a constant firm-specific marginal cost $w_i(k) = c_i/z^{(\alpha)}(k)$.

3. To enter any market, the firm pays per unit pair- and sector-specific variable trade costs $\tau_{ij}(k)$ as well as sector-specific fixed costs which are assumed to be pair-specific $f_{ij}(k)$.

4. Consumer preferences are assumed well-represented by a Cobb-Douglas utility function

[89] This is because we seek to estimate the trade elasticity parameter for aggregate trade. Crozet and Koenig (2010) show how the panel dimension of firm-level exports' data can be used to identify sector-specific γ_k, σ_k, and ρ_k. Non-trade data such as firm-level sales or employment has been used to estimate Pareto distribution parameters at the sectoral or economy-wide level in Luttmer (2007); Di Giovanni et al. (2011); Chatterjee and Rossi-Hansberg (2012).

at the intersectoral level and CES at the intrasectoral level. The substitutability parameter σ_k is sector-specific. Varieties within any sector are indexed by α.

$$U = q_0^{\mu_0} \prod_{k=1}^{K} \left[\int_0^1 q_k(\alpha)^{\frac{\sigma_k - 1}{\sigma_k}} \mathrm{d}\alpha \right]^{\mu_k \frac{\sigma_k}{\sigma_k - 1}}$$

5. Firms draw their productivity from a Pareto distribution with support $[1, \infty)$, and it is assumed that the shape parameter γ_k of the Pareto is common across countries but sector-specific.[90]

$$\Pr(Z \leq z) = 1 - z^{-\gamma_k}$$

6. As product differentiation is assumed costless, each firm optimally chooses to produce a unique variety. Varieties within each sector can be ranked by the efficiency $z^{(\alpha)}(k)$ of the producing firm. Given CES demand, the price of any variety is a constant mark-up charged over cif production costs:

$$p_{ij}^{(\alpha)}(k) = \frac{\sigma_k}{\sigma_k - 1} \left[c_i \tau_{ij}(k) / z^{(\alpha)}(k) \right]$$

7. Given CD intersectoral demand, expenditure in country j on goods in sector k is given by $\mu_k Y_j$. Using the properties of the CES aggregator at the intrasectoral level, demand for the firm-specific variety is:

$$x_{ij}^{(\alpha)}(k) = \left[\frac{p_{ij}^{(\alpha)}(k)}{P_j(k)} \right]^{1-\sigma_k} \mu_k Y_j$$

where $P_j(k)$ is the ideal price index in country j in sector k.

8. Define the firm-specific production cost inclusive of variable trade costs $v_{ij}^{(\alpha)}(k) = c_i \tau_{ij}(k) / z^{(\alpha)}(k)$. Using the zero profit condition for firms in sector k in source i exporting to destination j, we can express the cut-off cost $\overline{v_{ij}(k)}$ as a function of the sectoral ideal price index $P_j(k)$ and the fixed bilateral trade cost $f_{ij}(k)$. Net profits of a firm are given by:

$$\pi_{ij}^{(\alpha)}(k) = \left[p_{ij}^{(\alpha)}(k) - v_{ij}^{(\alpha)}(k) \right] q_{ij}^{(\alpha)}(k) - f_{ij}(k)$$

where $q_{ij}^{(\alpha)}(k) = x_{ij}^{(\alpha)}(k) / p_{ij}^{(\alpha)}(k)$. Defining the mark-up $M = \sigma_k / (\sigma_k - 1)$, replacing

[90] Parameter restrictions are $\sigma_k < \gamma_k + 1$.

$q_{ij}^{(\alpha)}(k)$ and $p_{ij}^{(\alpha)}(k)$ by their values, and setting $\pi_{ij}^{(\alpha)}(k) = 0$ for the cut-off firm:

$$f_{ij}(k) = \left(\frac{M\overline{v_{ij}(k)}}{P_j(k)}\right)^{1-\sigma_k} \frac{\mu_k Y_j}{\sigma_k}$$

Therefore, the cut-off cost for entering market j for firm from i in sector k is:

$$\overline{v_{ij}(k)} = \left[\frac{\mu_k Y_j}{f_{ij}(k)\sigma_k}\right]^{1/(\sigma_k-1)} \frac{P_j(k)}{M}$$

9. The sectoral ideal price index $P_j(k)$ is given by the summation across all countries of the firms with a cost below the cut-off:

$$P_j(k) = \left[\sum_{i=1}^{N} c_i L_i \int_0^{\overline{v_{ij}(k)}} \left[M v_{ij}(k)\right]^{1-\sigma_k} \mathrm{d}v_{ij}(k)\right]^{1/(1-\sigma_k)}$$

where $c_i L_i$ is total labor income in source i which is proportional to the potential mass of entrants in any sector k in i.

10. Using the assumption that the productivity distribution is sector but not country-specific, defining the cut-off productivity draw $\overline{z_{ij}(k)}$ as a function of $\overline{v_{ij}(k)}$, change the variable of integration to z to get:

$$P_j(k) = \left[\sum_{i=1}^{N} c_i L_i \int_{\overline{z_{ij}(k)}}^{\infty} \left[M c_i \tau_{ij}(k)\right]^{1-\sigma_k} \gamma_k z^{\sigma_k - 1} z^{-\gamma_k - 1} \mathrm{d}z\right]^{1/(1-\sigma_k)}$$

Rearranging and solving for the integral:

$$P_j(k) = M\left(\frac{-\gamma_k}{\sigma_k - (\gamma_k + 1)}\right)^{1/(1-\sigma_k)} \left[\sum_{i=1}^{N} c_i L_i \left(c_i \tau_{ij}(k)\right)^{1-\sigma_k} (\overline{z_{ij}(k)})^{\sigma_k - (\gamma_k+1)}\right]^{1/(1-\sigma_k)}$$

11. Working with the summation where we replace $\overline{z_{ij}(k)}$ by its value:

$$\sum_{i=1}^{N} c_i L_i \left(c_i \tau_{ij}(k)\right)^{1-\sigma_k} (c_i \tau_{ij}(k)/\overline{v_{ij}(k)})^{\sigma_k - 1 - \gamma_k} = \sum_{i=1}^{N} c_i L_i \left(c_i \tau_{ij}(k)\right)^{-\gamma_k} \left(\overline{v_{ij}(k)}\right)^{\gamma_k - (\sigma_k - 1)}$$

Defining $T_i = c_i L_i$, replacing $\overline{v_{ij}(k)}$ by its value, we get:

$$\sum_{i=1}^{N} T_i \left(c_i \tau_{ij}(k)\right)^{-\gamma_k} \left\{\left[\frac{\mu_k Y_j}{f_{ij}(k)\sigma_k}\right]^{1/(\sigma_k-1)} \frac{P_j(k)}{M}\right\}^{\gamma_k - (\sigma_k - 1)}$$

Rearranging gives:

$$\left[\frac{\mu_k Y_j}{\sigma_k}\right]^{\frac{\gamma_k-(\sigma_k-1)}{(\sigma_k-1)}}\left[\frac{P_j(k)}{M}\right]^{\gamma_k-(\sigma_k-1)}\sum_{i=1}^{N}T_i\left(c_i\tau_{ij}(k)\right)^{-\gamma_k}\left[f_{ij}(k)\right]^{1-\frac{\gamma_k}{(\sigma_k-1)}}$$

12. Define $\Psi_j(k)=\sum_{i=1}^{N}T_i\left(c_i\tau_{ij}(k)\right)^{-\gamma_k}\left[f_{ij}(k)\right]^{1-\frac{\gamma_k}{(\sigma_k-1)}}$. Replacing in the expression for the sectoral price index:

$$P_j(k)^{1-\sigma_k} = M^{1-\sigma_k}\frac{\gamma_k}{\gamma_k-(\sigma_k-1)}\left[\frac{\mu_k Y_j}{\sigma_k}\right]^{\frac{\gamma_k-(\sigma_k-1)}{(\sigma_k-1)}}\left[\frac{P_j(k)}{M}\right]^{\gamma_k-(\sigma_k-1)}\Psi_j(k)$$

13. Solving for the sectoral price index in j:

$$P_j(k)^{-\gamma_k} = M^{-\gamma_k}\frac{\gamma_k}{\gamma_k-(\sigma_k-1)}\left[\frac{\mu_k Y_j}{\sigma_k}\right]^{\frac{\gamma_k-(\sigma_k-1)}{(\sigma_k-1)}}\Psi_j(k)$$

14. Using the definition of exports for an individual firm, and the proportionality assumption on the mass of potential entrants in any country i, derive the expression for total bilateral sectoral exports.

$$X_{ij}(k) = c_i L_i\int_{\overline{z_{ij}(k)}}^{\infty}x_{ij}^{(\alpha)}(k)\gamma_k z^{-(\gamma_k+1)}\,\mathrm{d}z$$

Using $T_i=c_iL_i$ and replacing $x_{ij}^{(\alpha)}(k)$ by its value:

$$X_{ij}(k) = T_i\mu_k Y_j\int_{\overline{z_{ij}(k)}}^{\infty}\left(\frac{p_{ij}^{(\alpha)}(k)}{P_j(k)}\right)^{1-\sigma_k}\gamma_k z^{-(\gamma_k+1)}\,\mathrm{d}z$$

Replacing $p_{ij}^{(\alpha)}(k)$ by its value:

$$X_{ij}(k) = T_i\mu_k Y_j\gamma\left(\frac{Mc_i\tau_{ij}(k)}{P_j(k)}\right)^{1-\sigma_k}\int_{\overline{z_{ij}(k)}}^{\infty}z^{\sigma_k-1}z^{-(\gamma_k+1)}\,\mathrm{d}z$$

Solving for the integral:

$$X_{ij}(k) = T_i\frac{\overline{z_{ij}(k)}^{\sigma_k-1-\gamma_k}}{\gamma_k-(\sigma_k-1)}\mu_k Y_j\gamma_k\left(\frac{Mc_i\tau_{ij}(k)}{P_j(k)}\right)^{1-\sigma_k}$$

Using the definition of the cut-off productivity,

$$\overline{z_{ij}(k)} = Mc_i\tau_{ij}(k)P_j(k)^{-1}(\mu_k Y_j/\sigma_k)^{1/(1-\sigma_k)}f_{ij}(k)^{-1/(1-\sigma_k)}$$

total bilateral exports are given by:

$$X_{ij}(k) = T_i\left(\frac{Mc_i\tau_{ij}(k)}{P_j(k)}\right)^{\sigma_k-1-\gamma_k}\left(\frac{\mu_k Y_j}{\sigma_k}\right)^{\frac{\sigma_k-1-\gamma_k}{1-\sigma_k}} f_{ij}(k)^{\frac{\gamma_k-(\sigma_k-1)}{1-\sigma_k}}\frac{\gamma_k}{\gamma_k-(\sigma_k-1)}\mu_k Y_j\left(\frac{Mc_i\tau_{ij}(k)}{P_j(k)}\right)^{1-\sigma_k}$$

Simplifying:

$$X_{ij}(k) = T_i\frac{\gamma_k}{\gamma_k-(\sigma_k-1)}\left[\frac{Mc_i\tau_{ij}(k)}{P_j(k)}\right]^{-\gamma_k}\left[\mu_k Y_j\right]^{-\gamma_k/(1-\sigma_k)}\left[f_{ij}(k)\sigma_k\right]^{\frac{\gamma_k-(\sigma_k-1)}{(1-\sigma_k)}}$$

Replacing $P_j(k)^{-\gamma_k}$ by its value and simplifying:

$$X_{ij}(k) = T_i\mu_k Y_j\Psi_j(k)^{-1}\left[c_i\tau_{ij}(k)\right]^{-\gamma_k}\left[f_{ij}(k)\right]^{\frac{\gamma_k-(\sigma_k-1)}{(1-\sigma_k)}}$$

15. Define $\Psi_{ij}(k) = T_i\left[c_i\tau_{ij}(k)\right]^{-\gamma_k}\left[f_{ij}(k)\right]^{\frac{\gamma_k-(\sigma_k-1)}{(1-\sigma_k)}}$. The market share of country i in sector k in j is:

$$\frac{X_{ij}(k)}{\mu_k Y_j} = \frac{\Psi_{ij}(k)}{\Psi_j(k)}$$

16. Having obtained a market share equation at the sectoral level in terms of the Ψ-ratio, we want to express this ratio in terms of observed price distributions in our data. Having solved for the sectoral price index in the destination $P_j(k)$ as a function of $\Psi_j(k)$, we just need to solve for the ideal price index $P_{ij}(k)$ across varieties exported by i to j in sector k:

$$P_{ij}(k)^{1-\sigma_k} = c_i L_i\int_{\overline{z_{ij}}(k)}^{\infty} p_{ij}^{(\alpha)}(k)^{1-\sigma_k}\,\mathrm{d}z$$

Replacing $p_{ij}^{(\alpha)}(k)$ by its value and solving for the integral:

$$P_{ij}(k)^{1-\sigma_k} = \frac{\gamma_k}{\gamma_k-(\sigma_k-1)}T_i\left(Mc_i\tau_{ij}(k)\right)^{1-\sigma_k}\overline{z_{ij}(k)}^{\sigma_k-\gamma_k-1}$$

Replacing the cut-off productivity by its value:

$$P_{ij}(k)^{1-\sigma_k} = \frac{\gamma_k}{\gamma_k-(\sigma_k-1)}T_i\left[Mc_i\tau_{ij}(k)\right]^{-\gamma_k}P_j(k)^{\gamma_k-(\sigma_k-1)}\left[\frac{\mu_k Y_j}{f_{ij}(k)\sigma_k}\right]^{(\sigma_k-\gamma_k-1)/(1-\sigma_k)}$$

Replacing $P_j(k)^{\gamma_k}$ by its value, rearranging, and simplifying:

$$\left(\frac{P_{ij}(k)}{P_j(k)}\right)^{1-\sigma_k} = \frac{\Psi_{ij}(k)}{Psi_j(k)}$$

Thus, the observed price distribution from i to j can be used in the sectoral market share equation to get the substitutability parameter σ_k, but not the efficiency variability parameter γ_k.

17. Consider a reformulation of the utility function of consumers as Cobb-Douglas between the homogeneous good and the bundle constituted by the differentiated goods, and assume a two-tier CES function for the differentiated goods' bundle:

$$U = q_0^{1-\mu}\left[\sum_{k=1}^{K} Q(k)^{\sigma-1/\sigma}\right]^{\mu(\sigma/\sigma-1)}$$

where sectoral consumption across all varieties α is $Q(k) = \left[\int_0^1 q(\alpha)^{\frac{\sigma_k-1}{\sigma_k}}\,\mathrm{d}\alpha\right]^{\frac{\sigma_k}{\sigma_k-1}}$.

18. The parameter we seek to estimate is the trade elasticity at the level of aggregate trade in differentiated goods. This amounts to assuming away sector-specific dimensions: $\gamma_k = \gamma$ and $\sigma_k = \sigma$. This allows redefining the bundle constituted by differentiated goods as a continuum of differentiated varieties.

19. Suppressing sectoral subscripts, we get the following expressions for aggregate bilateral exports, the price index in the destination, and the Ψ-parameters:

$$X_{ij} = \frac{T_i\left[c_i\tau_{ij}\right]^{-\gamma}\left[f_{ij}\right]^{\frac{\gamma-(\sigma-1)}{(1-\sigma)}}}{\sum_{s=1}^{N} T_s\left(c_s\tau_{sj}\right)^{-\gamma}\left[f_{sj}\right]^{\frac{\gamma-(\sigma-1)}{(1-\sigma)}}}\mu Y_j$$

$$P_j = M\left(\frac{\gamma}{\gamma-(\sigma-1)}\right)^{-1/\gamma}\left[\frac{\mu Y_j}{\sigma}\right]^{\frac{\gamma-(\sigma-1)}{\gamma(1-\sigma)}}\Psi_j^{-1/\gamma}$$

$$\Psi_j = \sum_{s=1}^{N} T_s\left(c_s\tau_{sj}\right)^{-\gamma}\left[f_{sj}\right]^{\frac{\gamma-(\sigma-1)}{(1-\sigma)}}$$

$$\Psi_{ij} = T_i\left[c_i\tau_{ij}\right]^{-\gamma}\left[f_{ij}\right]^{\frac{\gamma-(\sigma-1)}{(1-\sigma)}}$$

These expressions are structurally similar to the ones obtained in the monopolistic set-up of the Eaton and Kortum model with a bounded number of available varieties because firm entry in foreign markets is determined by destination-specific characteristics. The difference resides in that Ψ_j in the Chaney set-up includes a pair-specific bilateral fixed cost while in the Eaton and Kortum set-up this fixed cost is invariant by source and therefore not in Φ_j.

20. In the Chaney set-up, the functional form of the bilateral fixed entry cost f_{ij} is not defined while in the Eaton and Kortum set-up the overhead cost is defined in terms of the number and cost of labor units specific to the destination: $f_{ij} = f_j = c_j F_j$ where F_j is the number of workers. Chaney (2008) provides some empirical evidence of distance-dependent fixed costs of trade. A simple way of making these fixed costs source- and distance-dependent would be to write: $f_{ij} = c_j F_{ij} = c_j dist^{\rho_1}$. The functional form for variable trade costs would still be defined as $\tau_{ij} = dist^{\rho_0}$. Bilateral exports are then given by:

$$X_{ij} = \frac{T_i c_i{}^{-\gamma} dist_{ij}{}^{-(\gamma\rho_0+\rho_1\gamma/(\sigma-1)-\rho_1)}}{\sum_{s=1}^{N} T_s c_s{}^{-\gamma} dist_{sj}{}^{-(\gamma\rho_0+\rho_1\gamma/(\sigma-1)-\rho_1)}} \mu Y_j$$

In this case, the distance elasticity estimated in gravity equations is a combination of σ, γ, and ρ parameters. Given that the substitutability parameter *dampens* the sensitivity of trade flows to trade barriers, our finding of an increasing σ only **deepens** the distance puzzle. Given our results for σ, an increase in the distance coefficient would have to be explained by changes in γ, ρ_0, and ρ_1, with at least one of these parameters strictly increasing in 1962-2009.

21. If however it is assumed that there is no intuitive reason for considering that the number of labor units required to export to some destination depends on the distance between the source and the destination, the model would become equivalent to the Eaton and Kortum set-up in monopolistic competition with a bounded number of available varieties due to an overhead cost defined as $f_{ij} = f_j = c_j F_j$. The distance coefficient would be defined as $\gamma\rho$. In this case, our results on the evolution of the substitutability would neither deepen, nor explain the distance puzzle.

C The Armington assumption, the Anderson and van Wincoop (2003) derivation of the gravity equation, and the generalization by Helpman et al. (2008) to the heterogeneous firms' framework.

[TO BE COMPLETED]

D The impact of zero trade flows

This appendix contains a discussion of the sensitivity of our results to constructing the relative price of the source-specific composite good at the highest disaggregation level. It is argued that the treatment of ztf is key, and that both procedures corroborate the finding that the evolution of the estimated substitutability parameter provides a lower bound on the increase in the underlying substitutability.

First, we present the estimates obtained when the hypothesis that unobserved prices are equal to the weighted average relative price constructed across observed prices is applied at the highest disaggregation level. Working at the SITC 4-digit level, the relative price of the composite good is constructed across observed prices using destination-specific expenditure weights at the 4-digit level. We make the assumption that this weighted average relative price corresponds to the relative price for those SITC 4-digit categories in which the source has zero trade flows. Figure 10 shows that in this case the evolution of the substitutability parameter is best described as flat in 1962-2009.

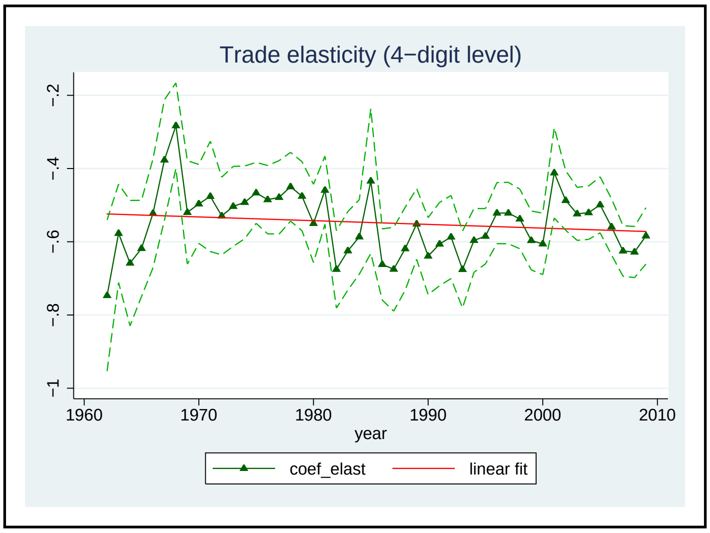


Figure 10: Estimated $(1-\widetilde{\sigma})$, relative price of composite good from 4-digit level)

However, if observed relative prices at the 4-digit level constitute a better proxy for unobserved prices at the 3-digit level within the same product category than of unobserved prices for a random 4-digit category, then constructing the 3-digit relative price as a weighted average of observed 4-digit relative prices, and reapplying this procedure at each level of aggregation until obtaining the relative price for the composite good, may allow capturing unobserved prices within each sector more accurately by reducing the extent of measurement error. As shown in Fig.8 in IV.3, when the relative price of the composite good is constructed hierarchically, with the final aggregation done at the highest possible level, i.e. at the SITC 1-digit level, the substitutability parameter is found to increase by 29%.

Table 4: **Predicted share of ztf for exporters with different market share, 1-digit level**

year	ms=0.02%	ms=1%	ms=10%	ms=28.7%
1962	0.60	0.35	0.26	0.22
2009	0.38	0.24	0.19	0.17

Columns (1) and (4) correspond to the mean and to 2 st. deviations above the mean in the distribution of log market share. Columns (2) and (3) correspond to the mean and to 2 st. deviations above the mean in the distribution of market share.

It is argued that the estimation conducted following hierarchical aggregation better captures the evolution of the true underlying elasticity. First, the higher the aggregation level, and the lower is the share of ztf as a fraction of sectoral bilateral trade flows as table 4 illustrates.[91]

Second, as shown in tables 5 and 4, the reduction in the gap between big and small exporters is more pronounced at the higher aggregation level.[92] Taking the example of an exporter with a 1% market share, the initial share of ztf is estimated to be .35 at the SITC 1-digit level, and it is reduced to .24 by 2009, e.g. a decrease of 11 percentage points while for an exporter with a 10% market share, the share of ztf is reduced from .26 to .19 in 1962-2009, e.g. a decrease of 7 percentage points. At the highest disaggregation level, these numbers are .80 and .71 for 1% ms, and .72 and .65 for 10% ms. It follows that not only is the share of zero trade flows substantially reduced for all exporters in this estimation, but the reduction in the share of ztf is more pronounced for small relatively to big exporters.

[91] At SITC 1-digit level, ztf correspond to 53% of all trade flows in 1962, and this is reduced to 46% in 2009.

[92] Taking column (4), the full year effect is given by $(-0.0065 + 0.0006 * ln(ms))$. The year effect is stronger for exporters of all types at the higher aggregation level. And the gap between small and big exporters in terms of the speed of the reduction in ztf is more pronounced at the higher aggregation level.

Table 5: **Proportion of zero trade flows as a function of market share**

depvar: Share of ZTF				
	(1)	(2)	(3)	(4)
ms	-0.1222***	-1.1178***	-0.1348***	-1.3800***
	(0.0002)	(0.0268)	(0.0002)	(0.026)
year	-0.0103***	-0.0055***	-0.0125***	-0.0065***
	(0.0000)	(0.0001)	(0.0000)	(0.000)
$ms * year$		0.0005***		0.0006***
		(0.0000)		(0.000)
constant	18.6511***	9.0486***	23.2635***	11.2598***
	(0.0832)	(0.2930)	(0.0842)	(0.283)
Destination FE	NO	NO	YES	YES
Observations	657001	657001	657001	657001

Notes: The share of ZTF is computed at the SITC 1-digit level. The estimation is conducted in PPML in order to include observations where ztf=0. The log of the market share is used in the estimation. Destination fixed effects are included in (3) and (4). Robust standard errors are in parentheses. *** p<0.01.

It cannot be argued in either estimation that the bias induced by the ztf treatment is time-invariant because both the level and the gap in the share of ztf between big and small exporters vary overtime. However, as shown by tables 3 and 4, it is at the higher aggregation level that ztf are less prevalent for all exporters which means that the extent of the bias due to ztf should be reduced. Further, it is at the higher aggregation level that the overestimation bias, due to the fact that the relative price is underestimated by more for small than for big exporters, is most strongly reduced overtime. The assumption that the relative price of the composite good is underestimated by a constant factor for all exporters is therefore most plausible by the end of the sample period in the estimation conducted at the highest aggregation level.

We expect that the reduction in the bias inherent to the estimation procedure, e.g. the reduction in the gap between the observed price distribution and the underlying true price dispersion, should result in a starker evolution of the estimated parameter if the true underlying trade elasticity increases in absolute value. We also expect that the level of the elasticity should be reduced in absolute value in the estimation in which the extent of the bias due to ztf treatment is reduced.

Figure 11 illustrates that the extent of the bias due to ztf treatment seems to be effectively

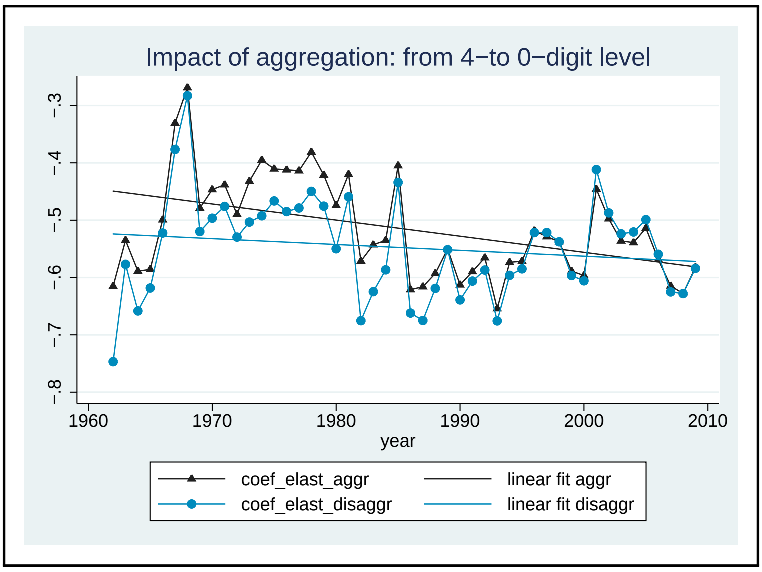


Figure 11: Estimated $(1-\widetilde{\sigma})$, relative price of composite good at 4- and 0-digit level

reduced in the estimation conducted after hierarchical price aggregation. First, the parameter estimated from prices aggregated directly at the 4-digit level is higher in absolute value than the parameter estimated using hierarchical aggregation in the construction of the relative price for the composite good, e.g. the true underlying trade elasticity is overestimated by less in all years when the estimation is conducted at a higher aggregation level. Second, even though in the beginning of the sample there is greater dispersion in the share of ztf between big and small exporters for estimation conducted at the higher aggregation level, the absolute value of the estimated parameter is lower which means that the bias due to the prevalence of ztf seems to dominate the bias due to the variation in ztf share across exporters. Third, variation in ztf share across exporters is similar for both procedures by the end of the sample while the share of ztf is less prevalent for all exporters at the higher aggregation level. The estimation procedure conducted at the higher aggregation level is preferred because the bias linked to ztf treatment should be reduced relatively to the estimation procedure conducted at higher disaggregation levels.

Given that the overestimation bias linked to ztf treatment is not eliminated in the estimation conducted at the higher aggregation level while the evolution of the trade elasticity parameter is more pronounced, we conclude that the evolution of $\widetilde{\sigma}$ in the estimation procedure conducted after hierarchical aggregation provides *a lower bound* on the increase in the absolute value of the true substitutability parameter in 1962-2009.

E Robustness checks for instrumented prices

This appendix shows that our results on the evolution of the substitutability parameter in the estimation with instrumented prices are robust to increasing the number of periods s for which the growth rate in the relative price of the exported composite good is predicted on the evolution of relative domestic prices.

Fig.(12) shows that, as we increase s from 3 to 5, the evolution of the estimated substitutability parameter becomes steeper, and the increase in the level of the parameter becomes more stable across the years. Thus, for $s = 5$, the parameter is found to increase by 16% in 1967-2009 while its level increases by about 20-30% relatively to the estimate obtained with observed relative prices.

Fig.(13) illustrates a similar finding for $s = 10$ relatively to $s = 5$. For $s = 10$, the substitu-

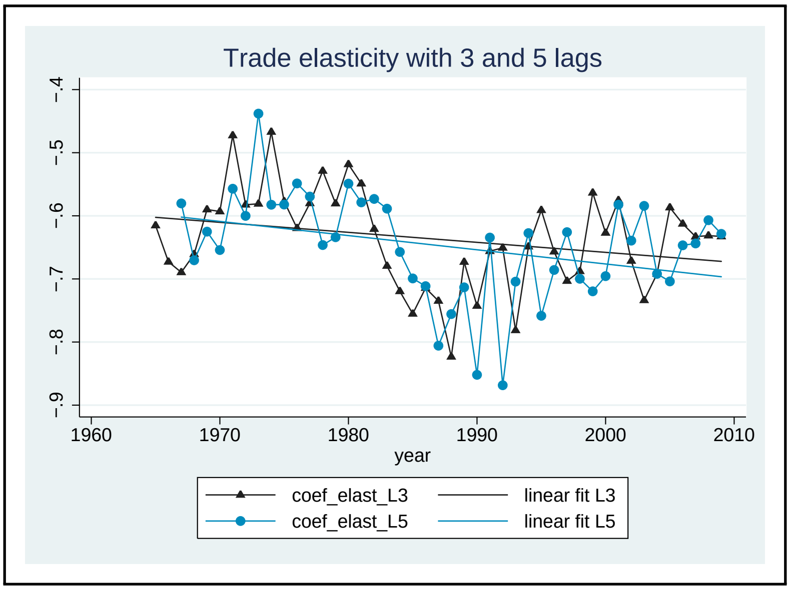


Figure 12: Estimated $(1-\widetilde{\sigma})$, instrumented relative price of composite good

tion elasticity increases by 21% in 1972-2009, and the absolute value of the parameter increases by 25-26% relatively to the estimate obtained with non-instrumented prices. Thus, in 2009 the instrumented estimate $(\widetilde{\sigma}-1)$ is .73 which is 26% higher than the non-instrumented estimate of .58.

The evolution of the substitutability parameter becomes steeper as the number of periods in which we predict the evolution of exports' prices increases. Partly, this result is due to the fact that the increase is steeper in 1970-2009. But as shown in fig.(13), this is only part of the explanation: for 1972-2009, the evolution is steeper for the parameter estimated with $s = 10$ than

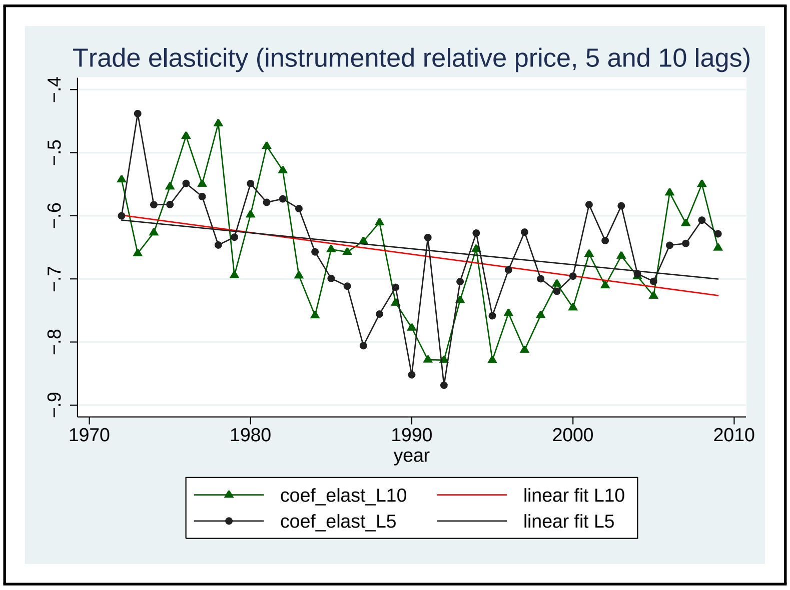


Figure 13: Estimated $(1-\widetilde{\sigma})$, instrumented relative price of composite good with $s = 5$.

F List of included FTAs

In squared brackets: *[years in which the FTA appears in the database, where 'F' stands for 'full' and 'S' for 'superbalanced' sample]*

N.B.: The GATT/WTO membership variable is present in the database in all years.

EC (European Communities), then **EU** (European Union): *[1962-2009 (F,S)]*

EFTA (European Free Trade Association): *[1962-2009 (F,S)]*

CACM (Central American Common Market): *[1963-69 and 1993-2009 (F)]*

COMECON (Union of Mutual Economic Assistance): *[1964-1990 (F)]*

CEMAC (Economic and Monetary Community of Central Africa): *[1964-2009, except 1981, 1988, 1991 and 1992 (F)]*

OCT (EC FTA with Overseas Countries and Territories): *[1971-2009 (F)]*

CARICOM (Caribbean Community and Common Market): *[1973-2009 (F)]*

EEA (European Economic Area: EC-EFTA FTA): *[1973-2009 (F,S)]*

PATCRA (Agreement on Trade and Commercial Relations between the Government of Australia and the Government of Papua New Guinea): *[1977-2009 (F)]*

EFTASPAIN(EFTA-Spain FTA): *:*[1980-1985 (F,S)]

SADC (Southern African Development Community): *[1980-1988 and 1990-2009 (F)]*

SPARTECA (South Pacific Regional Trade and Economic Cooperation Agreement): *[1981-2009 (F)]*

CER (Australia-New Zealand FTA): *[1983-2009 (F)]*

USISR(US-Israel FTA): *[1985-2009 (F,S)]*

USCAN(US-Canada FTA): *[1989-2009 (F,S)]*

NAFTA (North American Free Trade Agreement): *[1994-2009 (F,S)]*

EC-Andorra FTA: *[1991-2009 (F)]*

EFTA-CEEC FTA: *[1992-2006 (F)]*

EU-CEEC FTA: *[1992-2006 (F)]*

ASEAN (Association of South East Asian Nations FTA): *[1992-2009 (F,S)]*

CEFTA (Central European FTA): *[1993-2009 (F)]*

CIS (Commonwealth of Independent States): *[1995-2009 (F)]*

EAEC (Eurasean Economic Community): *[1997-2009 (F)]*

CEZ (Common Economic Zone): *[2004-2009 (F)]*

SAFTA (South Asian Free Trade Arrangement): *[2006-2009 (F)]*

WAEMU (West African Economic and Monetary Union): *[1996-2009 (F)]*

PAFTA (Pan Arab FTA): *[1998-2009 (F)]*

SACU (Sub Saharan South African Customs Union): *[2000-2009 (F)]*

EAC (East African Community): *[2000-2009 (F)]*

COMESA (Common Market for Eastern and Southern Africa): *[1995-2009 (F)]*

CAN (Andean Community FTA): *[1988-2009 (F,S)]*

MERCOSUR (Southern Common Market): *[1991-2009 (F,S)]*

DOMCAUSA (Dominican Republic - Central America - US FTA): *[2006-2009 (F)]*

TRANSPAC (Trans-Pacific Strategic Economic Partnership FTA): *[2006-2009 (F)]*

EFTASACU (EFTA-SACU FTA): *[2008-2009 (F)]*

ECSYR (EC-Syria FTA): *[1977-2009 (F)]*

ECTUR (EC-Turkey FTA): *[1996-2009 (F,S)]*

ECPAL (EC-Palestinian Authority FTA): *[1997-2009 (F)]*

ECFAR (EC-Faroe Islands FTA): *[1997-2009 (F)]*

ECTUN (EC-Tunisia FTA): *[1998-2009 (F)]*

ECMOR (EC-Morocco FTA): *[2000-2009 (F)]*

ECISR (EC-Israel FTA): *[2000-2009 (F,S)]*

ECSAFR (EC-South Afica FTA): *[2000-2009 (F)]*

EFTATUR (EFTA-Turkey FTA): *[1992-2009 (F,S)]*

EFTAISR (EFTA-Israel FTA): *[1993-2009 (F,S)]*

EFTAPAL (EFTA-Palestinian Authority FTA): *[1999-2009 (F)]*

EFTAMOR (EFTA-Morocco FTA): *[2000-2009 (F)]*

G Full and superbalanced samples

Table 6: List of countries in the full and superbalanced samples

Country name	Status	Country name	Status	Country name	Status
Afghanistan	*R;P*	French Polynesia	*R;P*	N. Mariana Islands	P
Albania	*R;P*	French S. Antartic terr.	P	Norway	*R;P*
Algeria	*R;P*	Gabon	*R;P*	Oman	*R;P*
Andorra	*R;P*	Gambia	*R;P*	Pakistan	*R;P*
Angola	*R;P*	Georgia	*R;P*	Palau	P
Anguilla	*R;P*	**Germany**	**R;P;S**	Panama	*R;P*
Antigua-Barbuda	*R;P*	Ghana	*R;P*	Papua New Guinea	*R;P*
Argentina	**R;P;S**	Gibraltar	P	**Paraguay**	**R;P;S**
Armenia	*R;P*	**Greece**	**R;P;S**	Peru	*R;P*
Aruba	*R;P*	Greenland	*R;P*	**Philippines**	**R;P;S**
Australia	*R;P*	Grenada	*R;P*	Pitcairn	P
Austria	*R;P*	Guadeloupe	*R;P*	Poland	*R;P*
Azerbaijan	*R;P*	Guatemala	*R;P*	**Portugal**	**R;P;S**
Bahamas	*R;P*	Guinea	*R;P*	Qatar	*R;P*
Bahrain	*R;P*	Guinea-Bissau	*R;P*	Reunion	*R;P*
Bangladesh	*R;P*	Guyana	*R;P*	Romania	*R;P*
Barbados	*R;P*	Haiti	*R;P*	Russian Federation	*R;P*
Belarus	*R;P*	Honduras	*R;P*	Rwanda	*R;P*
Belgium	**R;P;S**	**Hong Kong**	**R;P;S**	St. Helena	P
Belize	*R;P*	Hungary	*R;P*	St. Kitts and Nevis	*R;P*
Benin	*R;P*	**Iceland**	**R;P;S**	St. Lucia	*R;P*
Bermuda	*R;P*	India	*R;P*	St. Vincent-Grenadines	*R;P*
Bhutan	*R;P*	Indonesia	*R;P*	Samoa	*R;P*
Bolivia	*R;P*	Iran	*R;P*	San Marino	P
Bosnia-Herzeg.	*R;P*	Iraq	*R;P*	Sao Tome-Principe	*R;P*
Botswana	*R;P*	Ireland	*R;P*	Saudi Arabia	*R;P*
Brazil	**R;P;S**	**Israel**	**R;P;S**	Senegal	*R;P*
Br. Virgin Islands	P	**Italy**	**R;P;S**	Serbia-Montenegro	*R;P*
Brunei Darussalam	*R;P*	Jamaica	*R;P*	Seychelles	*R;P*
Bulgaria	*R;P*	**Japan**	**R;P;S**	Sierra Leone	*R;P*
Burkina Faso	*R;P*	Jordan	*R;P*	**Singapore**	**R;P;S**
Burma	*R;P*	Kazakstan	*R;P*	Slovakia	*R;P*
Burundi	*R;P*	Kenya	*R;P*	Slovenia	*R;P*
Cambodia	*R;P*	Kiribati	*R;P*	Solomon Islands	*R;P*
Cameroon	*R;P*	**Korea**	**R;P;S**	Somalia	*R;P*
Canada	**R;P;S**	DPR of Korea	P	South Africa	*R;P*
Cape Verde	*R;P*	Kuwait	*R;P*	Soviet Union	P
Cayman Islands	P	Kyrgyzstan	*R;P*	**Spain**	**R;P;S**
C.African Republic	*R;P*	Lao PDR	*R;P*	Sri Lanka	*R;P*
Chad	*R;P*	Latvia	*R;P*	St. Pierre and Miquelon	*R;P*
Chile	**R;P;S**	Lebanon	*R;P*	Sudan	*R;P*

Table 7: List of countries in the full and superbalanced samples: Contd.

Country name	Status	Country name	Status	Country name	Status
China	*R;P*	Lesotho	*R;P*	Suriname	*R;P*
Christmas Island	P	Liberia	*R;P*	Swaziland	*R;P*
Cocos Islands	P	Libya	*R;P*	**Sweden**	**R;P;S**
Colombia	**R;P;S**	Lithuania	*R;P*	**Switzerland**	**R;P;S**
Comoros	*R;P*	**Luxembourg**	**R;P;S**	Syria	*R;P*
Congo	*R;P*	Macau (Aomen)	*R;P*	Taiwan	*R;P*
Dem. Rep. of Congo	*R;P*	Macedonia	*R;P*	Tajikistan	*R;P*
Cook Islands	*R;P*	Madagascar	*R;P*	Tanzania	*R;P*
Costa Rica	*R;P*	Malawi	*R;P*	**Thailand**	**R;P;S**
Croatia	*R;P*	**Malaysia**	**R;P;S**	Togo	*R;P*
Cuba	*R;P*	Maldives	*R;P*	Tokelau	P
Cyprus	*R;P*	Mali	*R;P*	Tonga	*R;P*
Czech Republic	*R;P*	Malta	*R;P*	Trinidad-Tobago	*R;P*
Czechoslovakia	*R;P*	Marshall Islands	P	**Tunisia**	**R;P;S**
Côte d'Ivoire	*R;P*	Martinique	*R;P*	**Turkey**	**R;P;S**
Denmark	**R;P;S**	Mauritania	*R;P*	Turkmenistan	*R;P*
Djibouti	*R;P*	Mauritius	*R;P*	Turks-Caicos Islands	*R;P*
Dominica	*R;P*	**Mexico**	**R;P;S**	Tuvalu	*R;P*
Dominican Republic	*R;P*	Micronesia	P	Uganda	*R;P*
East Germany	**R;P;S**	Moldova	*R;P*	Ukraine	*R;P*
East Timor	*R;P*	Mongolia	*R;P*	United Arab Emirates	*R;P*
Ecuador	*R;P*	Montserrat	*R;P*	**United Kingdom**	**R;P;S**
Egypt	*R;P*	Morocco	*R;P*	**USA**	**R;P;S**
El Salvador	*R;P*	Mozambique	*R;P*	Uruguay	*R;P*
Equatorial Guinea	P	Namibia	*R;P*	Uzbekistan	P
Eritrea	*R;P*	Nauru	P	Vanuatu	*R;P*
Estonia	*R;P*	Nepal	*R;P*	**Venezuela**	**R;P;S**
Ethiopia	*R;P*	Netherland Antilles	*R;P*	Vietnam	*R;P*
Falkland Islands	P	**Netherlands**	**R;P;S**	Wallis-Futuna	*R;P*
Faroe Islands	*R;P*	New Caledonia	*R;P*	**West Germany**	**R;P;S**
Fiji	*R;P*	New Zealand	*R;P*	Western Sahara	P
Finland	*R;P*	Nicaragua	*R;P*	Yemen	*R;P*
Fm Vietnam DR	*R;P*	Niger	*R;P*	Yugoslavia	*R;P*
Fm Vietnam Rp	*R;P*	Nigeria	*R;P*	Zambia	*R;P*
France	**R;P;S**	Niue	P	Zimbabwe	*R;P*
French Guiana	*R;P*	Norfolk Island	P		

The full sample contains 207 reporters (R) and 230 partners (P) which are listed in tables 6 and 7. 'S' indicates that the country is present in the superbalanced sample.

In the full sample, several countries shift from reporting trade on an individual basis to reporting trade jointly with another country. This is the case of Belgium and Luxembourg, as well as Eritrea and Ethiopia. For consistency, we use a single country identifier for each of these two pairs. A single country identifier is used for Yougoslavia and for Serbia and Montenegro.

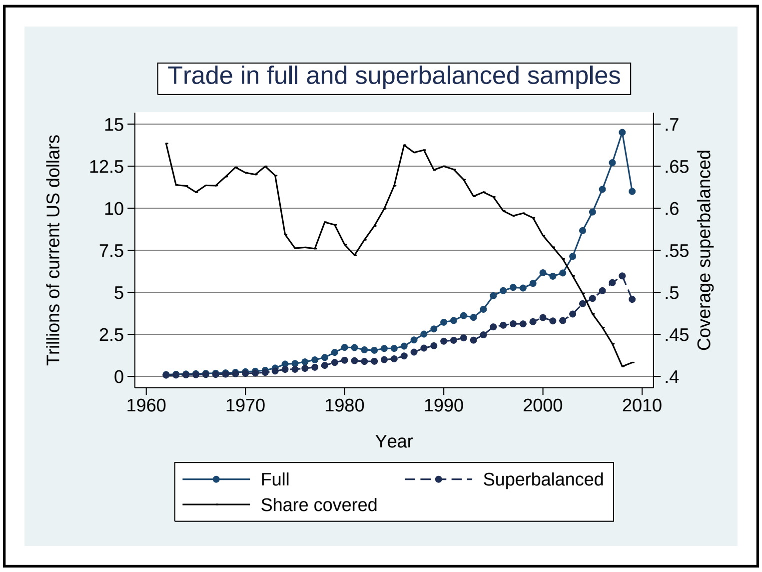


Figure 14: Trade coverage in the superbalanced sample

The superbalanced sample corresponds to the subsample of trading partners which trade both ways in each and every year of the sample. This corresponds to 32 reporters and partners, and 786 reciprocal pairs out of the 992 possible pairs. To avoid discarding reported trade for pairs which have reciprocal trade, but would fall out of the sample because of countries' split-up in several entities, or reunification, we introduce several additional single country identifiers before constructing the superbalanced sample. A single identifier is used for the Czech Republic, Slovakia, and Czechoslovakia; another single identifier for East, West, and reunited Germany; a single identifier for the USSR and the 15 countries which were formed after USSR split up. This brings Germany to the superbalanced sample, but the 15 countries which constituted the USSR still drop out because the USSR is never a reporter to COMTRADE. The Czech Republic and Slovakia also drop out because they do not have reciprocal trade in all years with another country of the sample.

As shown in Fig.14, the superbalanced sample covers 50-70% of full sample trade. Fig.15 shows the distribution of pairs in the full sample according to the number of years they are present in the sample.

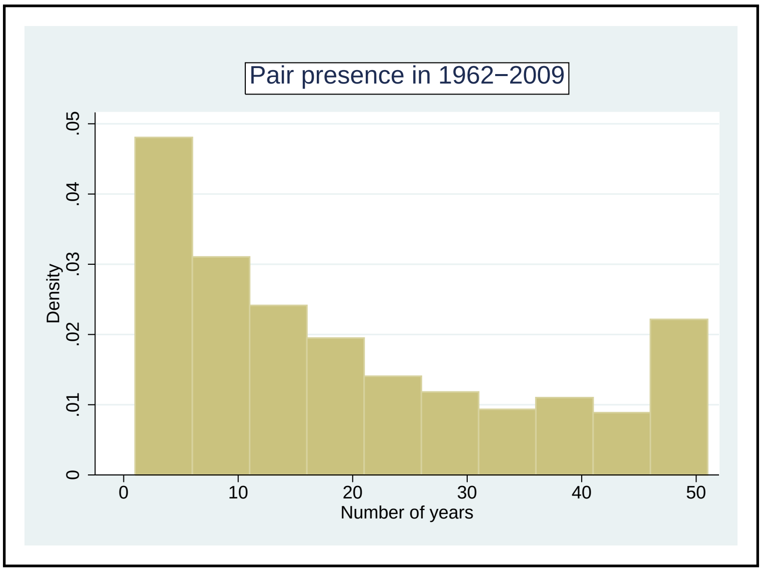


Figure 15: Number of years each pair is present in the sample

As a check on the way the superbalanced sample is constructed, we redefine the set of pairs which trade both ways in each year starting from 1970, e.g. the first year in which there are more than 100 reporters in the COMTRADE database.[93] This gives a sample of 1604 reciprocal pairs out of the 2162 possible pairs for 47 reporters (partners). Fig.16 shows that trade coverage is qualitatively similar.

Fig.17 shows the share of total trade covered by pairs which traded both ways in 1962, and as a check, the share of total trade by pairs which traded both ways in 1970. Qualitatively, trade coverage is similar to the superbalanced samples for, respectively, 1962 (1970). In the same figure, it is shown that one-way trade flows represent a marginal and decreasing share of world trade: more than 95% of total annual trade takes place between pairs which trade both ways in that year. This finding is complementary to Helpman et al. (2008) who find that the increase in world trade is driven by pairs which trade both ways in 1970. Helpman et al. (2008) find that the enlargement of the set of trading partners did not contribute in a major way to the growth of world trade in 1970-1997. We nuance this finding by showing that trade between partners who did not trade both ways in 1962 (1970) did contribute strongly to the increase in total

[93] There are 71 reporters in 1962, and 112 in 1970.

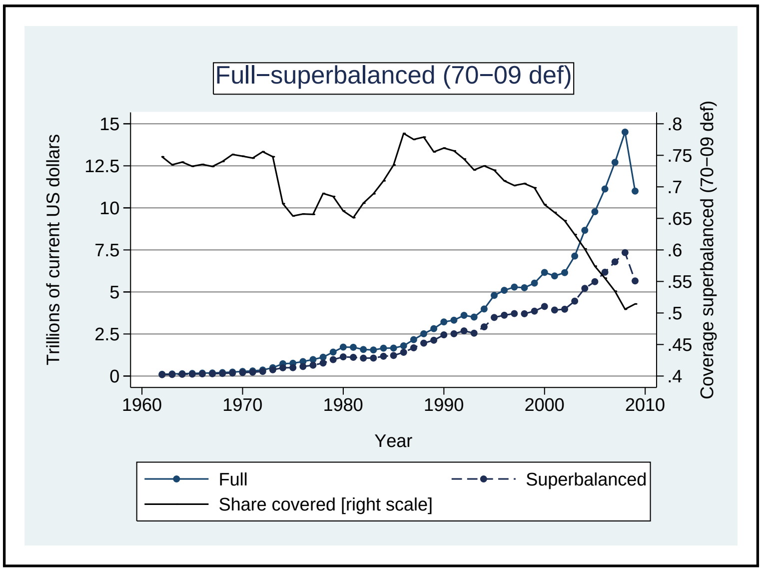


Figure 16: Trade coverage in superbalanced sample defined in 1970-2009

trade in 1995-2009. Furthermore, these new trade relationships were formed between countries trading both ways. Part of the explanation of the widening gap between the superbalanced and the full sample in the recent period is linked to the absence of Central and Eastern European countries and of China from the superbalanced sample as these countries did not report trade to COMTRADE until the recent period.

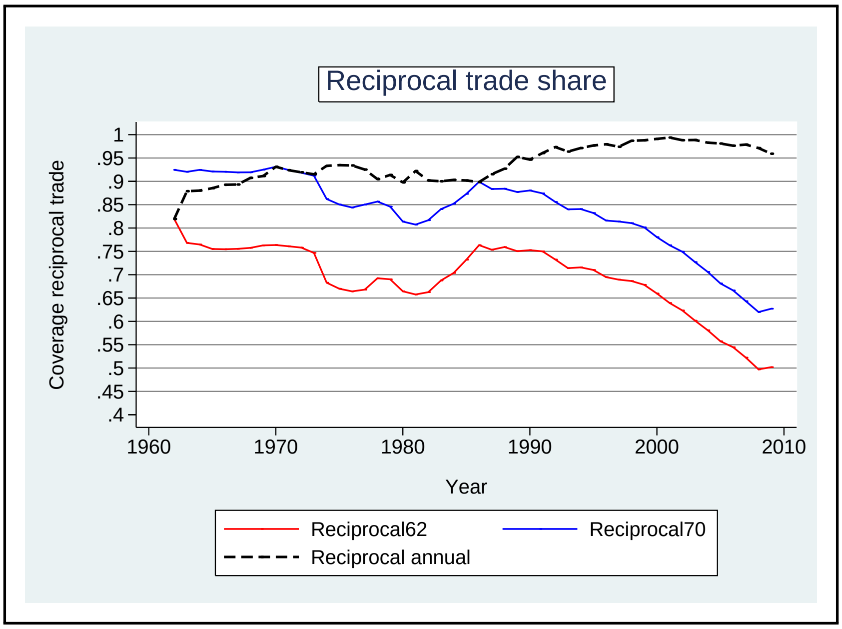


Figure 17: Trade coverage in superbalanced sample defined in 1970-2009