# Heterogeneity and the Distance Puzzle*

Elizaveta Archanskaia[†] Guillaume Daudin[‡]

July 2012

**Abstract**

This paper shows that reduced heterogeneity of exporter-specific goods provides a direct explanation of the distance puzzle. 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, i.e. the substitution elasticity in the Armington framework. The Armington elasticity has increased by 13% from 1963 to 2009 while the distance elasticity of trade has increased by just 7%. The evolution of the distance coefficient is thus compatible with a 5-7% 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, Anne-Célia Disdier, Peter Egger, Lionel Fontagné, Raphaël Franck, Guillaume Gaulier, 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, the international workshop on the Economics of Global Interactions (Bari), the Ljubljana Empirical Trade Conference (Forum for Research in Empirical Internation Trade). We thank Christian Broda, Natalie Chen, Dennis Novy, and David E. Weinstein for giving us access to their programs. The usual disclaimers apply.

†SciencesPo/OFCE

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

# Introduction

The estimated effect of distance in gravity equations has remained stable, or even slightly increased, in the past 50 years.[1] This is the "distance 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" (Cairncross (1997); Levinson (2006); Friedman (2007)).

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 (Santos Silva and Tenreyro (2006); Coe et al. (2007); Bosquet and Boulhol (2009)). See §I.2 for a discussion). 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 established trade relations has increased, and extensive, in the sense that new trade relations have been established (Helpman et al. (2008); Baldwin and Harrigan (2007)). 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 (Bosquet and Boulhol (2009); Duranton and Storper (2008); Egger (2008); Buch et al. (2004)). 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 and 2005 on the distance coefficient, but find that it has had a negligible effect. Using a log-linear estimation

[1] Disdier and Head (2008) find that distance impedes trade by 37% more since 1990 than it did from 1870 to 1969 Berthelon and Freund (2008) find that the distance coefficient has increased by 10% in absolute value over the period 1985-2005. See also Combes et al. (2006); Brun et al. (2005); Buch et al. (2004).

strategy, Brun et al. (2005) and Márquez-Ramos et al. (2007) find decreasing distance elasticities once the sample is restricted to developed countries. Secondly, it is not certain that transport costs have declined relative to the price of traded goods (Hummels (1999); Daudin (2003, 2005); Hummels (2007)). Furthermore, it might be the case that other distance-related components of trade costs, such as delays, have a growing importance (Hummels and Schaur (2012); Hummels (2001)).

Thirdly, the value of the distance coefficient in theoretically derived gravity models of trade is not directly related to the level of trade costs.[2] It 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 (Feyrer (2009)).

The discussion on the distance puzzle in this second strand of the literature has been mainly concerned with the evolution of the elasticity of trade costs to distance. One should look as well at the elasticity of trade flows to trade costs. Following Arkolakis et al. (2012), 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 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 the same set of goods as 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 composite goods, and hence increased elasticity of trade flows to trade costs, providing a direct explanation of the distance puzzle (See the Anderson and van Wincoop (2003) derivation of the gravity model which uses the demand framework pioneered by Armington (1969). See §II.2 for a discussion). In the Ricardian framework, it could be argued that countries' technological capabilities have become more homogeneous across sectors overtime. This evolution would reduce the strength of com-

[2] The level of trade costs, captured by country-specific fixed effects, influences the openness of each country.

parative advantage in determining trade flows, leading to increased sensitivity of trade flows to trade costs (Eaton and Kortum (2002).). In the framework of monopolistic competition with heterogeneous firms, a reduced efficiency dispersion of firms within any given sector, i.e. a relatively bigger share of firms situated near the export threshold, could have increased the sensitivity of trade flows to trade costs (Chaney (2008).).

II.3 shows that the data required to compute annual trade elasticities in 1962-2009 in the Ricardian and heterogenous firms frameworks are unavailable. Therefore, this paper's original contribution is to investigate whether the evolution of the trade elasticity in the Armington framework provides a direct explanation of the distance puzzle. We find that the elasticity of substitution between country-specific composite export goods, i.e. the aggregate trade elasticity in the Armington framework, has increased by at least 13% from 1963 to 2009. Two conclusions follow. First, the non-decreasing distance elasticity of trade is fully explained by the increase in the elasticity of trade flows to trade costs. Second, the distance puzzle ceases to exist insofar as the elasticity of trade costs to distance is found to have decreased by at least 5-7% from 1963 to 2009. The reduction in the elasticity of trade costs to distance is even more pronounced if we focus on the period in which air transportation plays an important role in world trade. In 1970-2009, the elasticity of trade costs to distance has decreased by 17%.

# I 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. This section measures the distance puzzle in our data and evaluates whether some combination of the factors identified by previous studies as explanatory of movements in the distance coefficient pins down the determinants of the distance puzzle. We focus on:[3]

- the composition effect defined as the evolution in the goods' composition of world trade;
- 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 Free Trade Agreements between existing countries.

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

## I.1 Data

We use the COMTRADE world bilateral trade dataset spanning the years 1962-2009[4]. We use the 4-digit SITC Rev.1 classification of goods (600-700 goods), as this provides the longest coverage of disaggregated bilateral trade flows. We restrict the sample to trade in goods which are attributed to specific 4-digit categories. As trade data is of better quality for imports, the estimation is conducted only on import flows. Our sample covers between 88% and 99% of reported trade in COMTRADE.

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 are taken from the CEPII.[5] We constructed the database on GATT/WTO and FTA membership.[6]

## I.2 Evolution of the distance elasticity of trade

We discuss in part II why our data only allow investigating the determinants of the distance puzzle in the Armington framework. Acknowledging this limitation, we use the Anderson and van Wincoop (2003) derivation of the gravity equation in this section, even though the estimation procedure is also valid for the other two main derivations of the gravity trade model: Eaton and Kortum (2002), and Krugman (1980), extended by Chaney (2008) to a setting with heterogeneous firms.[7]

$$X_{ij,t} = \left(\frac{Y_{i,t}Y_{j,t}}{Y_t}\right)\left(\frac{\tau_{ij,t}}{\Pi_{i,t}P_{j,t}}\right)^{-\zeta_t} \tag{1}$$

$X_{ij,t}$ is the value of goods from country $i$ consumed in country $j$ in year $t$, i.e. bilateral imports at cif prices. Imports are a function of the nominal income of each trading partner $Y_{ij,t}$ and $Y_{j,t}$, of world income $Y_t$, of bilateral trade costs $\tau_{ij,t}$, and of inward and outward multilateral trade resistance terms $\Pi_{i,t}$, $P_{j,t}$.

4 The 228 trading partners present in this dataset are listed in App.C.

5 See Mayer and Zignago (2011). The database is available at www.cepii.fr. We constructed bilateral distance and bilateral costs' controls for East and West Germany, USSR, and Czechoslovakia.

6 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. B for the list of included FTAs and the years they appear in our data.

7 See footnote 20 in Eaton and Kortum (2002) for a discussion of the equivalence.

Across frameworks, the parameter $\zeta_t$ is the trade elasticity. In the Armington framework, $\zeta_t = \sigma_t - 1$, where $\sigma_t$, the Armington elasticity of substitution, captures the degree of substitutability of composite goods of different national origin.

As is standard in the gravity literature, we model bilateral trade costs as a function of bilateral distance $dist_{ij}$, a vector $Z_1$ of additional bilateral trade cost controls such as adjacency, common language, colonial linkages, belonging or having once belonged to the same country, and a vector $Z_2$ of trade cost controls linked to trade policy such as common membership of GATT/WTO and common membership of an FTA.

$$\tau_{ij,t} = exp\left\{\rho_t \ln dist_{ij} + Z_{1,t}{}'\beta_{1,t} + Z_{2,t}{}'\beta_{2,t}\right\} \quad (2)$$

Working with bilateral imports data in cross-section and omitting time subscripts to simplify notation, the equation to be estimated is (3):

$$X_{ij} = \exp\left(\alpha_0 - \alpha_1 \ln dist_{ij} + Z_1{}'\beta_1 + Z_2{}'\beta_2 + fe_{exp} + fe_{imp}\right) \varepsilon_{ij} \quad (3)$$

where $fe_{exp}$ and $fe_{imp}$ are respectively exporter and importer fixed effects and $\varepsilon_{ij}$ is a multiplicative error term. The coefficient of interest is the absolute value of the distance elasticity of trade, $\alpha_1$, which in the Armington framework is the product of the distance elasticity of trade costs $\rho$ and the Armington elasticity of substitution $(\sigma - 1)$.

Santos Silva and Tenreyro (2006) (SST) have shown that the structure of trade data is such that in general the additive error term in the log-linearized model is heteroscedastic, and its variance depends on the regressors. In this case OLS does not generally give consistent slope estimates.[8] SST advocate the use of the Poisson pseudo maximum likelihood (PPML) estimator because it performs best along the consistency and efficiency dimensions relatively to alternative non linear estimators.[9]

Following SST we use the PPML to estimate the distance elasticity of trade in cross-section. A potential drawback of this estimator is that its implied functional form is not consistent with the prevalence of zero trade observations in the data even though it allows including zero trade observations in the estimation.[10] We therefore check in App.A that results are robust to switching to a two-part model in which the probability of forming a trade relationship is estimated in the first step.

[8] OLS gives consistent coefficient estimates in the case of $V[y_i|x] = \mu(x_i\beta)^2$, i.e. when the conditional variance of raw trade flows follows a Gamma distribution.

[9] See Manning and Mullahy (1999); Bosquet and Boulhol (2009) for a discussion.

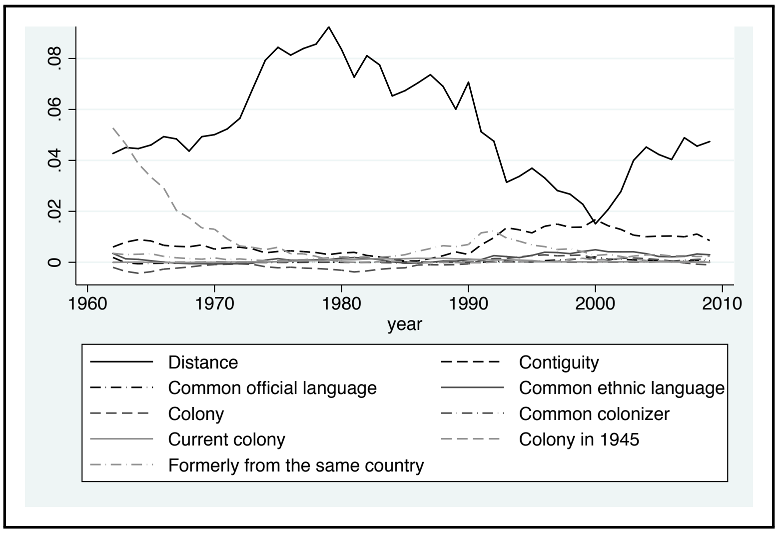
[10] See Egger et al. (2011) for a discussion of the basic PPML and the two-part PPML.

The baseline PPML regression excludes $Z_2$, the vector of trade cost controls linked to trade policy. We find that $\alpha_1$, the absolute value of the distance coefficient, has increased over 1962-2009 (see Fig.1). There is indeed a distance puzzle.

Figure 1: Distance coefficient in the baseline PPML regression


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

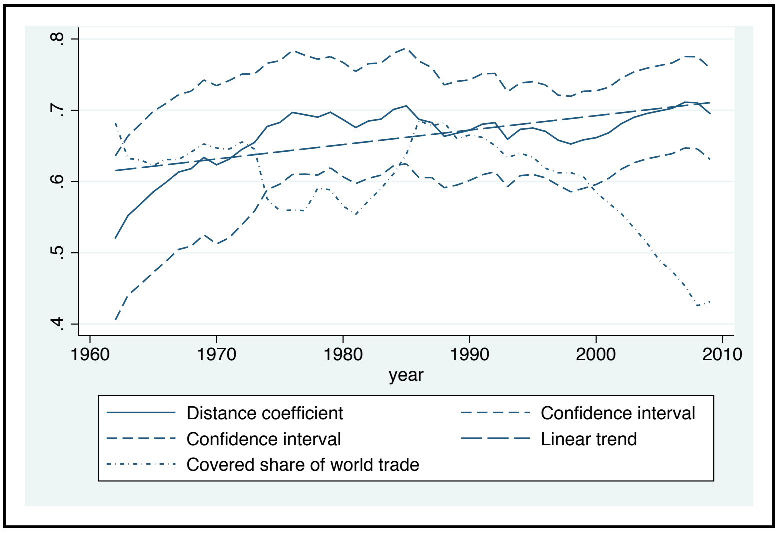
Figure 2: Coefficient of partial determination in the PPML baseline equation



## I.3 Robustness of the distance puzzle

To check the robustness of 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 786 stable two-way pairs.[11] Fig. 3 shows that the sample effect does not reduce the distance puzzle.

Figure 3: Distance coefficient in sample of stable pairs (PPML)




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 in total trade are fixed at their 1962 values. As shown by Fig. 4, the composition effect deepens the distance puzzle.[12]

Contrasting both Fig. 3 and Fig. 4 with Fig. 1, we note that most short-term fluctuations of the distance coefficient seem to be explained by sample and composition effects.

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.[13] As shown in Fig. 5, controlling for countries' participation into free trade agreements solves

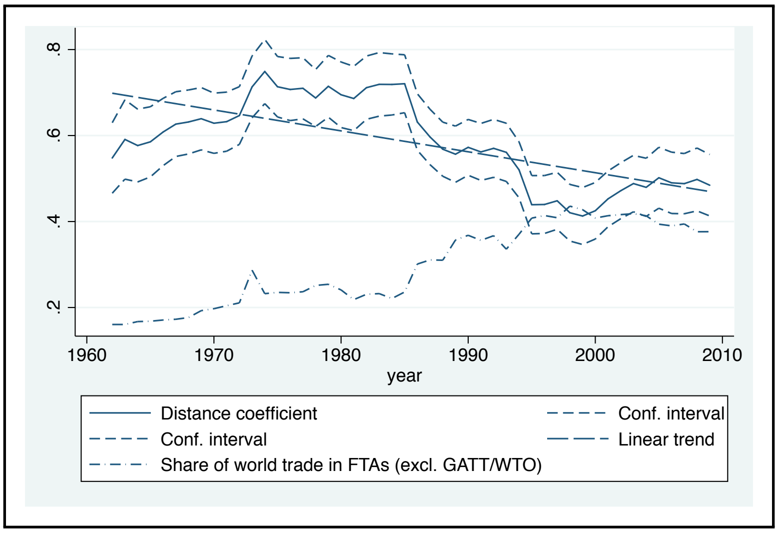
[11] This is a rectangular sample which includes 32 countries. See App.C for details.

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

[13] See Appendix B for the full list of included FTAs. Restricting this list to the "main" FTAs (EC and EU, USA-Canada FTA and NAFTA, the Comecon, EFTA, ASEAN, Mercosur, as well as membership of GATT-WTO) has no qualitative impact on the results.

Figure 4: Distance coefficient controlling for composition (1962 weights)


Figure 5: Distance coefficient controlling for FTAs (full sample)




the distance puzzle. However, this effect is mechanical. Indeed, Fig. 5 shows that in-FTA trade increased from less than 20% to approximately 40% of total trade from 1962 to 2009, and Fig. 6 shows that over the same period the share of intra-FTA trade among nearby (less than 2000 km) countries increased from 38% to 79% in 2009.[14] This means that including controls for FTA membership amounts to adding a growing number of proximity controls in the regression. This reduces the sensitivity of trade flows to distance.

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

A further argument for caution in interpreting results with FTA controls is the probable endogeneity of FTAs for which we do not correct in the 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 they control for FTAs correcting for endogeneity (fig. 11 in their paper).

Figure 6: Share of intra-FTA trade among nearby countries (2000km or less)

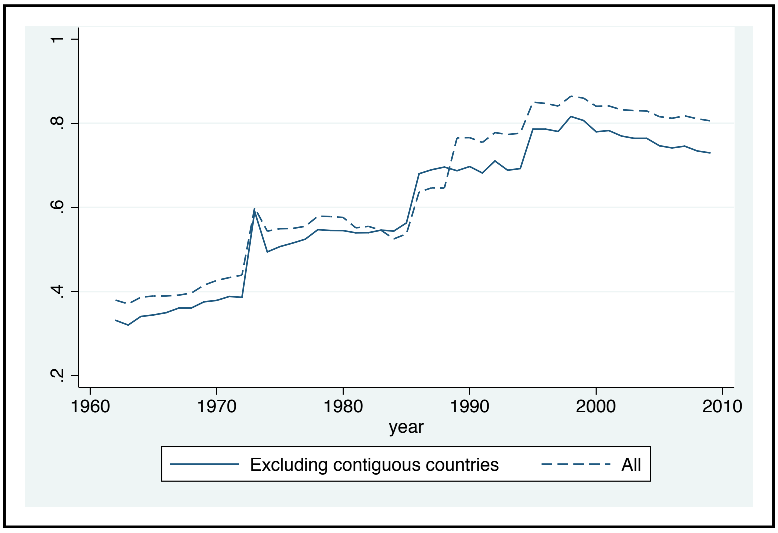



Table 1 summarizes our findings on the distance puzzle. Controlling separately and jointly for sample and composition effects does not change the magnitude of the distance puzzle. On the other hand, controlling for FTAs eliminates the distance puzzle. As argued, the explanation of the distance puzzle through FTAs is not satisfactory because it 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 a robust feature of the data, 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 in the PPML regression depending on sample, composition and FTA effects

| | Total change 62-09 (geometric trend) | % change relatively to baseline |
|---|---|---|
| Baseline | 1.07 | |
| Sample effect | 1.14 | 7% |
| Composition effect | 1.14 | 7% |
| FTA effect | 0.49 | -54% |
| Composition + Sample | 1.14 | 7% |
| Composition + FTA | 0.75 | -29% |
| Sample + FTA | 0.44 | -59% |
| Sample + Composition + FTA | 0.49 | -54% |

# II Interpreting the distance coefficient

The distance coefficient is the product of two elasticities: the elasticity of trade costs to distance $\rho_t$ and the elasticity of trade flows to trade costs $\zeta_t$. The question addressed in this paper is very simple. Have trade flows become more sensitive to trade costs (increasing $\zeta$), or have trade costs become more sensitive to distance (increasing $\rho$)? In III we estimate the evolution of $\zeta$ to deduce the evolution of $\rho$. The three main microfoundations of the gravity model of trade give structurally different interpretations to $\zeta$ but not to $\rho$. In this section we define $\zeta$ in each model and show that the estimation procedure for $\zeta$ differs across models. We provide details on the procedure used in this paper to estimate the evolution of $\zeta$ in the Armington framework, and show why the lack of domestic price data impedes the estimation of $\zeta$ in the Ricardian and the heterogeneous firms' frameworks.

## II.1 What does the trade elasticity measure in each model?

### II.1.1 Ricardian framework

In Eaton and Kortum (2002) the heterogeneity dimension captured by the trade elasticity $\zeta$ is intersectoral. Consumers maximize a CES utility function defined at the intersectoral level by shopping around the world for the cheapest supplier of each sectoral good. Thus, contrary to the Armington framework, goods of different national origin are perfect substitutes.

A combination of ingredients determines the cheapest supplier. The accumulated stock of know-how $T_i$ is the level of the overall technological ability of the exporting country. The fundamental cost component $c_i$ is the cost of the bundle of inputs in the exporting country. As in the Armington framework, this cost component is augmented by iceberg trade costs $\tau_{ij}$. The final ingredient is the source of intrinsic heterogeneity in the model, and it is due to the assumption of labor productivity heterogeneity across the continuum of goods.

A Fréchet distribution parameter $\theta$ common to all countries captures the degree of dispersion in productive efficiency across sectors within countries, i.e. the strength of comparative advantage. 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.

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 $\pi_{ij}$ that $i$'s cif price in country $j$ is lowest among the set of all possible suppliers to $j$:

$$\pi_{ij} = \Pr\left[p_{ij}(k) \leq \min\left\{p_{sj}(k)\right\} s \neq i\right] = \int_0^\infty \prod_{s\neq i} \left[1 - G_{sp}(p)\right] dG_{ij}(p) = \frac{T_i(c_i\tau_{ij})^{-\theta}}{\Phi_j} \quad (4)$$

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 $p$.

Defining bilateral trade costs $\tau_{ij}$ as in (2), the distance coefficient $\alpha_1$ estimated by (3) is $\alpha_1^{RICARDO} = \rho\theta$, where $\theta = \zeta$ is the trade elasticity in the Ricardian framework.

### II.1.2 Melitz-Chaney framework

In the Melitz-Chaney framework the heterogeneity dimension captured by the trade elasticity $\zeta$ is intrasectoral. Consumers value firm-specific varieties of sectoral goods which they acquire in monopolistic competition markets. The fraction of total expenditure allocated to the differentiated sector is $\mu$, with the remaining income spent on the homogeneous freely tradable good produced with constant returns to scale technology. Trade in differentiated goods is costly: on top of $\tau_{ij}$ the firm pays fixed costs $f_{ij}$ to acquire exporter status.

Firms draw a productivity characteristic from a Pareto distribution with shape parameter $\gamma$ which determines the magnitude of the trade elasticity.[15] The lower the degree of efficiency dispersion in domestic markets, the higher the trade elasticity. The intuition for this result is given in Chaney (2008). The parameter $\gamma$ 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. Bilateral exports are given by (5):

$$X_{ij} \;=\; \mu\frac{Y_iY_j}{Y}\frac{\left(c_i\tau_{ij}\right)^{-\gamma}}{\Psi_j}f_{ij}^{-\left(\frac{\gamma}{\sigma_f-1}-1\right)} \tag{5}$$

where $\Psi_j = \sum_{i=1}^{N} L_i c_i \left(c_i\tau_{ij}\right)^{-\gamma}\left[f_{ij}\right]^{1-\frac{\gamma}{(\sigma_f-1)}}$ is the overall price distribution parameter in country $j$, $c_i$ is the wage, i.e. the fundamental cost component in the exporting country, $L_ic_i$ is labor income in $i$, and $\sigma_f$ captures the degree of substitutability between firm-level varieties.

Defining variable trade costs $\tau_{ij}$ as in (2) and fixed trade costs as a function of trade cost controls which do not include distance $f_{ij} = exp\left[Z'\beta\right]$ where $Z = \{Z_1, Z_2\}$, the distance coefficient $\alpha_1$ estimated by (3) is $\alpha_1^{HET-FIRMS} = \rho\gamma$, where $\gamma = \zeta$ is the trade elasticity in the Melitz-Chaney framework.[16]

### II.1.3 Armington framework

In Anderson and van Wincoop (2003) there is no heterogeneity in productive efficiency. The heterogeneity dimension comes from the assumption that consumers perceive products of

[15] See the technical appendix to Chaney (2008) for the proof that the elasticity of trade flows to variable trade costs is common across countries and equal to $\gamma$ under the small country assumption.

[16] If distance is included in the definition of fixed bilateral trade costs as in Chaney (2008), the distance coefficient is no longer a product of two elasticities: $\alpha_1^{HET-FIRMS-FULL} = \rho\gamma + \rho_2\left(\frac{\gamma}{\sigma_f-1}-1\right)$. Throughout this paper, bilateral components of fixed costs are identified through the $Z$-vector of bilateral proximity controls and $\rho_2 = 0$.

different national origin as intrinsically imperfect substitutes. Each country is specialized in production of country-specific varieties which on aggregate give a country-specific composite good. The parameter $\sigma$ is the lower tier Armington elasticity of substitution. It measures the degree of substitutability of goods of different national origin.[17] Bilateral exports are:

$$X_{ij} = \nu_i Y_j \left(\frac{c_i \tau_{ij}}{P_j}\right)^{-(\sigma-1)} \tag{6}$$

where $c_i$ is the cost of producing domestic varieties in country $i$, $\tau_{ij}$ is the bilateral trade cost, $\nu_i$ is the preference parameter for country $i$ products, and $P_j$ is the aggregate price index in country $j$. Defining bilateral trade costs $\tau_{ij}$ as in (2), the distance coefficient $\alpha_1$ estimated by (3) is $\alpha_1^{ARMINGTON} = \rho(\sigma-1)$, where $(\sigma-1) = \zeta$ is the trade elasticity.

To sum up, a common specification of variable trade costs $\tau_{ij}$ across models leads to a common definition of the elasticity of trade costs to distance $\rho$ while each model gives a different structural interpretation to $\zeta$, the elasticity of trade flows to trade costs. Across frameworks this trade elasticity is decreasing in the degree of heterogeneity observed in a single dimension which is framework-specific.

## II.2 What do we know about the evolution of the trade elasticity?

In these models there is no theoretical mechanism to bring about a change in the trade elasticity overtime. A shock to consumer preferences or to the shape parameter of the productivity distribution would be required. Nonetheless, the heterogeneity parameter measured on aggregate trade data could have evolved over 1962-2009 without any shock to the underlying heterogeneity, either through changes in the range of traded goods, time-sensitive aggregation issues linked to estimating a single parameter across sectors, or agents' adaptation to a changing economic environment.

However, there is little empirical evidence on the evolution of sector-specific and aggregate trade elasticites in either model. Three theoretically grounded methodologies have been used to estimate aggregate trade elasticities in the Ricardian framework for a specific year(Eaton and Kortum (2002); Simonovska and Waugh (2011); Costinot et al. (2012); Caliendo and Parro (2011)), but only Levchenko and Zhang (2011) study the evolution of intersectoral productivity

[17] The upper-tier elasticity measures substitutability of domestic products and an aggregate import good. The lower-tier one measures substitutability between importers of a given good. See Sato (1967); Reinert and Shiells (Reinert and Shiells); Saito (2004)

dispersion. They find evidence of within-country convergence in sectoral knowledge stocks in 1960-2010. As there is less heterogeneity in producer efficiency across the set of goods comparative advantage exerts a weaker force against trade resistance imposed by trade barriers. In terms of the heterogeneity parameter, this result would be consistent with an increasing $\theta$ in 1960-2010, reducing the distance puzzle.[18]

To the best of our knowledge, there is no direct empirical evidence on the evolution of sectoral or aggregate efficiency dispersion parameters in the heterogenous firms framework.

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) provide evidence on 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 at all levels of product disaggregation, i.e. at the 10-digit, 5-digit, and 3-digit levels. These results indicate that the parameter estimated on aggregate trade 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 bilateral trade while constraining the parameter to be the same across destination markets.

## II.3 Which trade elasticity can be measured over the whole period?

In this section, we explain the methodology used to estimate the trade elasticity parameter in the Armington framework. And we provide details on data limitations which impede estimating the trade elasticity parameter in the framework with efficiency heterogeneity.

### II.3.1 Estimation procedure for the Armington framework

The UN COMTRADE bilateral trade database covers the majority of countries over 1962-2009. It gives information on trade flows and unit values at the SITC 4-digit level. This data is sufficient to estimate the trade elasticity in the Armington framework.

The distance puzzle concerns an elasticity estimated on aggregate trade data. As shown

[18] In Levchenko and Zhang (2011) reduced dispersion in sector-specific technology stocks is obtained under the assumption of a constant $\theta$. To make sense of this result a sector-specific technological spillover function across trade partners would have to be specified.

by Imbs and Méjean (2011) this parameter cannot generally be mimicked by a theoretically grounded weighted average of sector-specific trade elasticities. Hence we need an estimation procedure that works directly with aggregate data. To estimate the trade elasticity parameter in the Armington framework, we apply to the SITC 4-digit data a consistent aggregation procedure to get relative prices of country composite goods assuming equality of sector-specific trade elasticities.

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: $X_{ij} = \sum_k X_{k,ij}$.[19] Given CES utility at the intersectoral level, sectoral demand in $j$ in sector $k$ is given by:

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

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:

$$X_{k,ij} = \left(\frac{P_{k,ij}}{P_{k,j}}\right)^{-(\sigma_k-1)} Y_{k,j}$$

Replacing $Y_{k,j}$ by its value gives:

$$X_{k,ij} = \left(\frac{P_{k,ij}}{P_{k,j}}\right)^{1-\sigma_k} \left(\frac{P_{k,j}}{P_j}\right)^{1-\sigma} Y_j \tag{7}$$

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

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

Summing over all SITC 4-digit sectors:

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

Going back to the expression for $X_{k,ij}/Y_j$, use (7) to write:

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

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

Replacing $\frac{P_{k,j}}{P_j}$ by its value and defining $\frac{Y_{k,j}}{Y_j} = \omega_{k,j}$, we get:

$$\frac{X_{k,ij}}{Y_{k,j}} \frac{Y_{k,j}}{Y_j} = \left[ \frac{P_{k,ij}}{P_{k,j}} \left( \omega_{k,j} \right)^{1/1-\sigma} \right]^{1-\sigma}$$

Summing over all SITC 4-digit sectors:

$$\sum_{k=1}^{K} \frac{X_{k,ij}}{Y_{k,j}} \frac{Y_{k,j}}{Y_j} = \sum_{k=1}^{K} \omega_{k,j} \left[ \frac{P_{k,ij}}{P_{k,j}} \right]^{1-\sigma}$$

Multiplying and dividing the right hand side of the expression by $\omega_{k,j}^{1-\sigma}$ and taking logs:

$$\ln \left[ \frac{X_{ij}}{Y_j} \right] = \ln \left\{ \sum_{k=1}^{K} \omega_{k,j}^{\sigma} \left[ \omega_{k,j} \frac{P_{k,ij}}{P_{k,j}} \right]^{1-\sigma} \right\} \quad (8)$$

Working with the right hand side of (8), and using the approximation that for a large number of sectors $k$, $\ln \sum_k x_k \approx \sum_k \ln x_k$:

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

The first term disappears because:

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

Using the same approximation as previously for the second term:

$$(1-\sigma) \sum_{k=1}^{K} \ln \left[ \omega_{k,j} \frac{P_{k,ij}}{P_{k,j}} \right] \approx (1-\sigma) \ln \left[ \sum_{k=1}^{K} \omega_{k,j} \frac{P_{k,ij}}{P_{k,j}} \right]$$

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

$$\ln \left[ \frac{X_{ij}}{Y_j} \right] \approx -(\sigma-1) \ln \left[ \sum_{k=1}^{K} \omega_{k,j} \frac{P_{k,ij}}{P_{k,j}} \right] \quad (9)$$

Exponentiating gives the equation which is estimated in cross-section:

$$X_{ij}/Y_j = \exp \left[ \lambda_0 - (\sigma-1) \ln \left( \sum_{k} \omega_{k,j} \frac{P_{k,ij}}{P_{k,j}} \right) + fe_{exp} + fe_{imp} \right] \eta_{ij} \quad (10)$$

where $fe_{exp}$ and $fe_{imp}$ are source and destination fixed effects, $\eta_{ij}$ is a multiplicative error term, and $\lambda_0$ 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 I.I.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 function of the weighted average of sectoral relative prices of each source in the destination. Each sectoral relative price is weighted according to the share of the sector in total expenditure of the destination which means that identical weights are applied on both sides of the equation.[20]

Using COMTRADE, source-specific sectoral prices $P_{k,ij}$ 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 $k$. Destination-specific sectoral prices $P_{k,j}$ 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: $P_{k,j} = \sum w_{k,ij} P_{k,ij}$ with $w_{k,ij} = X_{k,ij}/Y_{k,j}$. Exporters for which some trade but no unit value is observed in $k$ are excluded from the computation of $P_{k,j}$.

This procedure allows allows estimating the perceived substitutability of country composite goods in all three frameworks, as all assume CES utility function for consumer preferences. Aggregate bilateral trade can 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 (11):

$$X_{ij} = \left(\frac{P_{ij}}{P_j}\right)^{-(\sigma-1)} Y_j \tag{11}$$

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 $(\sigma - 1)$ captures substitutability of country-composite goods across frameworks. However, it corresponds to the aggregate trade elasticity $\zeta$ only in the Armington framework.

### II.3.2 Estimation procedure for the framework with producer heterogeneity

The COMTRADE database does not allow estimating the trade elasticity parameter in the Ricardian or the Melitz-Chaney framework because it does not contain information on sectoral domestic prices in the destination. The intuition for this result is the following: the threshold which determines the fraction of producers which enter any market is destination-specific in both frameworks. The observed distribution of source-specific prices is therefore not informative of the underlying efficiency heterogeneity but rather of the degree of substitutability across

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

effectively exported goods. On the other hand, the price distribution in each destination across all sources including itself informs on the shape parameter of the productivity distribution.[21]

Costinot et al. (2012) have developed an empirical approach to get unbiased estimates of the efficiency heterogeneity parameter $\theta$ using producer price indices in the destination instrumented by R&D expenditure and export shares normalized by the fraction of domestic expenditure. To implement this approach in our investigation of the distance puzzle we need data on gross output, aggregate producer price indices, and R&D expenditures on top of the COMTRADE database. These data are unavailable for the majority of countries in 1962-2009.

# III Measuring the evolution of heterogeneity

## III.1 Missing unit values

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 83% to 90% between 1962-2000, and a subsequent decrease back to 82% in 2001-2009.[22] We assume that information on quantities is missing due to imperfections in the data collection procedure, and that bilateral trade flows are observed with a similar degree of precision whether or not quantity had been recorded. To deal with missing uv, we impute prices from similar products[23].

The stepwise price imputation procedure is as follows. The relative price of each source in the destination is constructed at the highest disaggregation level for each product and quantity unit in which the source is active, the *4'-digit* level. We then proceed level by level for

[21] The technical note derives this result formally under perfect and monopolistic competition.

[22] 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.In 2001-2006 it is 85-87%, and about 82% in 2007-2009.

[23] The weighted average price at a higher aggregation level for this sector and source will be used for the unobserved price. The alternative procedure consists in imputing the relative price observed at the same disaggregation level for another source with a similar market share in this sector and destination. Results are not sensitive to the procedure used.

aggregation: the relative price of the composite sectoral good of the source is constructed at the 4-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. Given relative prices constructed at the 4-digit level, destination-specific weights are used to aggregate these up to the 3-digit level, and so on until the relative price for the composite good is constructed using relative prices at the 1-digit level. This improves the estimation of prices if one assumes that missing destination-specific relative prices at the 4'-digit can be approximated by the mean observed destination-specific relative price among the corresponding 4-digit group (and similarly at each aggregation level).

## III.2 Zero trade flows

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 $ij$.[24]

We assume 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 (in UN COMTRADE this threshold corresponds to 1000 USD). Such flows, if recorded, would not substantially modify the distribution of observed market shares in the destination (the left hand side of (10)) because they are an order of magnitude smaller than observed trade.

We use the same stepwise price imputation as in the case of missing unit values. This is problematic because the constructed relative price of the composite good systematically underestimates the true underlying relative price. Under model assumptions, statistically unobserved trade values must correspond to a higher cif price than the maximum observed price in the destination across all sources and sectors while by construction we postulate that unobserved relative prices in ztf sectors are equal to a weighted average relative price across sectors in which trade is observed.[25]

The assumption we make on unobserved prices would not be problematic if the underestimation factor were constant across exporters. This scalar would cancel out across sources,

[24] 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.

[25] 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.

and the estimated substitutability parameter would correspond to the true parameter. Table 3 shows that this is not the case. The share of ztf is strongly decreasing in market share, i.e. the underestimation factor is of greater magnitude for small exporters.[26] As the relative price of the composite good is underestimated by more for small exporters, 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. This means that the estimated parameter $\widetilde{\sigma}$ overestimates the true substitutability parameter $\sigma$.

Table 2: **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.

Table 3 shows that the reduction in the share of ztf proceeds at quicker pace in 1962-2009 for small exporters: the coefficient for the interaction term for the market share and year is significant and positive. Table 2 presents the predicted share of ztf for four 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. As the gap between the share of ztf for big and small exporters is reduced overtime, the overestimation bias of $\widetilde{\sigma}$ is progressively reduced.

Thus, the hypothesis we make on unobserved sectoral prices in ztf sectors does not always impede interpreting the evolution of the underlying substitutability parameter. In particular, if it is found that the estimated parameter increases in absolute value, this evolution necessarily

[26] The technical note derives formally the result that the relative price of the composite good is underestimated by more for small exporters.

provides a lower bound on the increase in the underlying substitutability parameter. This is because the overestimation bias is reduced overtime.

Table 3: **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.

Figure 7 presents the results on the evolution of $(1-\widetilde{\sigma})$ obtained when (10) is estimated on annual crossections of the COMTRADE dataset. The absolute value of trade elasticity has increased by 33% from 1962 to 2009. This corresponds to an annual increase of .6% per year.[27]

[27] The coefficient of the geometric fit is significant at 1% level.

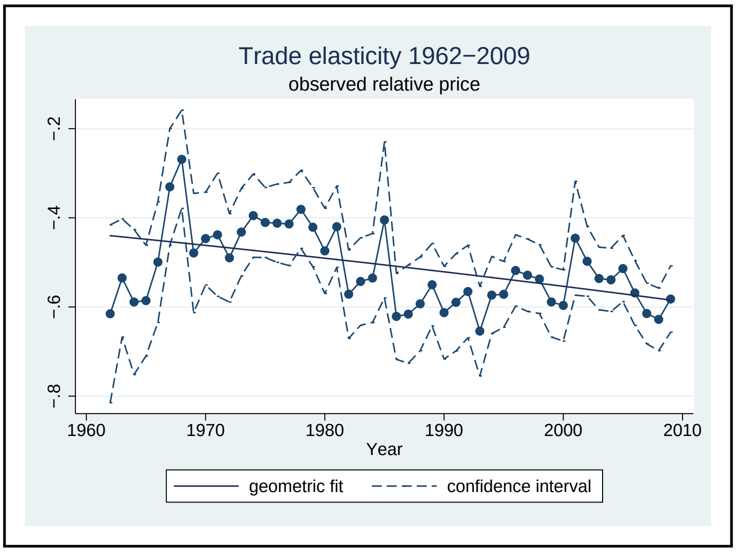



Figure 7: Estimated $(1-\widetilde{\sigma})$

## III.3 Robustness checks

### III.3.1 Changing the dataset

We provide a robustness check by estimating the evolution of the heterogeneity parameter for aggregate bilateral trade on a different dataset. We use the BACI dataset which reports bilateral trade data at the HS-1992 6-digit disaggregation level for 1995-2009. The accuracy of the relative prices of country-composite goods constructed with this dataset is improved because the harmonization procedure applied by Gaulier and Zignago (2010) in constructing BACI yields much better-quality unit values while substantially reducing the number of observations with lacking unit value. As a result, at the 6-digit level, less than 7% of total reported trade in BACI has missing unit values. This is reduced to 1-3% of total trade when the data is aggregated to the 4-digit level, as opposed to more than 10% in the raw COMTRADE data we originally used. Another advantage is that the share of ztf in BACI is stable in 1995-2009 as opposed to relatively strong fluctuations in the share of ztf overtime in our original dataset. The disadvantage of BACI is that it covers only a relatively short period compared to the years over which the distance puzzle exists. Obviously, we do not expect to reproduce exactly the results obtained with our original dataset because the trade classification and its level of aggregation are different.

Fig. 8 shows that our results hold: the elasticity parameter is found to increase in absolute value from 1995-2009. This can be compared with the equivalent period in our original dataset: the increase in the elasticity is much steeper on the BACI dataset. This finding supports the idea

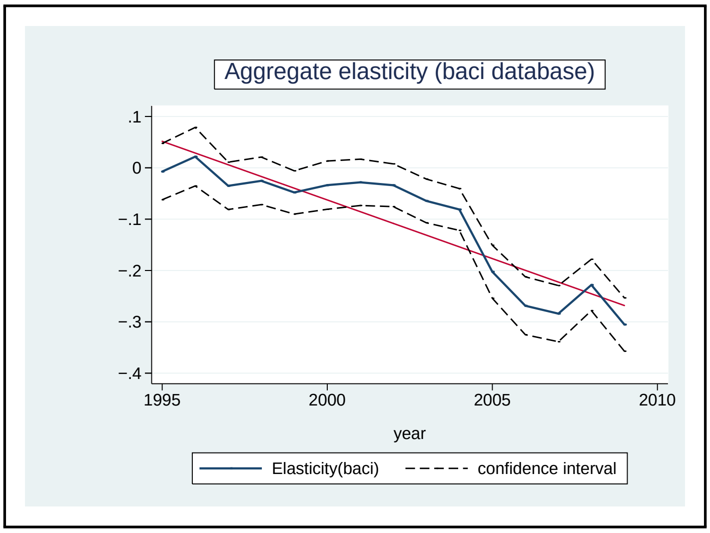



Figure 8: Estimated $(1-\widetilde{\sigma})$, BACI database

that our benchmark estimation likely provides a lower bound on the increase in the aggregate trade elasticity. However, the level of the elasticity estimated in 1995-1999 on BACI data is puzzling and suggests the existence of an attenuation bias. This is the focus of our second robustness check.

### III.3.2 Instrumenting: motivation and results

The results just presented are subject to caution if supply schedules are not horizontal.[28] The demand elasticity parameter estimated in the market share equation would then be subject to 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), the attenuation bias also impacts the evolution of the parameter.

Feenstra has developed an instrumental variables' approach which solves this problem (see Feenstra (1994), and refinements in Broda and Weinstein (2006) and Imbs and Méjean (2011)). This method exploits year-to-year variations in relative prices and market shares over 10- to 20-year estimation windows to compute the Armington elasticity. We refrain from this approach for two sets of reasons. First, Feenstra's method relies on the assumption that the elasticity

[28] Broda et al. (2006) find that supply elasticities are finite at the 4-digit level. On the other hand, Magee and Magee (2008) find that the small country assumption may hold in the data in which case there would be no attenuation bias.

parameter remains constant through time, whereas we allow the parameter to vary in each year. Second, more fundamentally, Feenstra's elasticity parameter determines short-run, marginal, longitudinal effects whereas we are interested in the elasticity parameter which determines long-run, equilibrium, cross-section outcomes. It is not immediate that these two elasticities should be the same.

We adopt a different approach to preserve the time dimension which is central to our analysis. We need 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 inputs' prices. One possibility 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 current US dollars as reported in the Penn World Tables for 189 countries in 1950-2009.[29]

The instrumenting procedure is the following. First, we compute relative prices for exporter-specific composite goods in each destination market using the stepwise price imputation procedure (see III.1). Second, for each destination market, we compute the mean evolution of GDP price levels in current US dollars of its trading partners, weighted by their market shares in this destination. This amounts to computing the evolution of the relevant real exchange rate for each specific bilateral trade relation. Third, we compute a hypothetical relative price at time $t$ for each exporter in each market as the product of its relative price at time $(t-1)$ and the evolution of its GDP price level between $t$ and $(t-1)$ relatively to all other trading partners in this destination. Fourth, we predict the relative price of each exporter in each destination at time $t$ by regressing its observed relative price on this hypothetical relative price. This gives an instrumented relative price for each exporter which depends only on its past relative price and the relative evolution of its GDP price level. Finally, we estimate (10) using these instrumented relative prices instead of the observed relative prices.

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, the instrumenting procedure would amount to little more than replacing observed prices in $t$ with lagged observed prices in $(t-1)$. We therefore

[29] 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.

also estimate (10) using as instrument the evolution of each exporter's GDP price level relatively to all other trading partners in the destination between $(t-s)$ and $t$ where $s=1,...,10$.

Results obtained with one lag ($s=1$) are shown in Fig.9. The absolute value of the substitutability parameter has increased by 13% in 1963-2009 while the level of the estimated parameter increases by 9% on average relatively to the estimate obtained with non-instrumented prices.

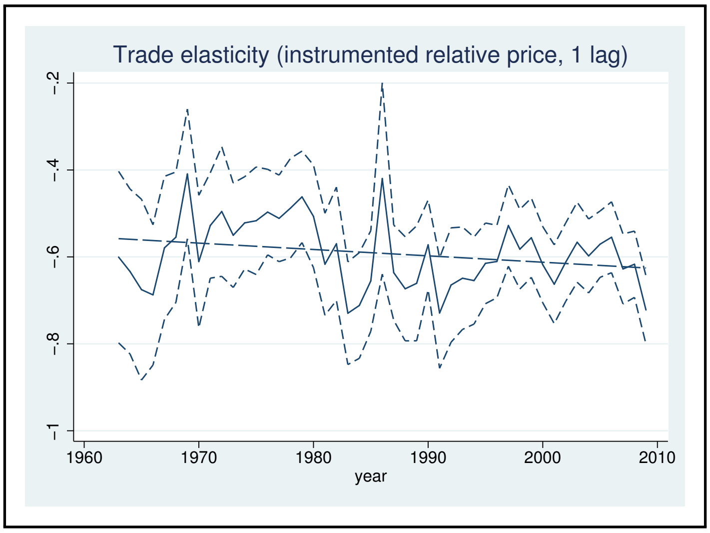



Figure 9: Estimated $(1-\widetilde{\sigma})$, instrumented relative price of composite good, 1 lag

This result is robust to increasing the number of periods in which the evolution of exports' prices is predicted with the evolution of domestic prices. Thus, in 1972-2009, the elasticity increases by 20% when the instrument is constructed with one lag, and by 23% when the number of lags is 10 (see Appendix D). 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.

The increase in the substitutability parameter is less pronounced in the instrumented specification. This result informs on the evolution of the attenuation bias. As shown by Feenstra (1994), the correlation parameter which determines the magnitude of the attenuation bias is increasing in the inverse supply elasticity and in the Armington elasticity. As the Armington elasticity increases, the reduction in the attenuation bias implies that the inverse supply elasticity decreased in 1962-2009. This finding suggests that the small country assumption has become more consistent with the data overtime.

### III.4 Is there a distance puzzle left?

This section has provided empirical evidence on the evolution of the aggregate substitutability parameter for world trade in 1963-2009. This substitutability parameter corresponds to the aggregate trade elasticity in the Armington framework. We find that this parameter has increased by 33% between 1963 and 2009 in the benchmark estimation, and by 13% when relative prices are instrumented. Both estimates are likely to be providing lower bounds on the increase in the true substitutability parameter. Section I has shown that the distance elasticity of trade has increased by 7% over the same period. Combining these two results to evaluate the magnitude of the distance puzzle redefined as increasing elasticity of trade costs to distance, 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 at least 5-7% in 1963-2009.[30] The increasing distance elasticity of trade is fully explained by increased perceived substitutability of country-specific composite goods.

The reduction in the elasticity of trade costs to distance is even more pronounced if we focus on 1970-2009. As shown by Hummels (2007), this period is characterized by a new phenomenon: the fact that air transportation starts playing a substantial role in world trade. The instrumented Armington elasticity increases by 19% in this period while the evolution of the distance elasticity is best described as flat. It follows that the elasticity of trade costs to distance has decreased by at least 17% in 1970-2009.

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.[31]

Second, composition effects may have 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 com-

[30] The elasticity decreases by 5% when the evolution of $\rho$ is computed from the ratio of trends, and by 7% when it is computed as the trend of the ratios.

[31] 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.

position 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 (2011), 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).

# Conclusion

The estimated effect of distance in gravity equations has 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 increasing 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 the baseline estimation the distance coefficient has increased by 7% from 1962 to 2009. This result 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 in the main theoretical foundations of the gravity equation 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 composite goods. It uses 4-digit unit values as proxies for sectoral prices. In our method, unobserved unit values lead to an overestimation bias that is reduced over time. As the estimated elasticity still increases in absolute value this evolution provides a lower bound on the increase in the absolute value of the underlying trade elasticity. Once instrumented, the estimated elasticity increases by 13% between 1963 and 2009. The evolution of the distance coefficient is thus compatible with a decrease of the elasticity of trade costs to distance of at least 5 to 7%. This reduction in the elasticity of trade costs to distance is even more pronounced if we focus on the period in which air transportation starts playing an important role in bilateral trade. We find that from 1970 to 2009 the elasticity of trade costs to distance has decreased by 17% while the perceived substitutability of countries' product bundles has increased by at least 19%.

## 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 (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

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

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.

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., N. Limão, and D. Weinstein (2006). Optimal tariffs: The evidence. *NBER Working Paper Series 12033*.

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.

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 (2012). What goods do countries trade? A quantitative exploration of Ricardo's ideas. *Review of Economics and Statistics (forthcoming)*.

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

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.

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.

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.

Egger, P. (2008). On the role of distance for bilateral trade. *The World Economy 31*(5), 653–662.

Egger, P., M. Larch, K. E. Staub, and R. Winkelmann (2011). The trade effects of endogenous preferential trade agreements. *American Economic Journal: Economic Policy 3*(3), 113–143.

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

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.

Gaulier, G. and S. Zignago (2010). Baci: International trade database at the product level. *CEPII working papers* (23).

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? *Unpublished manuscript, Purdue University*.

Hummels, D. (2001). Towards a geography of trade costs. *Unpublished manuscript, Purdue University*.

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

Hummels, D. and G. Schaur (2012). Time as a trade barrier. *NBER Working Paper Series 17758*.

Imbs, J. and I. Méjean (2011). Elasticity optimism. *CEPR Discussion Paper 7177*.

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

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.

Magee, C. S. and S. P. Magee (2008). The united states is a small country in world trade. *Review of International Economics 16*(5), 990–1004.

Manning, W. G. and J. Mullahy (1999). Estimating log models: To transform or not to transform? *NBER Technical Working Paper Series 246*.

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*.

Mayer, T. and S. Zignago (2011). Notes on CEPII's distances measures: the GeoDist database. *MPRA Working Paper Series 36347*.

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

Reinert, K. A. and C. R. Shiells. Trade substitution elasticities for analysis of a north american free trade area. *Staff Research Study, USITC* (19).

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. *UC Davis Working Paper 11*.

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.

# A Choice of the non-linear estimator

Three questions have to be addressed in the choice of the estimator: is its functional form consistent with the structure of the data? Does it give consistent point estimates? Is it efficient? In this appendix we provide details on why the PPML estimator used in the body of this paper is likely to perform best along these three dimensions. We also show that results obtained with the PPML in the body of the paper are not sensitive to switching to an estimator which underlying functional form takes into account the prevalence of zero trade flows in the data.

## A.1 PPML performs best in terms of the consistency-efficiency trade-off

Manning and Mullahy (1999) have shown that it is generally preferable to estimate the parameters of interest in multiplicative models directly in multiplicative rather than in log-linear form if the data is characterized by nonnegative outcomes, prevalence of zero trade flows, and skewness. This is because the expected value of the logarithm of a random variable depends on 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. Santos Silva and Tenreyro (2006) have put forward that these features characterize bilateral trade data which means that slope estimates in the log linear estimation of gravity equations will generally be biased. Considering that the most important criterion for our study of the evolution of the distance elasticity is the consistency of the parameter estimate, we opt for a non-linear estimator.[32]

The trade-off in switching to a non-linear estimator is the loss of efficiency. As shown by Manning and Mullahy (1999), all non-linear estimators entail a significant loss in efficiency relatively to OLS, and it is therefore important to choose the non-linear estimator which fits best the conditional variance function of the data. 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 $\lambda_1 = 2$. Manning and Mullahy (1999) have shown that among these alternatives, the PPML is likely to be a more efficient estimator because it

[32] OLS would give consistent coefficient estimates if the conditional variance of raw trade flows followed a Gamma distribution, i.e. was proportional to the square of the conditional mean: $V[y_i|x] = \mu(x_i\beta)^2$. But there is no a priori reason to justify this assumption.

assumes that the conditional variance of the raw data is proportional to the conditional mean. For trade data, Santos Silva and Tenreyro (2006) and Bosquet and Boulhol (2009) find that the Gamma PML estimator performs worse than PPML because it is sensitive to certain forms of measurement error, such as the rounding of the dependent variable, likely to be present in trade data. The NLS estimator is inefficient because the assumption that the conditional variance is independent of the regressors does not fit trade data.

In our cross-sectional estimation of the distance elasticity, the PPML actually gives very precise results. The loss in precision is due to the Huber-White correction we implement following Santos Silva and Tenreyro argument that the conditional variance function of the PPML is not the true conditional variance function of the data.

## A.2 Two-part and two-stage models

The remaining question is whether the functional form of the PPML is consistent with the structure of the data and particularly with the prevalence of zero trade flows.[33]

There are two possible approaches to zero trade flows. The two-stage approach, followed by Helpman et al. (2008) is to modify the theoretical model to generate zero trade flows. They achieve this in a setting with heterogeneous firms with bounded support distribution.[34] The alternative, two-part approach, followed in this paper, is to stay within the three canonical frameworks which do not allow for zero trade flows, and to treat such observations as a statistical problem due to data registration thresholds. Nil trade flows correspond to underlying positive flows which are below the data collection threshold.[35]

The two-part model deals adequately with the sample selection bias, and is consistent with the underlying model. Even if the true model had structural zero trade flows, opting for a two-part model instead of a two-stage one would not be a problem. Inteed, it has been shown by Mullahy (1986) that two-stage and two-part models give qualitatively similar results when there

[33] At the level of aggregate bilateral trade, the share of observed non-zero trade flows relatively to the number of potential trade flows if each source traded with each destination increases from 47% in 1962 to 73% in 2009. A Vuong test indicates that the number of zero trade observations is excessive and that a two-part model such as the Zero Inflated Poisson (ZIP) is preferable to the PPML.

[34] Following this approach, the equation estimated in the second stage has to be augmented with a term which controls for the pair-specific fraction of exporting firms. See Helpman et al. (2008).

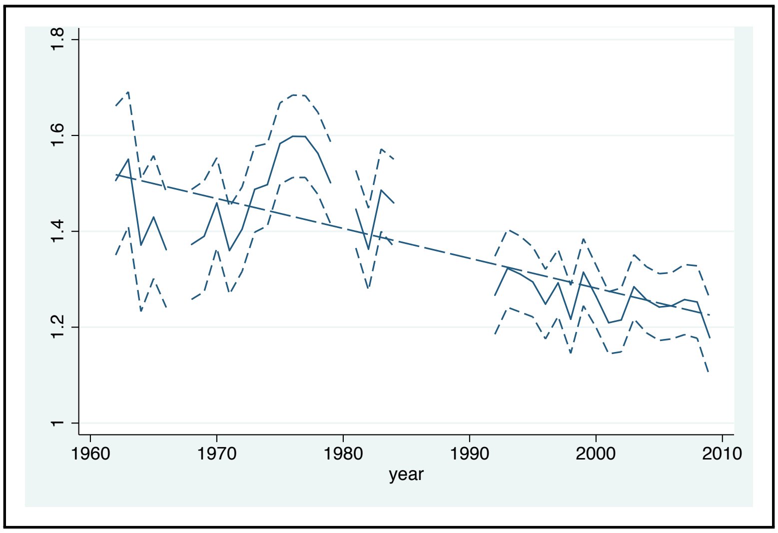
[35] In the framework with heterogeneous firms this corresponds to the assumption of no upper bound on the productivity distribution or to an upper bound on productivity combined with sufficiently low fixed costs of trade.

is clarity on the first-stage selection variable, and that the two-part approach performs better if there is uncertainty on the choice of the selection variable. In our case we consider that there is strong uncertainty on the selection variable.

In the two-part model the PPML estimator is coupled with a logit estimation of the chances of being in the zero category by doing a zero-inflated Poisson regression (ZIP).[36] The two-part model uses the same set of explanatory variables in the first step (logit) and in the second step (PPML restricted to predicted non-zeros) as the baseline PPML regression.[37]

The distance coefficient in the logit, first-step, regression has decreased, i.e. the formation of a trading relationship is less sensitive to distance (see Fig. 10).

Figure 10: Distance coefficient in the logit baseline ZIP regression (first part)




The distance puzzle seems to be specific to existing trade relationships. Conditional on trading, trade volumes have become more sensitive to distance since 1962 (see Fig. 11). The second stage ZIP regression yields very similar results to the PPML regression. This is also the case in the case when sample, composition and FTA effects are controlled for (see 4). But the ZIP

[36] For details on the ZIP specification, see Greene (2003), p. 750. The name comes from Lambert (1992). It is named 'With zeros' model by Mullahy (1986) and 'Zero-Altered Poisson' by Greene (1994). This approach is advocated in Burger et al. (2009) and in Martin and Pham (2008). For a critical approach of the ZIP, see the page of J.M.C. Santos Silva at http://privatewww.essex.ac.uk/~jmcss/LGW.html.

[37] In a recent paper, Egger et al. (2011) suggest a semi-parametric estimation of the two-part model whereby zero trade flows are considered a structural feature of the data. We perform the robustness check using the ZIP procedure because it is consistent with our treatment of zero trade flows as a statistical and not structural feature of the data.

method has an important drawback: convergence of the estimation is difficult to achieve consistently over all years, and especially in the 1980s. We conclude that the results on the evolution of the distance elasticity presented in the body of the paper are qualitatively unchanged by switching to a two-part model.

Figure 11: Comparing ZIP and PPML results

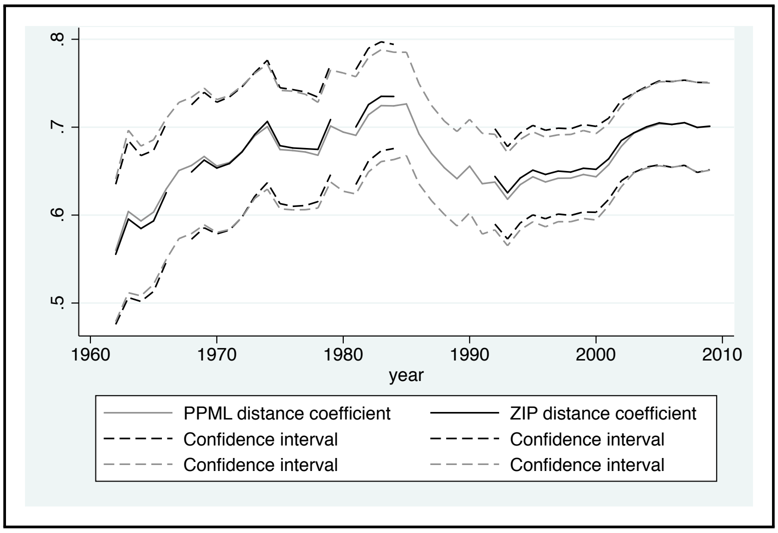



Table 4: Evolution of the distance coefficient depending on sample, composition, FTA effects

| | **% change relatively to baseline** | | **Total change 1962-2009** | |
|---|---|---|---|---|
| | PPML | ZIP | PPML | ZIP |
| Baseline | | | 1.07 | 1.08 |
| Sample effect | 7% | 5% | 1.14 | 1.14 |
| Composition effect | 7% | 6% | 1.14 | 1.15 |
| FTA effect | -54% | -52% | 0.49 | 0.52 |
| Composition + sample | 7% | 5% | 1.14 | 1.14 |
| Composition + FTA | -29% | -29% | 0.75 | 0.77 |
| Sample + FTA | -59% | -59% | 0.44 | 0.44 |
| Sample + Composition + FTA | -54% | -55% | 0.49 | 0.49 |

Results reported in the last two columns correspond to a geometric trend.

# B 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)]*

# C Full and superbalanced samples

The full sample contains 207 reporters (R) and 230 partners (P) which are listed in tables 5, 6, and 7 below. '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 also used for Yougoslavia and for Serbia and Montenegro.

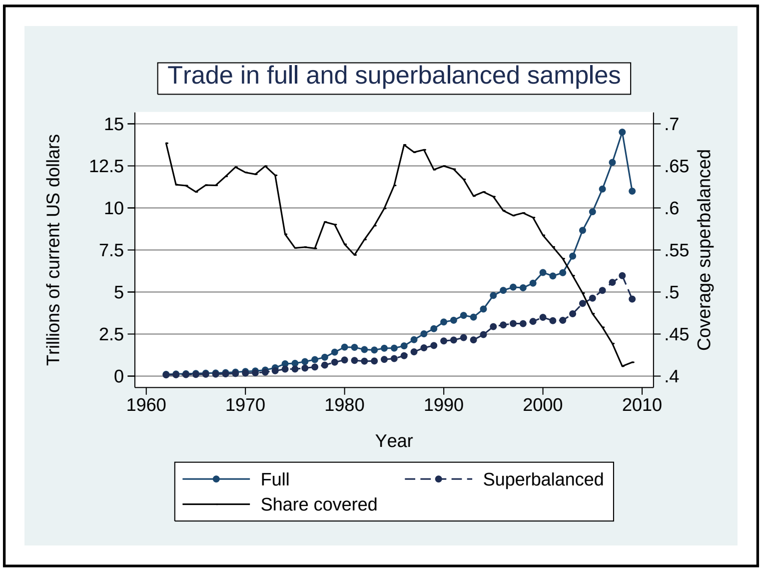



Figure 12: Trade coverage in the superbalanced sample

The superbalanced sample corresponds to the subsample of trading pairs which trade both ways in each and every year of the sample. This corresponds to 32 countries, 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.12, the superbalanced sample covers 50-70% of full sample trade. Fig.13 shows the distribution of pairs in the full sample according to the number of years they are present in the sample.

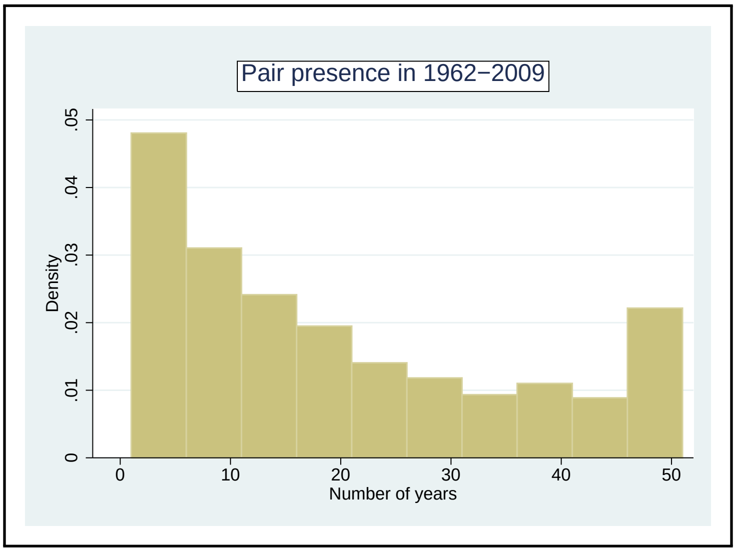



Figure 13: Number of years each pair is present in the sample

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.[38] This gives a sample of 1604 reciprocal pairs out of the 2162 possible pairs for 47 reporters (partners). Fig.14 shows that evolution of trade coverage is qualitatively similar.

Fig.15 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

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

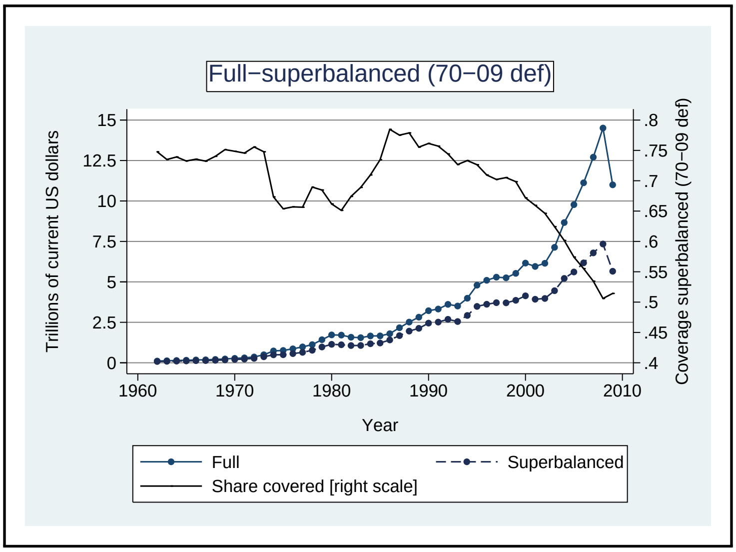



Figure 14: Trade coverage in superbalanced sample defined in 1970-2009

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.

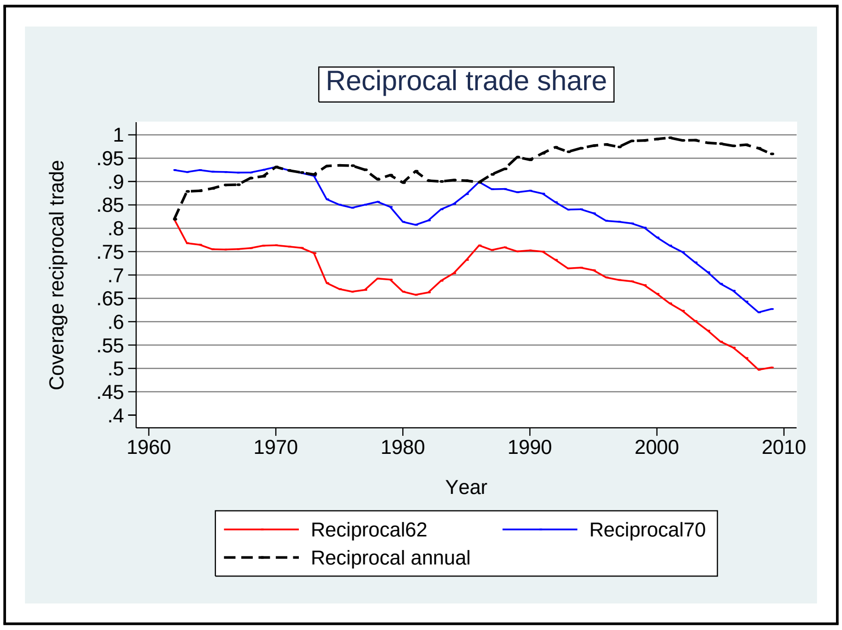



Figure 15: Trade coverage in superbalanced sample defined in 1970-2009

Table 5: 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* |

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

| Country name | Status | Country name | Status | Country name | Status |
|---|---|---|---|---|---|
| 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* |
| 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* |

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

| Country name | Status | Country name | Status | Country name | Status |
|---|---|---|---|---|---|
| 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 | | |

# D 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.

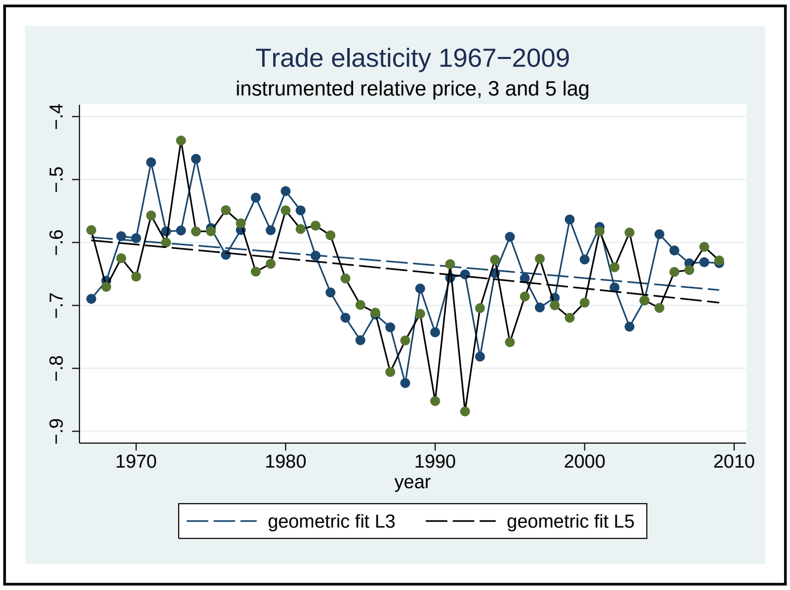



Figure 16: Estimated $(1-\widetilde{\sigma})$, instrumented relative price of composite good

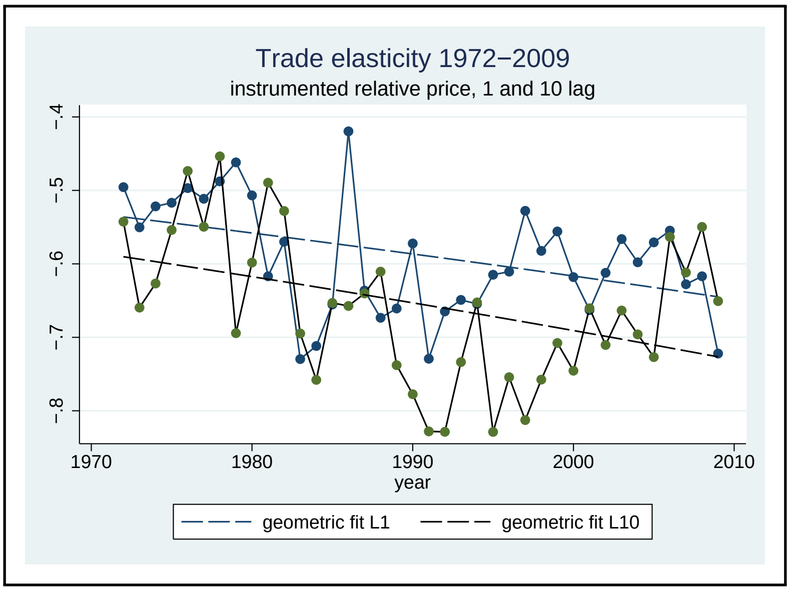



Figure 17: Estimated $(1-\widetilde{\sigma})$, instrumented relative price of composite good

Fig.(16) and Fig.(17) show that as we increase the number of lags, the evolution of the parameter becomes steeper and the level of the parameter relatively to the non-instrumented specification becomes more stable across the years. Thus, for $s = 5$, the level of the parameter increases by 20% on average relatively to the estimate obtained with observed relative prices while for $s = 10$ the level increases by 22%.

In terms of the slope, we find that for example in 1972-2009, the substitution elasticity increases by 23% when the instrument is constructed with 10 lags while it increases by 20% in the specification with just one lag. Similarly, for 1967-2009, the slope increases as we increase the number of lags from 3 to 5.