Product substitutability and the Distance Puzzle*

Elizaveta Archanskaia† Guillaume Daudin‡

April 2019

Abstract

We propose a novel explanation of the non-decreasing distance elasticity of trade over 1963-2013, commonly referred to as 'the distance puzzle'. Consistently with previous work, we find that the changing composition of world trade has contributed to making trade less rather than more sensitive to distance.

Our original contribution consists in documenting a 35% increase in the perceived substitutability of products traded on the world market over the same period. We find that the Armington elasticity of substitution increased from about 1.7 to 2.3 between 1963 and 2013. This elasticity corresponds to the elasticity of trade flows to trade costs on the intensive margin. In the Armington framework it determines the elasticity of trade flows to trade costs. The evolution of this parameter suffices to explain the non-decreasing distance elasticity of trade.

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

JEL codes: F15, N70

*This paper has greatly benefited from helpful discussions with Thomas Chaney, Anne-Célia Disdier, Peter Egger, Lionel Fontagné, Raphaël Franck, Joseph Francois, Guillaume Gaulier, Samuel Kortum, Jacques Le Cacheux, Philippe Martin, Thierry Mayer, and suggestions by seminar participants at the ETSG, the AFSE, the Ljubljana Empirical Trade Conference (Forum for Research in Empirical Internation Trade), the international workshop on the Economics of Global Interactions (Bari), at the Economics Department of SciencesPo, and at the International Trade Working Group at the University of Chicago. We thank Christian Broda, Natalie Chen, Dennis Novy, and David Weinstein for giving us access to their programs. The usual disclaimers apply.

†KU Leuven.

‡Université Paris-Dauphine, PSL University, CNRS 8007, IRD 260, LEDa, DIAL, 75016 Paris; SciencesPo, OFCE, 75007 Paris; email: guillaume.daudin@dauphine.psl.eu (Corresponding author)

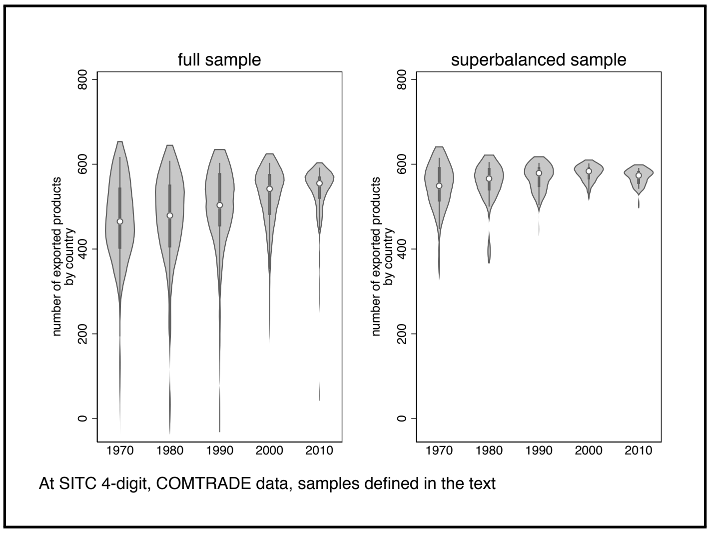
Introduction

It is a well-established fact that trade has not become less sensitive to distance over the last 60 years.

Technological developments in transportation and communication, e.g. the airplane, the container, and the internet were initially expected to lead to the "death of distance" by the end of the 20th century.[1] Yet, a larger share of world trade is conducted at short distances now than in the 1960s. This increase in the distance elasticity of trade, notwithstanding technological developments, has been dubbed the "distance puzzle". More precisely, Disdier and Head (2008) adopt a meta-analytical approach and find that distance impedes trade by 37% more in the 1990s than it did from 1870 to 1969.[2] Head and Mayer (2013) estimate the distance elasticity of trade in successive cross-sections and find that it has doubled in 1960-2005.

We investigate the empirical relevance of a mechanism that may help to rationalize the distance puzzle. We make the simple point that the flattening out of the world may go hand in hand with a persistent impeding effect of distance on trade if consumers perceive product bundles shipped out by each country to the world market as increasingly substitutable.[3]

Figure 1: Dispersion of the number of products exported by each country



Figures 1 and 2 are the simplest way to make our point. The figures illustrate that an increasing number of countries exports a larger range of products, so that the export bundle of

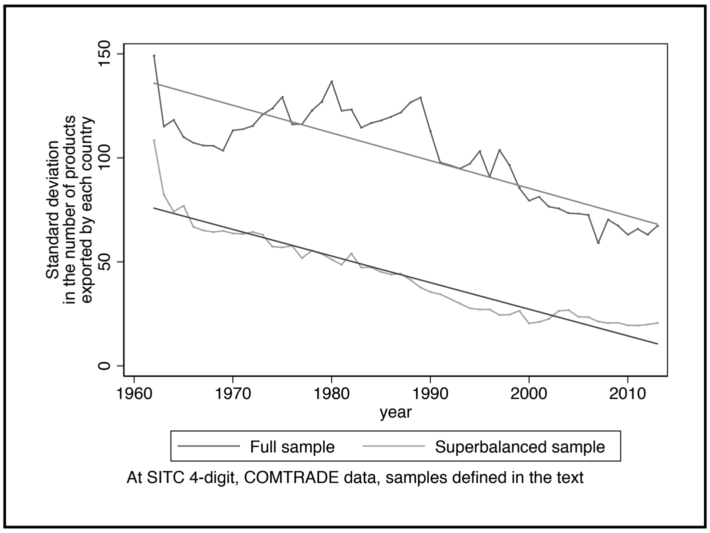
[1] Cairncross (1997); Levinson (2006); Friedman (2007)

[2] See also Berthelon and Freund (2008); Combes et al. (2006); Brun et al. (2005); Buch et al. (2004).

[3] Equation (4) in Head and Mayer (2013): the distance effect is a product of two elasticities, the elasticity of trade to trade costs and the elasticity of trade costs to distance. We investigate the evolution of the former.

all countries is converging toward the entire range of goods. The figures also illustrate that the increasing similarity of export bundles is robust to the examination of a more limited sample of exporters.[4]

Figure 2: Standard deviation in the number of products exported by each country



The increasing similarity of product bundles corresponds to an increasing number of varieties of any given product that is available in the market. A more densely populated product space may translate into an increase in the perceived substitutability of product varieties, as well as into an increase in the perceived substitutability of product bundles. We work with the canonical Anderson and van Wincoop (2003) framework to quantify the contribution of this mechanism. We find that the non-decreasing distance elasticity of trade can be explained by the increase in the perceived similarity of product bundles of different national origin.

Prominent explanations of the distance puzzle

Recent work rationalizes the distance puzzle in three complementary ways: by pointing out a possible misspecification of the econometric model, by refining the specification of the trade cost function, and, more recently, through the lens of network analysis.

The first strand of the literature investigates the incidence of the estimation method on the distance puzzle. Santos Silva and Tenreyro (2006) advocate estimating the gravity model in multiplicative form using a specific non-linear estimator, the Poisson Pseudo Maximum Likeli-

[4] The full sample comprises all pairs. The superbalanced sample focuses on pairs which trade both ways in all years. See Appendix A.

hood (PPML). Contrary to the canonical loglinear approach, this estimator provides consistent estimates and is robust to rounding error and overdispersion which are both likely features of trade data.[5] The magnitude of the distance puzzle is reduced when the gravity model is estimated in multiplicative form (Bosquet and Boulhol (2015); Head and Mayer (2013)).

The sensitivity of the distance puzzle to the estimation method is likely due to sample composition effects (Head and Mayer (2013); Larch et al. (2016)). Indeed, 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 (2011)). If trade relations have in priority been established between small and distant partners, the reduction in the number of zeros may have gradually reduced the underestimation of the distance coefficient in the loglinear specification. [6] This explanation echoes Felbermayr and Kohler (2006) who pointed out that the log-linear specification was subject to sample selection bias due to the exclusion of zero trade flows. They conjectured that the distance puzzle was an artefact of reduction in this bias through the extensive margin of trade.[7]

The second and most prominent strand of the literature singles out the underpinnings of the trade cost function as key to understanding the distance puzzle. The basic point formulated by Buch et al. (2004) is that the distance elasticity of trade is invariant to reductions in transportation and communication costs if their distribution over distance remains unchanged.[8] Furthermore, while the distance elasticity of transport costs may have decreased (Hummels (2007)), other cost components, such as delays, may have become more distance-elastic (Hummels and Schaur (2013)). More generally, as argued by Head and Mayer (2013), if freight costs account for an ever smaller fraction of distance-dependent trade costs, the distance elasticity of trade will be determined by other, possibly persistent, cost components.[9]

Krautheim (2012) presents a complementary mechanism in the heterogeneous firms' frame-

[5] Santos Silva and Tenreyro (2011) and Fally (2015) provide evidence on desirable properties of the PPML. Head and Mayer (2014) review properties of alternative estimators. See also Bosquet and Boulhol (2015, 2014).

[6] Larch et al. (2016) attribute the puzzle to the growing bias of the OLS estimates. Using a nonlinear estimator and controlling for the number of exporting firms, they find a decreasing distance elasticity in 1980-2006.

[7] This discussion leaves open the question of the estimator which correctly captures the level of the distance elasticity. Head and Mayer (2013) argue that PPML gives little weight to small trade flows, characteristic of more distant partners. For Santos Silva and Tenreyro (2006) small flows are more prone to measurement error.

[8] Egger (2008) argues that the marginal effect of distance on trade is fundamentally non uniform across trade partners and decreasing in the level of bilateral trade. This nonlinearity reinstates increased openness as a possible explanatory factor because a uniform reduction in trade costs has non uniform effects on bilateral trade.

[9] Head and Mayer (2013) propose a typology of persistent but unobserved trade costs. Daudin (2003, 2005) put forward that trade costs may have remained stable as a share of value added.

work. He models the informational component of trade costs as a fixed cost which decreases in the number of exporting firms. This refinement of the trade cost function magnifies the distance elasticity of trade because the number of exporters is decreasing in variable trade costs which increase with distance. This magnification mechanism may have been reinforced by the increasing weight of information costs in total fixed costs.

An alternative explanation put forward in models with input-output linkages is that the relationship between total trade costs and transport costs may be non-monotonic. An increasing distance elasticity may be an endogenous outcome of transport cost reductions if they engender a reoptimization of the production process which ends up increasing the relative cost of long-distance trade. One possible mechanism is trade cost magnification through multiple border crossings by goods as a consequence of increased production fragmentation (Yi (2010); Daudin et al. (2011); Noguera (2012)). Another mechanism formalized by Duranton and Storper (2008) works through quality upgrading. Lower transport costs shift trade towards higher-quality inputs which are more distance-sensitive because their customization requires intensive communication, i.e. more back-and-forth travelling, between upstream and downstream firms.

A shift of focus to the trade elasticity

The focus of the literature on the shape of the trade cost function mirrors the expectation that the distance coefficient moves together with the elasticity of trade costs to distance. But Chaney (2018) provides a theoretical foundation for the gravity equation through the lens of network analysis which demonstrates that the distance coefficient can be invariant to the trade cost function. In Chaney (2018) the rate of distance decay in aggregate trade is linked to the rate of decay in the density of firms which cover that distance with their network of contacts. As the geographic dispersion of the network is increasing in firm size, the shape parameter of the firm size distribution plays a key role in explaining movements in the distance coefficient. Thus, technological advances in transportation increase the geographic dispersion of exports at the level of the firm but have no incidence on the distance elasticity of aggregate trade as long as the stationary firm size distribution verifies Zipf's law.

The link between the distance coefficient and the parameter which captures the degree of structural heterogeneity in the economy is not specific to Chaney (2018). Every theoretical foundation of the gravity model delivers a functional relationship of the distance elasticity with the intensity of the incentive to trade, i.e. the degree of structural heterogeneity in some model-specific dimension. The combination of empirical evidence on the changing shape of the trade cost function with evidence on the stability of the distance distribution of trade indicates that structural heterogeneity may have contributed to the evolution of the distance coefficient. However, empirical evidence on the evolution of structural heterogeneity in the economy since the 1960s is notoriously scarce (Head and Mayer (2013)).

We pursue the idea that a key parameter for understanding movements in the distance coefficient is the one measuring the degree of structural heterogeneity. Following Arkolakis et al. (2012) we refer to this parameter as the 'trade elasticity'. Because of data limitations inherent to our 60-year perspective, we can only estimate the trade elasticity in the Armington framework. Structural heterogeneity is determined in this framework by perceived product substitutability, i.e. the degree of product differentiation by place of origin.[10]

We estimate the distance elasticity of aggregate trade and the Armington elasticity of substitution between goods of different origin in each year between 1963 and 2013. These two elasticities are identified separately. We deduce the implied evolution of the elasticity of trade costs to distance from the two estimated elasticities.[11] Our main result is that the increase in the Armington elasticity not only rationalizes the non-decreasing distance elasticity of trade but also hints at a reduction in the elasticity of trade costs to distance. We conclude that in the Armington framework the distance puzzle is fully explained by the increasing sensitivity of consumers to price differences. More generally, our results suggest that the distance puzzle may be due to a reduction in structural heterogeneity.

To the best of our knowledge, Berthelon and Freund (2008) is the only paper that studies the impact of changes in perceived product substitutability on the distance coefficient. Using estimates of sectoral Armington elasticities obtained by Broda et al. (2006), Berthelon and Freund (2008) find a positive relationship between the variation in sectoral distance coefficients and the variation in Armington elasticities between 1985-1989 and 2001-2005.[12] Our approach is different from Berthelon and Freund (2008) because we focus on the aggregate Armington elasticity and provide direct estimates of this parameter in each year between 1963 and 2013. Our approach is different from Broda et al. (2006) because we exploit cross-sectional variation in

[10] This parameter plays no role in determining the trade elasticity in the Melitz-Chaney or Eaton and Kortum frameworks. The demand parameter comes back into the picture under alternative assumptions on the productivity distribution (Bas et al. (2017), Feenstra (2018)).

[11] Erkel-Rousse and Mirza (2002) do this exercise for just one point in time on a subsample of world trade flows.

[12] Berthelon and Freund (2008) work with 776 sectors defined at the SITC Rev.2 4-digit level. Broda et al. (2006) use time-series variation in prices and market shares for the set of exporters to the US market to get one value for the Armington elasticity in 1972-1988 and another value in 1990-2001.

prices and trade shares to infer the extent of substitutability while these authors exploit changes in prices and trade shares over time to identify one point estimate. Our approach enables us to trace out changes in the perceived similarity of the product mix that countries supply to the world market.

We proceed in three steps. First, we examine the hypothesis that the distance puzzle is a by-product of compositional changes in the set of traded goods or of trading pairs. We refute this hypothesis by showing that the distance puzzle is more pronounced in the stable set.

Second, we point out that a straightforward nonlinear estimation approach allows identifying the aggregate Armington elasticity in cross-section. The intuition follows Imbs and Méjean (2015) who show that a consistent estimate of the aggregate elasticity is obtained on sectoral data by constraining sectoral elasticities to equality in the estimation.[13] Our estimation approach differs from Imbs and Méjean (2015) - who generalize Feenstra (1994) - in that these authors use price and market share variation over time to estimate a single elasticity. We relax the assumption that the parameter is time invariant and work with the cross-sectional price distribution.

Third, we use non linear least squares on pooled sectoral data aggregated in a theoretically sound way to provide estimates of the Armington elasticity in each year between 1963 and 2013. We then carry out a series of robustness checks to deal with missing unit values, zero trade flows, low data quality and endogeneity. We document an increase of between 22% and 60% of the Armington elasticity.

Finally, we discuss the implications of the increasing Armington elasticity for the distance puzzle. The Armington elasticity has increased faster than the distance elasticity of aggregate trade between 1963 and 2013. The evolution of the distance coefficient is thus compatible with a reduction of the elasticity of trade costs to distance.

1 The magnitude of the distance puzzle

In this section we evaluate the sensitivity of the distance puzzle to compositional changes in the set of traded goods and in the set of trading pairs over 1962-2013. Such changes were identified in previous estimations of the loglinearized gravity model as explanatory of movements

[13] The focus of Imbs and Méjean (2015) is on documenting the heterogeneity bias, i.e. that the estimated aggregate elasticity differs from the 'true' elasticity defined as the weighted average of the sectoral parameters. Our focus is on obtaining an estimate consistent with the aggregate trade data.

in the distance coefficient.[14] We estimate the gravity model in multiplicative form and document that the distance puzzle is magnified in the sample of stable pairs and robust to fixing the sectoral composition of world trade.

1.1 Baseline estimate of the distance puzzle

We use the COMTRADE dataset to make our investigation of the distance puzzle directly comparable to Head and Mayer (2013) and Berthelon and Freund (2008). We work with the 4-digit SITC Rev.1 product classification (600-700 goods) because it provides the longest and most comprehensive coverage of disaggregate bilateral trade (1962-2013). Data on bilateral distance, bilateral trade cost 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.[15]

We conduct the estimation on CIF import flows. We restrict the sample to trade in goods which are attributed to specific 4-digit categories and to pairs for which we have data on bilateral trade cost controls. App. A lists the resulting set of countries. For each active pair attributed sectoral flows are summed to obtain total bilateral trade. We refer to the resulting sample as 'the full sample'. It covers between 88% and 99% of reported trade in COMTRADE.

We follow the canonical Anderson and van Wincoop (2003) derivation of the gravity model to express aggregate bilateral trade X_{ij} as a function of bilateral trade barriers τ_{ij}, multilateral trade resistance terms in source i and destination j (resp. Π_i and P_j), and nominal incomes Y_n with $n \in \{i, j, w\}$ where w is world income.[16]

$$X_{ijt} = \left(\frac{Y_{it}Y_{jt}}{Y_{wt}}\right)\left(\frac{\tau_{ijt}}{\Pi_{it}P_{jt}}\right)^{\epsilon_t} \tag{1}$$

We include the time subscript t on the elasticity of trade flows to trade costs ϵ in (1) to underline that this parameter is subject to change. The goal of this paper is to quantify the magnitude of this change. In the Armington framework the elasticity of substitution between goods of different national origin, σ_t, determines the level of the trade elasticity: $\epsilon_t = 1 - \sigma_t$. We pin down the evolution of the perceived substitutability of country-specific composite goods over 1963-2013, thereby quantifying the evolution of the trade elasticity in the Armington framework.

[14] Berthelon and Freund (2008) find that changes in the sectoral composition of world trade do not help to explain movements in the distance coefficient in 1985-2005. Head and Mayer (2013) find that the distance puzzle is reduced in the set of stable pairs between 1960 and 2005 when the model is estimated in loglinear form.

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

[16] This formulation is not specific to the Armington framework. See footnote 20 in Eaton and Kortum (2002) and subsequent discussions of equivalence in Arkolakis et al. (2012); Head and Mayer (2013).

As total bilateral trade costs τ_{ijt} are not directly observed for each pair and year, we model them as a function of observable time-invariant bilateral controls which are distance, adjacency, and common language together with persistent but time-varying controls standard in the gravity literature which are historical and current colonial linkages as well as belonging or having once belonged to the same country. We include an unobserved bilateral trade cost component ν_{ijt} assumed to have mean zero conditional on the observables.[17] We denote distance $\eth_{ij}$, group the other time-invariant observables in the vector Z and time-varying observables in the vector S_t to get the following specification of the trade cost function:

$$\tau_{ijt} \quad = \quad \exp\left\{\rho_t \ln \eth_{ij} + Z'\zeta_t + S_t{}'\varsigma_t + \nu_{ijt}\right\} \tag{2}$$

Replacing (2) in (1), substituting source and destination specific variables with country fixed effects (resp. f_{it} and f_{jt}), defining a constant ξ_t and specifying a multiplicative error term ξ_{ijt} which includes the exponentiated unobserved bilateral trade cost gives the equation to be estimated on aggregate bilateral trade:

$$X_{ijt} \quad = \quad \exp\left(\xi_t - \delta_t \ln \eth_{ij} + Z'\zeta_t + S_t{}'\varsigma_t + f_{it} + f_{jt}\right)\xi_{ijt} \tag{3}$$

The parameter of interest is the distance elasticity, $-\delta_t$, which corresponds to the product of the distance elasticity of trade costs ρ_t and of the trade elasticity ϵ_t.[18] The estimation is conducted in cross section. To ensure consistency of the point estimates, we do not loglinearize the model although working with a non-linear estimator may entail a loss of efficiency (Manning and Mullahy (1999)). Following Santos Silva and Tenreyro (2006), we opt for the PPML estimator.

Figure 3 gives the results for the full sample. The distance elasticity δ_t increased by 4.5% between 1962-2013 but this increase is only marginally significant.

Consistently with Head and Mayer (2013) and Bosquet and Boulhol (2015), we find that the distance sensitivity of trade is best described as stable when the gravity model is estimated using the PPML.

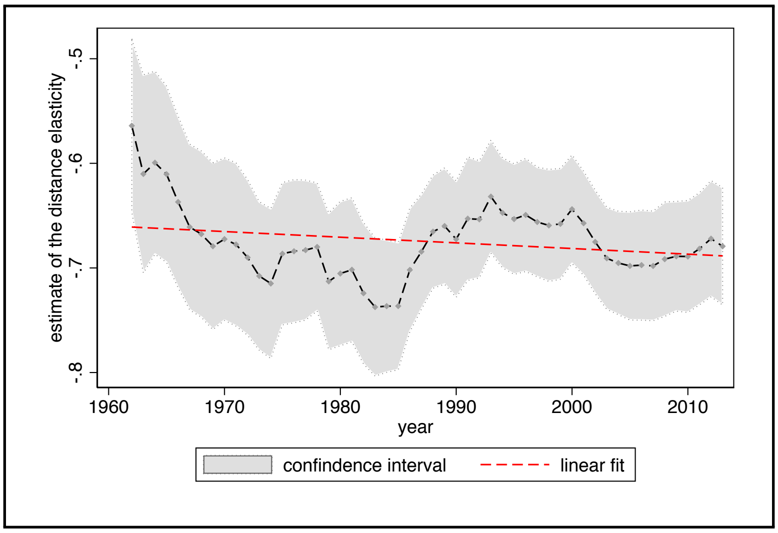
The large confidence intervals shown in Figure 3 are linked to the implementation of the Huber-White correction of standard errors.[19] Standard errors are reduced by an order of magnitude without this correction. Considering that we are looking at these point estimates as if

[17] The error term contains bilateral variation in trade costs due to trade policy. The question of possible changes in the distance distribution of trade costs as a consequence of policy decisions is beyond the scope of this paper.

[18] We follow the notation in Head and Mayer (2013) (see equation (4), p.1205).

[19] Switching to the panel approach with the full set of country-year dummies does not achieve significant improvement in precision.

Figure 3: Baseline estimate of the distance puzzle



they were descriptive statistics on the whole population, we take the results shown in Figure 3 as sufficient evidence for the existence of the distance puzzle.

1.2 The magnitude of the sample composition effect

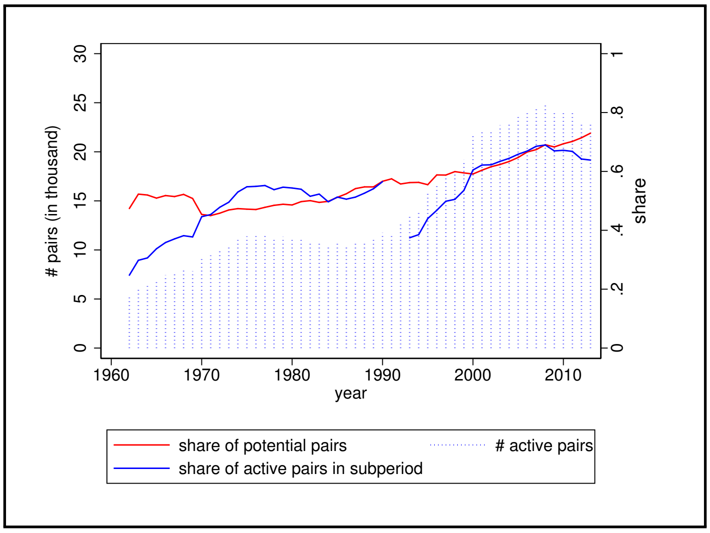
Mayer et al. (2019) and Head and Mayer (2013) argue that the main consequence of choosing to estimate gravity equations in multiplicative rather than in loglinearized form is the relative weight given to small and big trade flows. The loglinearized model picks up a stronger increase in the distance elasticity of trade, mainly because of the increasing number of new low volume long-distance trade relationships (Head and Mayer (2013)).

Consequently, in this section we first evaluate the magnitude of sample composition effects and second check to what extent our PPML estimates of the distance elasticity are sensitive to changes in the set of trading pairs.

Figure 4 summarizes the main features of the country coverage of the data. The number of active pairs increases more than fourfold in 1962-2013 (in dash, left scale), both because more countries report trade to COMTRADE and because more pairs have non-zero trade flows (Helpman et al. (2008)). Active pairs make up between 45% and 73% of the total number of possible trade relationships, with a clear upward trend (in red, right scale).

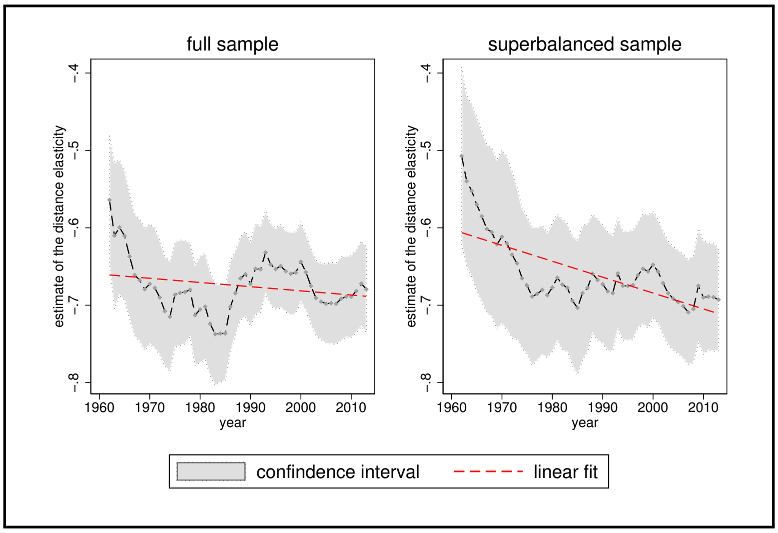
If we focus on the set of pairs that report non-zero trade in at least one year, the share of active pairs increases by 22 percentage points between 1962 and 1990 and by 27 percentage points between 1993 and 2013 (in blue, right scale). [20] By the end of the sample about 2/3 of pairs which trade at least once between 1962 and 2013 are reporting non-zero trade.

Figure 4: Active pairs in COMTRADE (1962-2013)



Hence, sample composition effects are substantial. Nonetheless, the bulk of total trade is attributable to the 786 pairs which trade both ways in every year. We refer to this set of stable reciprocal pairs as the 'superbalanced' sample and use it to investigate the magnitude of the sample composition effect.[21]

Figure 5: Sample composition effect on the distance puzzle



---

[20] We split the sample in two subperiods, 1962-1990 and 1993-2013, to take into account country creation and disappearance in the early 1990s.

[21] The superbalanced sample includes 32 countries listed in App. A. The appendix discusses trade coverage.

In the superbalanced sample (right pane) the increase in the distance effect δ_t is magnified to 18.2%, and this increase is strongly significant.[22] The distance puzzle is exacerbated in the sample of stable trade relationships.[23] We take the results shown in Figure 5 as more evidence for the existence of the distance puzzle.

1.3 The magnitude of sectoral composition effects

In this section we evaluate the sensitivity of the distance elasticity to changes in the sectoral composition of trade in two complementary ways. The first exercise consists in fixing the sectoral composition of world trade. The second exercise consists in fixing the sectoral composition of the bundle supplied by each exporter to the world market.

In the first exercise we fix the sectoral composition of total trade to the initial year of the sample. Denoting each 4-digit sector k and the annual share of the sector in world trade (w,t) by $s^k_{w,t}$, the reweighting procedure fixes the share of each 4-digit sector in world trade to its share in 1962. The reweighted sectoral bilateral flow is $\tilde{X}^k_{ijt} = X^k_{ijt} * \frac{s^k_{w,1962}}{s^k_{w,t}}$. The reweighted sectoral flows are summed for each pair, and the gravity equation is estimated in each year for aggregate bilateral trade.

Results are shown in Figure 6. The evolution of the distance coefficient becomes much more linear in the full sample (left pane), and this exacerbates the distance puzzle to a 14.5% increase in δ_t. The impact of this reweighting procedure is about nil in the sample of stable pairs (right pane) where δ_t increases by 18.4%.[24] Indeed, the main incidence of fixing the sectoral composition of world trade is the elimination of short-term fluctuations in the distance coefficient due to fluctuations in the weight of the energy sector. As this sector plays a relatively minor role in trade of stable reciprocal partners, the reweighting procedure has little incidence on the distribution of trade over distance in this sample.

In the second exercise we fix the composition of the bundle supplied by each exporter i to the world market. Denoting the annual share of the sector in world imports from i by $s^k_{i,t}$, the reweighting fixes the share of each 4-digit sector in world imports from i to its share in 1962. The reweighted sectoral bilateral flow is $\tilde{X}^k_{ijt} = X^k_{ijt} * \frac{s^k_{i,1962}}{s^k_{i,t}}$. The resulting sectoral flows are

[22] The annualized growth rate is 0.09% in the full sample and 0.33% in the superbalanced sample.

[23] The opposite result is documented by Head and Mayer (2013) for the full and balanced samples when the model is estimated in logs. The likely explanation is that the reduction in the distance elasticity due to the elimination of small trade flows between distant partners trumps the increase in the distance elasticity of stable trade relationships.

[24] The annualized growth rate is 0.26(0.33)% in the full (superbalanced) sample. Confidence intervals are reported for the Huber-White correction of standard errors. Without it standard errors are reduced by a factor of 10.

summed for each pair, and the gravity equation is estimated on total reweighted bilateral trade.

Figure 6: Product composition effect: fixing the world bundle

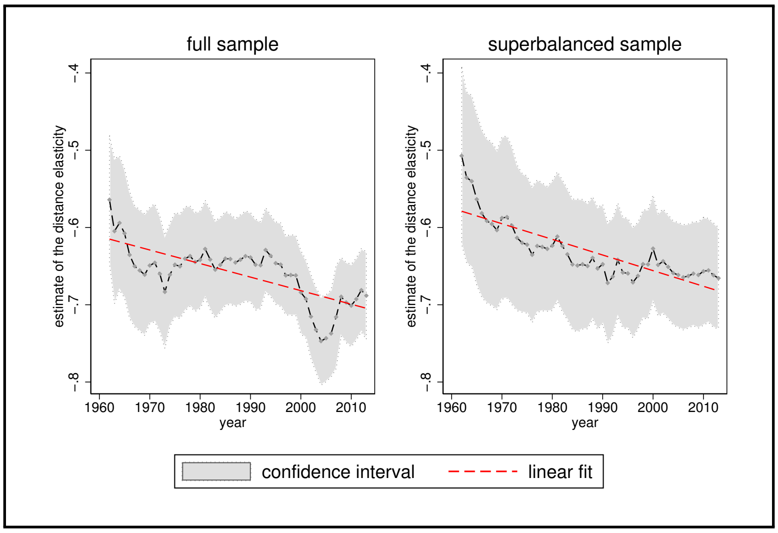
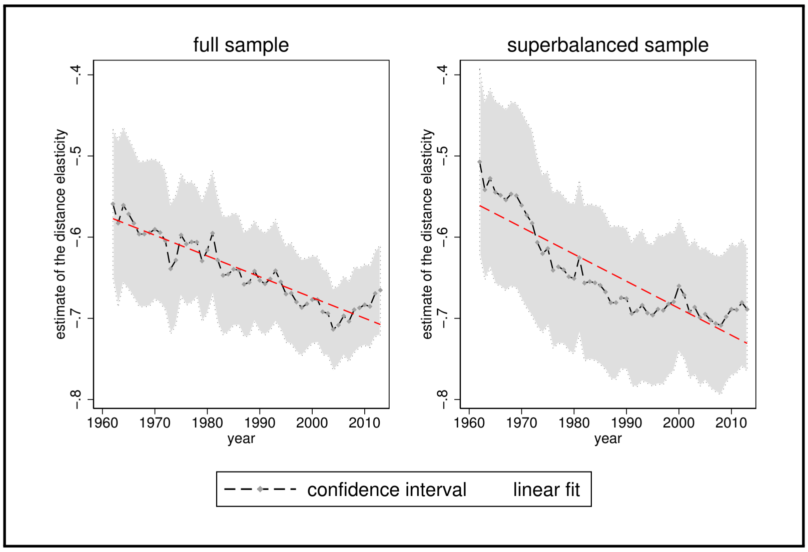


Figure 7: Product composition effect: fixing the country bundle



Results are shown in Figure 7. The magnitude of the distance puzzle is amplified when we fix the composition of the country-specific composite good. Although the degree of precision in the estimation of the gravity equation with the Huber-White correction of standard errors remains similar to the benchmark specification, 87% of the variation in the distance coefficient is here attributable to the time trend, against just 7% in the baseline specification for the full sample.[25] The corresponding annualized growth rate is .40% in the full, and .54% in the stable sample. This corresponds to a 22.7% increase in the distance sensitivity of trade between 1962 and 2013 in the full sample, and to a 31.4% increase in the sample of stable pairs.

We conclude that short-term fluctuations in the estimated distance elasticity δ_t are likely attributable to product and sample composition effects. At the same time, the long-term evolution of the distance elasticity appears linked to structural changes. These structural changes are more pronounced in the set of stable trade relationships.

1.4 Summing up: the robustness of the distance puzzle

Table 1: Evolution of δ_t: sample, composition and FTA effects

	FULL			STABLE		
	rate (%)	R-sq	total change	rate (%)	R-sq	total change
Baseline	.09*	.07	1.045	.33***	.49	1.182
World bundle	.26***	.53	1.145	.33***	.68	1.184
Country bundle	.40***	.87	1.227	.54***	.77	1.314

Note: Estimated annualized growth rates reported in col.2 and col.5 are obtained as a geometric fit on the basis of annual point estimates of the distance coefficient in 1962-2013. Col.3 and col.6 report the share of time variation in the point estimate explained with the annualized growth rate.

Table 1 summarizes our findings. The distance puzzle is magnified in the sample of stable pairs and robust to fixing the product composition of world trade. Hence, the non-decreasing distance elasticity of trade is likely to be a structural outcome rather than an artefact of composition effects. These results motivate our focus on structural heterogeneity as a possible alternative explanation of the non-decreasing distance elasticity.

2 Interpreting the distance coefficient

In the three canonical microfoundations of the gravity model, the distance coefficient δ is the product of two elasticities: the elasticity of trade costs to distance ρ and the elasticity of trade flows to trade costs ϵ. The question addressed in this paper is very simple. Have trade flows become more sensitive to trade costs (increasing $|\epsilon|$), or have trade costs become more sensitive to distance (increasing ρ)? The three main microfoundations of the gravity model

[25] In the superbalanced sample 77% of the variation is attributable to the time trend when the product bundle is fixed against 49% when the product bundle is not fixed.

of trade give structurally different interpretations to ϵ but not to ρ. In this section, we briefly discuss the interpretation of the structural parameter ϵ in each model and review the evidence on its evolution.[26] We then provide details on the procedure used in section 3 to pin down the evolution of ϵ in one of these canonical microfoundations, namely the Armington framework.

2.1 What does the trade elasticity ϵ actually measure?

In Eaton and Kortum (2002) the heterogeneity dimension captured by the trade elasticity ϵ 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. $|\epsilon|$ is decreasing in the degree of intersectoral productivity dispersion. If sectoral productivity draws become more similar, trade flows become more responsive to trade costs.

In the Melitz (2003)-Chaney (2008) framework the heterogeneity dimension captured by the trade elasticity ϵ is intrasectoral. Consumers value firm-specific varieties of goods which they acquire in monopolistic competition markets. $|\epsilon|$ is decreasing in the degree of intrasectoral productivity dispersion. If firm productivity draws become more similar, trade flows become more responsive to trade costs.[27]

In Anderson and van Wincoop (2003) there is no heterogeneity in productive efficiency, and product markets in each country are competitive. So production costs are equalized across goods within a country. The heterogeneity dimension comes from the assumption that consumers perceive goods of different national origin as intrinsically imperfect substitutes. Each country is specialized in production of country-specific goods which on aggregate give a country-specific composite good. $|\epsilon|$ is increasing in the perceived substitutability of goods of different origin. The trade elasticity ϵ equals $1-\sigma$, where σ denotes the lower tier Armington elasticity of substitution.[28]

The defining feature of the trade elasticity is that it captures the extent of structural heterogeneity in the economy. Specific functional form assumptions in the canonical derivations of the gravity equation deliver the result that the trade elasticity depends either on the dissimilarity of

[26] We refer the reader to Head and Mayer (2014) for a comprehensive review of theories underpinning structural gravity and a meta-analysis of available estimates of the trade elasticities.

[27] Gabaix et al. (2016) study transition dynamics in random growth processes which stationary distributions follow a power law in the upper tail. They show that shocks to the parameters of the random growth process translate very slowly into changes in the exponent of the power law. Speeding up the transition requires deviations from Gibrat's law, i.e. allowing the parameters of the random growth process to depend on firm types.

[28] In classical work on the price sensitivity of import demand, 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 (1991), Saito (2004), and Feenstra et al. (2018).

production technologies or on the dissimilarity of goods as perceived by consumers. The trade elasticity depends simultaneously on supply- and demand-side heterogeneity under less restrictive assumptions.[29] The bottom line is that trade flows become more responsive to trade costs in more homogeneous economies. The open question is whether trading partners' economies have become more homogeneous since the 1960s, and if so, whether shocks to structural parameters can be identified with the trade data at hand.

2.2 What do we know about the evolution of the trade elasticity?

Recent work on the trade elasticity has been mainly concerned with the level of the aggregate trade elasticity and with the heterogeneity underpinning the aggregate parameter.[30] Evidence on the evolution of either sectoral or aggregate trade elasticities is scant. Data constraints play a role in explaining this situation. An additional explanatory factor is the focus on counterfactual analysis for which stability in structural parameters is the standard assumption.

On the supply side, there is no direct evidence on the evolution of the structural parameter that captures technological dissimilarity in production either between sectors or between firms. Levchenko and Zhang (2016) is the only paper that tracks technological dissimilarity at the intersectoral level over a 50 year horizon. Levchenko and Zhang (2016) do not estimate the structural heterogeneity parameter in the Eaton and Kortum (2002) framework, but they do find evidence of within-country convergence in sectoral knowledge stocks over 1960-2010.[31] Increasing similarity of sectoral production technologies would translate into an increasing trade elasticity $|\epsilon|$ in the Ricardian framework.

Andrews et al. (2016) track technological dissimilarity at the intrasectoral level over a 25 year horizon. Andrews et al. (2016) do not estimate the structural heterogeneity parameter in the Melitz (2003)-Chaney (2008) framework, but they do find evidence of divergence in labour and multifactor productivity over 1997-2014 in a sample covering 24 OECD countries. Increasing dissimilarity of firm production technologies would translate into a decreasing trade elasticity $|\epsilon|$ in the heterogeneous firms' framework. The difficulty in interpreting Andrews et al. (2016)' findings in terms of structural heterogeneity resides in the fact that measured productivity

[29] See for example Feenstra (2018) and Bas et al. (2017), as well as discussion in Head and Mayer (2014).

[30] On the level of the trade elasticity, see the discussion in Allen et al. (2019) as well as Head and Mayer (2013, 2014); on the heterogeneity underpinning the aggregate trade elasticity, see for example Feenstra et al. (2018), Bas et al. (2017), Imbs and Méjean (2015).

[31] Different theoretically grounded methods have been used to estimate the trade elasticity in the Ricardian framework, but only in a specific year. See Eaton and Kortum (2002); Simonovska and Waugh (2014); Costinot et al. (2012); Caliendo and Parro (2015).

dispersion increases mechanically in the extent of market integration in a sample which includes exporting and non-exporting firms.[32] Autor et al. (2017) document the increasing importance of foreign sales in the sample of the 500 biggest U.S. firms over 1982-2012.[33] Hence, the findings of Andrews et al. (2016) do not suffice to conclude that firm production technologies have become more dissimilar since the mid-1990s.

On the demand side, evidence on the evolution of lower-tier Armington elasticities of substitution measured on aggregate data is scarce. 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 et al. (2006) provide evidence on the evolution of Armington lower-tier elasticities between 1972-1988 and 1990-2001 for American imports.[34] They find that the mean and median elasticities decreased for all types of goods at all levels of product disaggregation, i.e. at the 10-digit, 5-digit, and 3-digit levels. Lower perceived substitutability of products of different national origin would translate into a decreasing trade elasticity in the Armington framework, deepening the distance puzzle.

2.3 A method to measure the trade elasticity in the Armington framework

To tease out the aggregate trade elasticity in the Armington framework that can contribute to solving the aggregate distance puzzle, it is necessary to constrain the trade elasticity to be invariant across sectoral goods and destination markets. To the best of our knowledge, no paper has as yet provided evidence on the evolution of the Armington elasticity for aggregate bilateral trade in that context.[35]

We focus on the Armington framework for two main reasons. First, as we want to measure the trade elasticity from 1963 for all countries involved in world trade, we operate in a data-poor environment. We believe that the biases entailed by this lack of information do not make the exercise worthless because we are interested in the evolution of the parameter rather than in its exact value. We do not have sufficient information on disaggregated bilateral tariffs to carry out

[32] See Di Giovanni et al. (2011) who show that measured productivity dispersion among firms of which some are exporters is systematically higher than the underlying 'true' extent of technological heterogeneity among domestic firms. This wedge is more pronounced in sectors that are more open to trade.

[33] Foreign sales accounted for 30% of total sales in the 1980s and grew to 60% of total sales by 2010.

[34] See also discussion in Head and Mayer (2013).

[35] Head and Mayer (2013) put forward that the heterogeneity in the trade elasticity across trading pairs and changes in trade participation led to an increase in the aggregate trade elasticity. Our focus is on an alternative - and possibly complementary - mechanism, namely an increase in the perceived similarity of products of different origin.

the estimation in a model-independent way.[36] Nor do we have information on the distribution of domestic prices or disaggregated data on domestic production needed to estimate the trade elasticity in the Ricardian framework. The available trade data is amenable to the estimation of the Armington elasticity only in cross-section.

Second, as put forward in recent work, the disappearance of the demand parameter from the trade elasticity is a consequence of opting for an unbounded Pareto in modelling the firm productivity distribution in the Meltiz framework. Under less restrictive assumptions, the trade elasticity is a function of 2 or more parameters, and in particular combinations of demand and supply side parameters.[37] The coefficient is increasing in the magnitude of the Armington elasticity.

Measuring the Armington elasticity is a perennial issue in the trade literature. Feenstra et al. (2018) discuss the difficulties involved and implement a state-of-the art method to estimate both the upper-tier import demand elasticity (between the domestic good and the imported composite good) and the lower-tier Armington elasticity (between different foreign goods). [38]

We depart from this literature and suggest a cruder estimation method in part because we operate in a data-poor environment but also for more fundamental reasons. First, Feenstra's method and its developments rely on the assumption that the elasticity parameter remains constant through time, whereas our research question implies that the parameter can vary from year to year. Second, 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. While we admit that in most tractable theoretical settings these parameters would be the same, it is useful to bring this question to the data.

The UN COMTRADE database gives information on bilateral trade flows and cif unit values at the SITC 4-digit level since 1962 for the majority of countries. These data are sufficient to estimate the trade elasticity in the Armington framework. The perceived substitutability of imported sectoral goods can be estimated on the basis of the distribution of prices and of market shares in each market and each destination.

[36] See Head and Ries (2001) and Caliendo and Parro (2015) for the methodology. Caliendo et al. (2015) build a comprehensive dataset on bilateral tariffs that covers 1990-2010. A novel model-independent approach is proposed in Allen et al. (2019) that only requires data on income and gross output of each country. Although in principle feasible, this approach also requires constructing a set of instruments, the hypothetical values of income and gross output, and it delivers an estimate of the trade elasticity that is relatively imprecise.

[37] See Costinot and Rodríguez-Clare (2014), Bas et al. (2017), Feenstra et al. (2018), Feenstra (2018).

[38] The seminal paper using time variation in prices and quantities across a set of destination markets to pin down the lower-tier Armington elasticity for a specific product is Feenstra (1994). For subsequent refinements see Broda and Weinstein (2006), Imbs and Méjean (2015), and Soderbery (2015).

The basic intuition of the method we use starts from the demand equation for an exporter-specific sectoral good in the one-good Armington framework (assuming a CES utility function):

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

where X_{ij} is the cif value of the exports from i to j, P_{ij} is the cif price of the good shipped from i to j and P_j is the price index in the destination and Y_j total import demand in the destination. The exponent $(1-\sigma)$ captures the substitutability of country-composite goods across frameworks. It is also the aggregate trade elasticity ϵ in the Armington framework. Briging this equation to the data is difficult, however, as we do not observe aggregate prices, but unit values at the SITC 4-digit category level. Still, the distance puzzle concerns an elasticity estimated on aggregate trade data. As shown by Imbs and Méjean (2015) 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.

Define aggregate imports from source country i to destination country j as the sum of imports from each sector k where a sector corresponds to a SITC 4-digit category: $X_{ij} = \sum_k X_{k,ij}$. Given CES utility at the intersectoral level, sectoral demand in country in sector for imported goods is given by:

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

Where $P_{k,j}$ and P_j are price indexes, $\beta_k > 0$ is a sector-specific preference parameter, Y_j is total demand for imported goods, $\sigma > 1$ is the elasticity of substitution between sectors.

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 with the same σ parameter as the intersectoral CES utility function. Intrasectoral demand for varieties exported by i in j in sector k is:

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

Where $\gamma_i > 0$ is a origin-country-specific preference parameter and $P_{k,j}$ is the CES price index:

$$P_{k,j} \quad = \quad \left[\sum_{i \neq j} \left(\frac{p_{k,ij}}{\gamma_i}\right)^{1-\sigma}\right]^{1/(1-\sigma)}$$

Defining $\frac{Y_{k,j}}{Y_j} = \omega_{k,j}$, we get:

$$\frac{X_{k,ij}}{Y_j} \quad = \quad \omega_{k,j} \left(\frac{p_{k,ij}}{\gamma_i P_{k,j}}\right)^{1-\sigma}$$

Summing over all SITC 4-digit sectors:

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

Changing notation to $\kappa_i = \gamma_i^{\sigma-1}$, the market share equation for aggregate bilateral trade in a destination market is defined as a function of the weighted average of sectoral relative prices:

$$\frac{X_{ij}}{Y_j} \quad = \quad \kappa_i \sum_{k=1}^{K} \omega_{k,j} \frac{p_{k,ij}^{1-\sigma}}{\sum_{l\neq j} \kappa_l p_{k,lj}^{1-\sigma}} \tag{4}$$

As this is a market share, it is reasonable to assume that the errors are multiplicative. Measurement errors on a market share of 1% cannot be the same in percentage points as measurement errors on a market share of 10%. We have to fit the following model to the data:

$$\frac{X_{ij}}{Y_j} \quad = \quad \kappa_i \sum_{k=1}^{K} \omega_{k,j} \frac{p_{k,ij}^{1-\sigma}}{\sum_{l\neq j} \kappa_l p_{k,lj}^{1-\sigma}} . e^{\varepsilon_{i,j}} \tag{5}$$

Notice that we only have non-zero observations, as $p_{k,ij}$ is only observed when there is a trade flow. We take logs to transform the errors into additive ones and estimate the following equation for each year with a non-linear least square procedure in STATA:

$$\ln\left(\frac{X_{ij}}{Y_j}\right) \quad = \quad \ln \kappa_i + \ln\left(\sum_{k=1}^{K} \frac{Y_{k,j}}{Y_j} . \frac{p_{k,ij}^{1-\sigma}}{\sum_{l\neq j} \kappa_l p_{k,lj}^{1-\sigma}}\right) + \varepsilon_{i,j} \tag{6}$$

This approach yields annual estimates of κ_i and σ.[39]

3 Evolution of the Armington trade elasticity from 1963 to 2013

In this section we first discuss the results of the baseline estimation of equation (6). We then carry out a series of robustness checks to deal with missing unit values, zero trade flows, low data quality and endogeneity. We conclude with an interpretation of the distance puzzle through the lens of the Armington framework by decomposing the in the distance coefficient into the part attributable to the change in the trade elasticity and the part attributable to the change of the elasticity of trade costs to distance ρ.

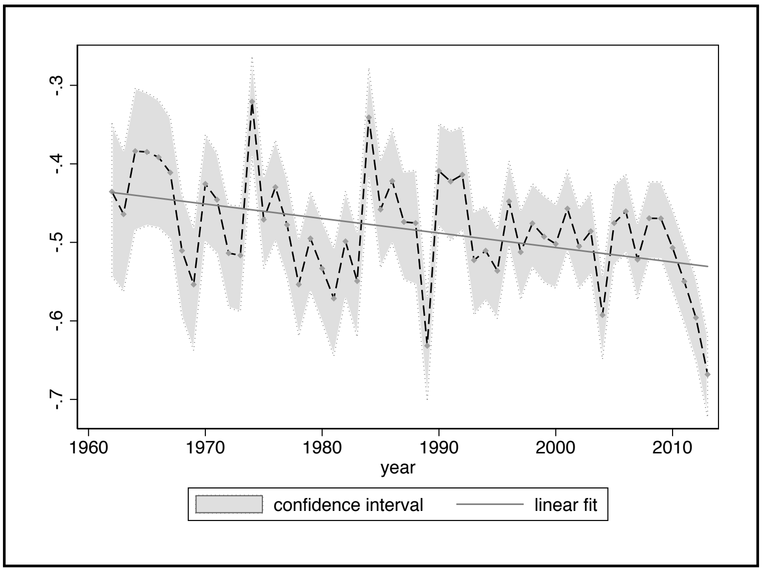
[39] Because of the way the procedure *nl* works in STATA, instead of minimizing $\sum_{i,j} \left(\varepsilon_{ij}\right)^2$, it minimizes $\sum_{i,j} n_{ij} \left(\varepsilon_{ij}\right)^2$ where n_{ij} is the number of 4-digit sectors in which imports from i to j are recorded. As a result, we have to weight each observation in the regression by $1/n_{ij}$ to make its results comparable to those in part one.

3.1 Results

The baseline estimation of equation (6) is carried out on annual cross-sections of the UN COMTRADE dataset. We work with the data at the 4-digit level, and we consider as separate goods trade flows with quantities measured in different units (kg,l,items). We therefore refer to the level of disaggregation as 4′-digit. We drop bilateral trade flows with missing information on quantity as well as trade flows and unit values with negative or zero values. We drop information on exporters whose market share is smaller than .1% of global trade. We drop bilateral trade flows whenever the associated unit value is in the top or bottom 5 percentiles of the unit value distribution in this product and destination market at the 4′-digit level. We also drop trade flows whenever the associated unit value is more than two orders of magnitude smaller or bigger than the median unit value in the product and destination market.

Figure 8 presents the results. The absolute value of the estimated trade elasticity $|1-\sigma|$ has increased by 22% from 1962 to 2013. This corresponds to an annual increase of .4% per year.[40]

Figure 8: Estimated $(1-\sigma)$, baseline



The magnitude of the estimated trade elasticity (in absolute value) is in the low range, i.e. $|1-\sigma| = \{.4, .5\}$. Working with US data, Feenstra et al. (2018) obtain a point estimate in the $\{0.5, 3\}$ range, depending on the estimator used, while Imbs and Méjean (2015) obtain a point

[40] The estimated annualized growth rate is significant at 1% level, the 95% confidence interval is between 0.15% and 0.63% (not taking into account though the uncertainty around yearly estimates).

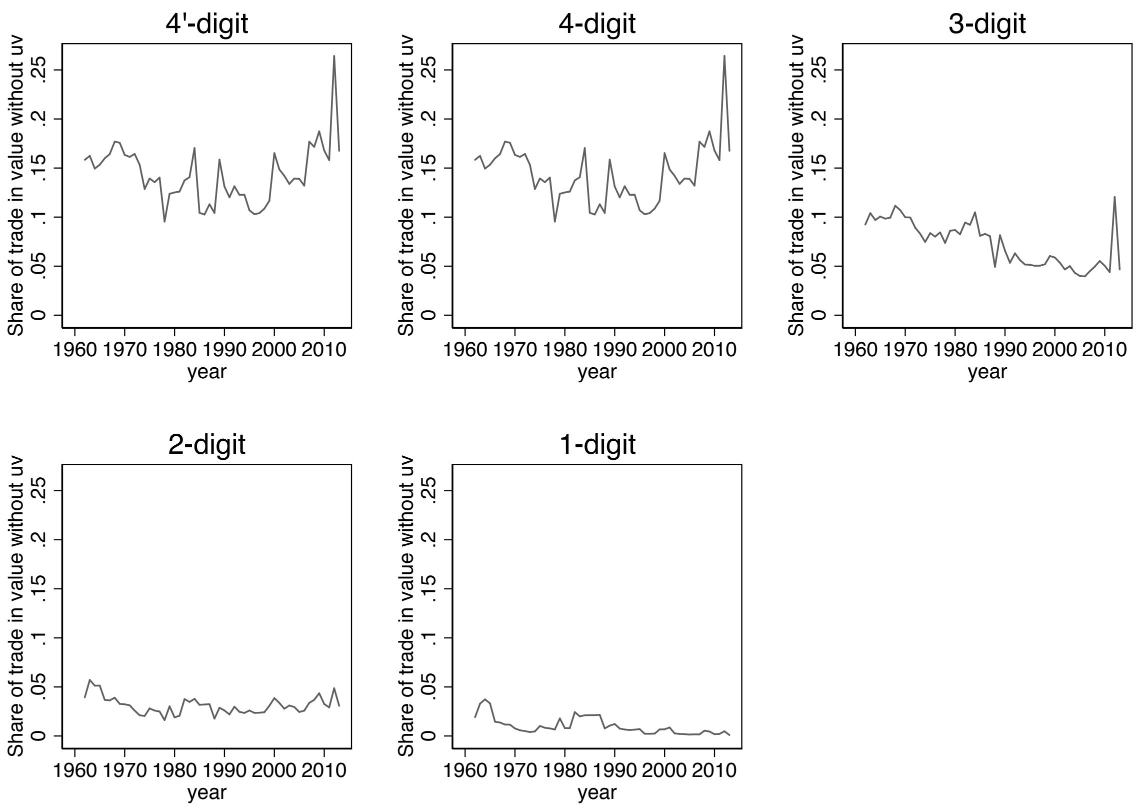
estimate of $1-\sigma=-2$.[41]

3.2 Robustness checks

3.2.1 The incidence of missing unit values

The COMTRADE dataset does not provide quantities for all non-zero trade flows. It is not possible to compute unit values in these cases. Figure 9, *4'-digit* indicates how often that happens from 1962 to 2013. The problem arises for a small share of the total trade value (never more than 25%). In the baseline estimation, we have simply dropped these observations. As a robustness check, it is possible to approximate missing unit values by unit values from similar products using a stepwise price imputation procedure.

Figure 9: Trade with missing unit values



The way to read this graph is : In 1962, 10% of trade is in a 3-digit bilateral trade category that does not mention at least one unit value

In this procedure, whenever possible, we first compute unit values at the highest disaggre-

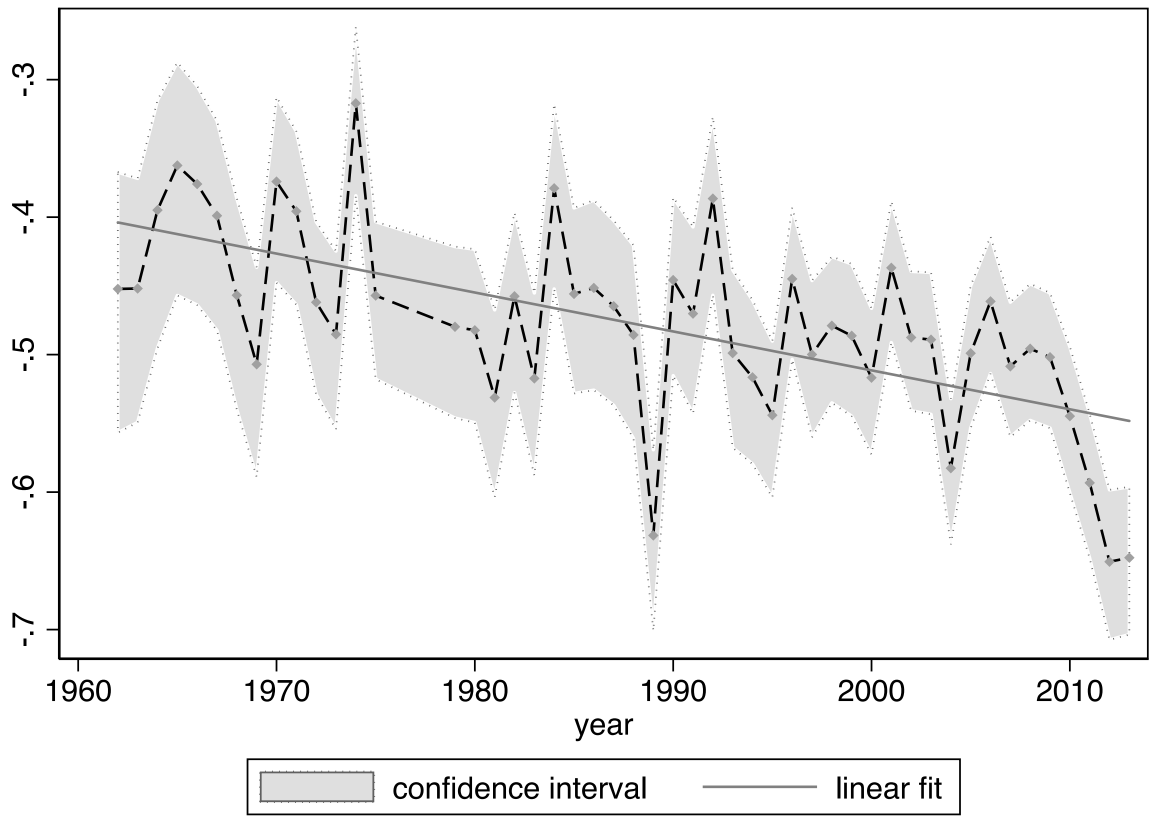
[41] The time variation in market shares and prices across US trade partners is used in these papers while we work off market share and price variation in a single cross-section.

gation level for each product and quantity unit provided by the source (the *4'-digit* level). When this is missing, we compute de trade-weighted mean relative unit value at the 4-digit level for this importer and use it to impute an absolute unit value at the *4'-digit* level.

If no relative unit value is available at the 4-digit level, we move at the 3-digit level, etc. As Figure 9 shows, that procedure allows to reduce considerably the share of trade without unit values. This procedure is justified if one assumes that missing destination-specific relative unit values at the 4'-digit can be approximated by the mean observed destination-specific relative price among the corresponding 4-digit group, or 3-digit...

Figure 10 shows that our result hold to the use of imputed prices. The yearly increase of the absolute value of the estimated $|1-\sigma|$ is 0.6% instead of 0.4%, for a total increase of 35% over the whole period.

Figure 10: Estimated $(1-\sigma)$, with imputed prices



3.2.2 Zero trade flows

A second difficulty arises when both quantity and value data are missing. Zero trade flows (ztf) are a prevalent feature of the data even though under model assumptions some trade should be observed in every sector k between all pairs ij. The share of observed SITC 4-digit flows relatively to the total number of potential SITC 4-digit bilateral trade flows increases from c. 3.5% to c. 9% between 1962 and 2013 (see Figure 11). The existence of zero trade flows can be reconciled with our theoretical framework by assuming that the underlying trade flows are strictly positive but so small that they do not pass the threshold applied by the data collecting authorities (in UN COMTRADE this threshold corresponds to 1000 USD). Still, the implied biais (if any) in our estimate is not clear.

To test the robustness of our results to this issue, we restrict the estimation to the "superbalanced" sample (see note 21). Figure 11 shows that the share of zero trade flows is smaller in the superblanced sample. The evolution of non-zero trade flows is similar in both samples. In both cases, it is multiplied by approximately 2.5. One would thus expect that any biais linked to the presence of zero trade flows would attenuated by the study of the superblalanced sample. Figure 12 shows that the increase in the absolute value of estimated $|1-\sigma|$ is faster: 1% a year instead of 0.4% a year, for a predicted total increase of 69% over the whole period. That suggests that, in itself, the large prevalence of zero trade flows does not lead to an over-estimation of the increase of $|1-\sigma|$.

Figure 11: Share of zero trade flows

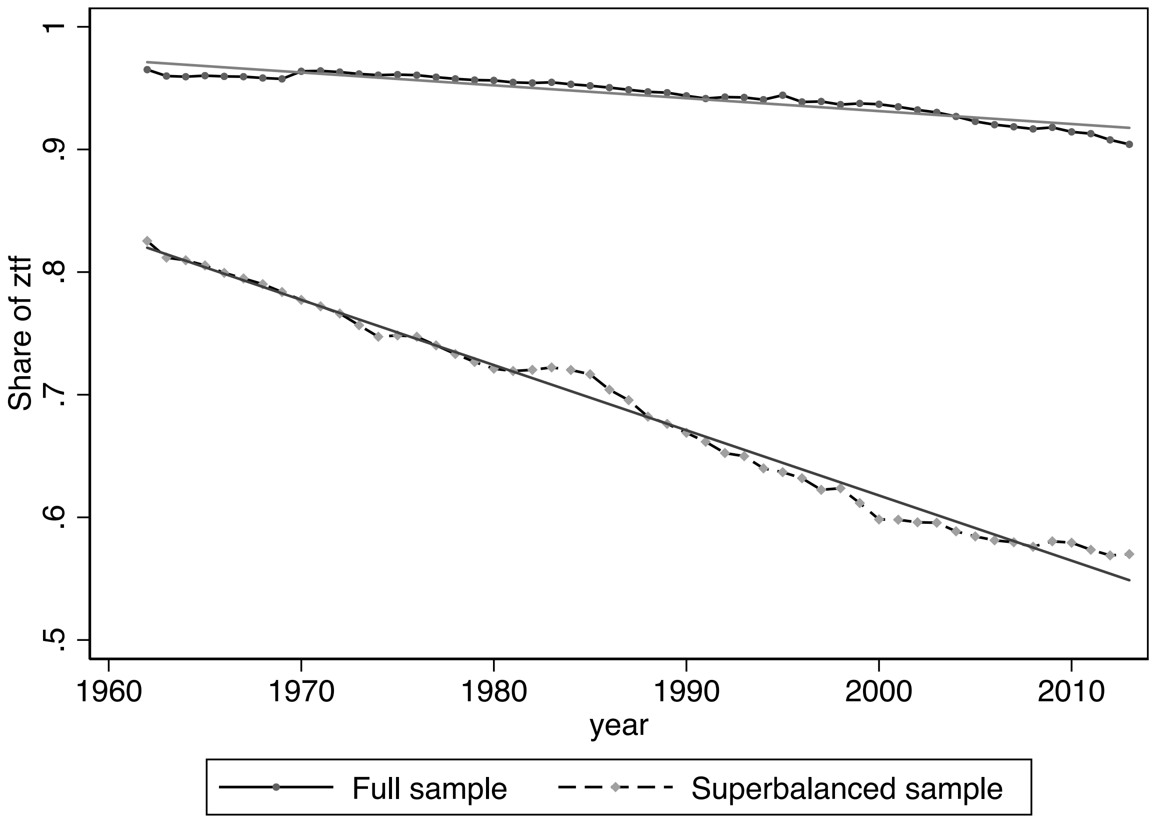
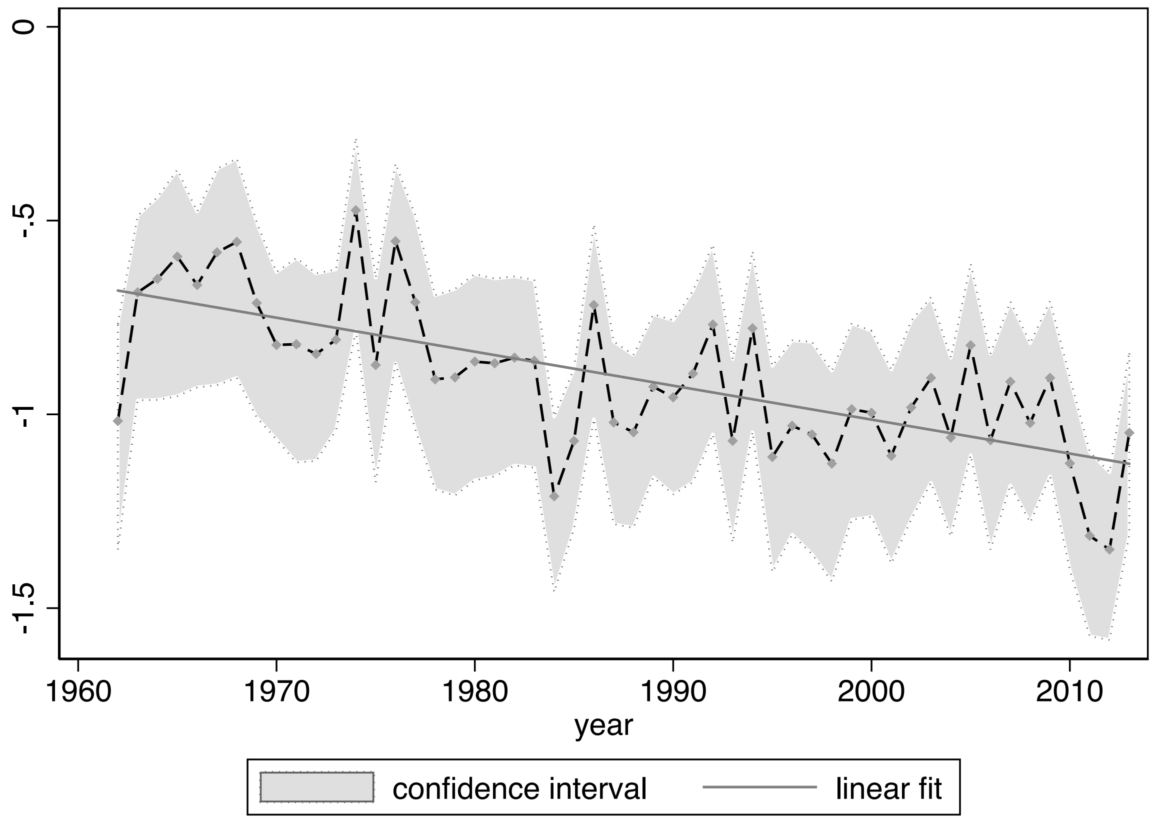


Figure 12: Estimated $(1-\sigma)$, superbalanced sample

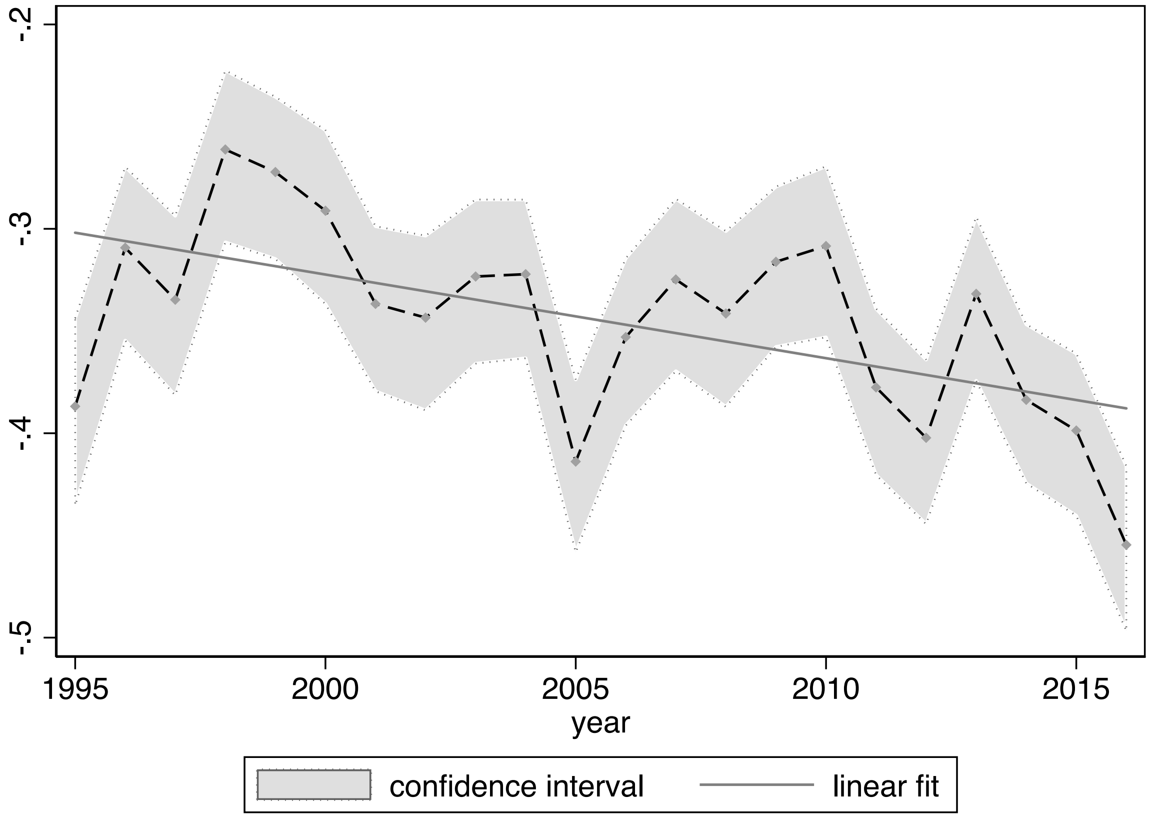


3.2.3 Changing the dataset

The UN COMTRADE dataset mostly simply brings together trade reported by individual countries. For example, it does not try to reconcile contradictory declarations for mirror flows. It is possible to increase the quality of the data by conducting an harmonization procedure. That has been don by the CEPII with the BACI dataset that reports bilateral trade data at the HS-1992 6-digit disaggregation level for 1995-2016 (Gaulier et Zignago 2010). As a result, BACI includes much better-quality unit values while substantially reducing the number of observations with lacking unit value. At the 6-digit level, less than 7% of total reported trade in BACI has missing unit values.

The disadvantage of BACI is that it covers only a relatively short period compared to the years over which the distance puzzle exists. Still, it is interesting to see if our estimated evolution of $|1-\sigma|$ is robust to the use of this better dataset. Obviously, we do not expect to reproduce exactly our baseline results because the trade classification and its level of aggregation are different Figure 13 shows that our results hold: the absolute value of $|1-\sigma|$ increases from 1995 to 2016. The speed of increase is faster than in the baseline estimate : 1.18% a year instead of 0.4%.

Figure 13: Estimated $(1-\sigma)$, BACI database



3.2.4 Estimation of $(1-\sigma)$ with predicted unit values

The estimate of σ that we report in section 3.1 may be downward-biased because of unobserved demand shifters that may be correlated with prices. Specifically, whenever the supply curve has a finite slope, unobserved demand shocks will result in a simultaneous increase in the price and in the expenditure on sector k products.[42] Moreover, even if the supply curve is perfectly elastic, measurement error may lead to a downward bias in the estimate of σ. This downward bias occurs whenever there is a systematic link between unobserved quality and observed unit values whereby higher observed unit values in sector k are associated with higher

[42] Soderbery (2018) provides estimates of exporter-specific supply elasticities. Broda et al. (2006) found that supply elasticities were finite at the 4-digit level.

underlying quality and, consequently, higher expenditure on sector k products.[43]

To check whether unobserved demand shifters correlated with prices are a source of concern in the estimation, in particular for evaluating the rate of change in the estimated parameter,[44] we implement a two-step procedure akin to an instrumental variable approach. That requires an instrument that adequately captures some exporter-specific shocks to sector k prices that are not demand-driven.

The shocks we focus on are exogenous shocks on the cost of material inputs. Producer price indices (PPI) would be our preferred data source on changes in the cost of domestic inputs. Unfortunately, information on producer price indices is not provided at the SITC 4-digit level. Even at the aggregate level, PPI information is not available for most countries and years in our sample. We therefore settle for the exporter price level of GDP and the exporter price level of capital formation (investment) as proxies of shocks to exporter-specific prices.

The price level of GDP is a valid instrument for exporter-specific domestic supply shocks if demand shocks in the importing country are the main source of price endogeneity, and the small country assumption holds.[45] In this case, demand shocks in any market for products of a particular exporter will have no incidence on world prices of these products and, consequently, on the price level of GDP in the exporting country. Thus the exclusion restriction would be satisfied as shocks would be independent of idiosyncratic demand shocks in the importing country. If however the small country assumption does not hold for certain importers, the price level of GDP could itself be a function of demand for exporter-specific products. Therefore, as a robustness check, we use the price level of investment as an alternative proxy for cost shocks to domestic production because the price of investment is more likely to be determined by global demand for industrial goods and to be exogenous to unobserved demand shocks in any given market.

Information on the price level of GDP and of investment is taken from the Penn World Table.[46] Each price level variable in the database is normalized relatively to the price level of US GDP in 2011. Changes in the domestic price level are constructed in such a way as to reflect

[43] Another potential problem is linked to using per kg instead of per item prices. As discussed in Hummels and Schaur (2013), unit values reflect product bulkiness, and bulkier products are more likely to be shipped via cheaper means of transportation. If exporter-specific sectoral goods differ in bulkiness, unit value differences will overstate price differences and lead to a downward bias in the estimated elasticity.

[44] Feenstra (1994) shows that the attenuation bias also affects the evolution of the estimated parameter.

[45] Magee and Magee (2008) provide evidence that the small country assumption likely holds in the data.

[46] We use the 9.0 release of the PWT described in Feenstra et al. (2015). This release contains data for 182 countries in 1950-2014.

the change in the full vector of international prices. [47] Information on the price level of GDP and of investment is provided in the database for 113 countries in every year. For the remaining countries between 10 and 51 years are covered. Thus, our sample of countries does not coincide perfectly with the countries in the Penn World Table. However, as it is the smallest exporters that drop out, the sample adjustment in terms of world trade coverage is minor.

The instrumenting procedure consists in exploiting past information on unit values together with information on changes in the cost of domestic inputs to replace observed unit values $p_{k,ij,t}$ with predicted unit values $\hat{p}_{k,ij,t}$. We assume a stylized cost function in which the landed cost of any sectoral good $c_{k,ij,t}$ is determined by an economy-wide cost measure C_{it} together with sector-producer specific characteristics summarized by the index $z_{k,i,t}$ (> 0) and bilateral characteristics of trade costs $\tau_{ij,t}$ (> 0):

$$c_{k,ij,t} \quad = \quad C_{i,t} z_{k,i,t} \tau_{ij,t}$$

We assume that the common cost component is well captured by the GDP or investment price level: $C_{i,t} = P^{\nu}_{i,t}$ where $\nu = \{gdp, i\}$. We denote by $\alpha_t > 0$ the sensitivity of prices to costs.[48] Prices are also dependent on consumer-sector-producer specific characteristics $d_{k,ij,t} > 0$ that capture, for example, shocks to demand in j for products from i in sector k. The landed price of each sectoral good is given by:

$$p_{k,ij,t} \quad = \quad \left(P^{\nu}_{i,t} z_{k,i,t} \tau_{ij,t}\right)^{\alpha_t} d_{k,ij,t}$$

Denoting the time lag by l, we express the unit value in t as a function of the unit value in $t-l$ and of the supply and demand shocks intervening between $t-l$ and t:

$$p_{k,ij,t} \quad = \quad p_{k,ij,t-l} \left(\frac{P^{\nu}_{i,t}}{P^{\nu}_{i,t-l}} \frac{z_{k,i,t}}{z_{k,i,t-l}} \frac{\tau_{ij,t}}{\tau_{ij,t-l}} \right)^{\alpha_{t,l}} \frac{d_{k,ij,t}}{d_{k,ij,t}} \tag{7}$$

The log difference of the unit value for each time lag l is given by:

$$\ln p_{k,ij,t} - \ln p_{k,ij,t-l} \quad = \quad \alpha_{t,l} \ln\left(\frac{P^{\nu}_{i,t}}{P^{\nu}_{i,t-l}}\right) + \alpha_{t,l} \ln\left(\frac{z_{k,i,t}}{z_{k,i,t-l}}\right) + \alpha_{t,l} \ln\left(\frac{\tau_{ij,t}}{\tau_{ij,t-l}}\right) + \ln\left(\frac{d_{k,ij,t}}{d_{k,ij,t}}\right) \tag{8}$$

We estimate the elasticity $\alpha^{\nu}_{t,l}$ for each instrument $\nu = \{gdp, i\}$ and time lag $l = \{1, 2, 3\}$, assuming that the sum of the three other terms on the right hand side of (8) follows a normal

47 Real GDP is constructed in chained PPPs in the PWT 9.0, i.e. keeping prices constant across countries and also over time. This approach entails that price levels are comparable within but also across countries over time.

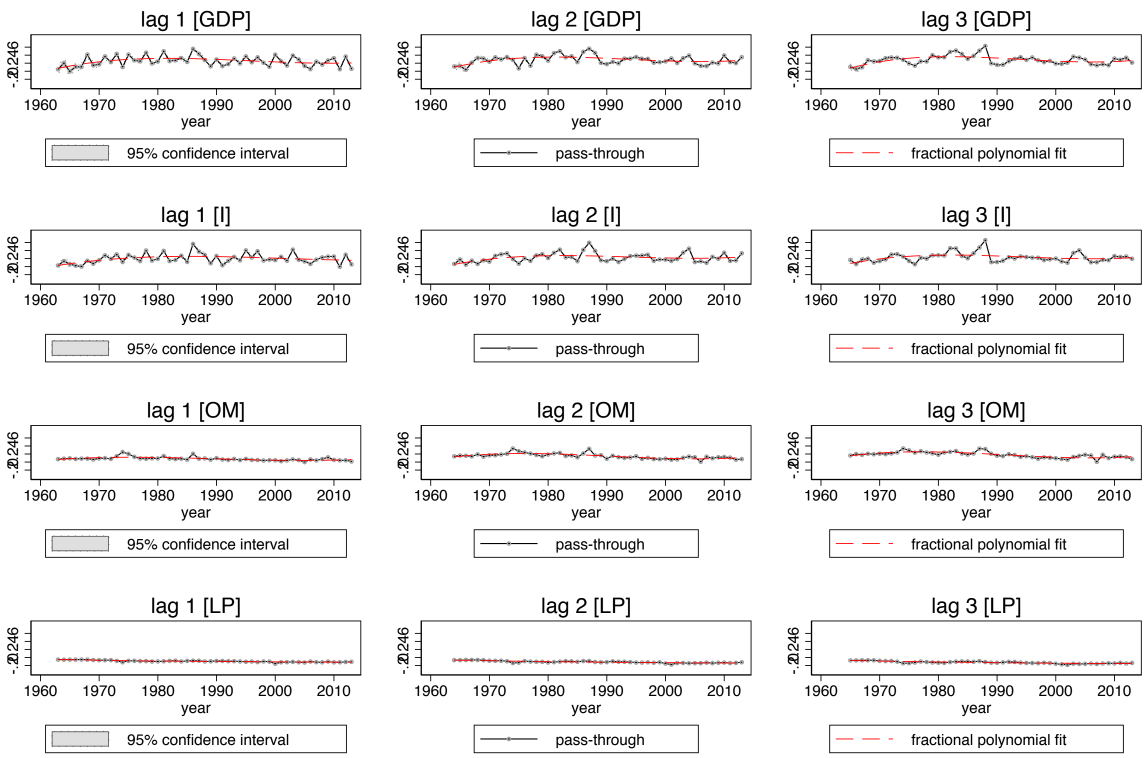
48 This measure of pass-through may depend on the competition conditions in the sector specific to the origin country. But for each individual exporter-sector combinations, our observed set of destinations is relatively low. Therefore we make the simplifying assumption of common pass-through.

law. The identifying assumption is that pair-sector specific shocks $\varepsilon_{k,ij,t-l}$ are independent of supply shocks that increase the cost of production. Our baseline equation for the first stage of the estimation for a given lag is:

$$\ln p_{k,ij,t} - \ln p_{k,ij,t-l} \quad = \quad \alpha^{\nu}_{t,l} \ln\left(\frac{P^{\nu}_{i,t}}{P^{\nu}_{i,t-l}}\right) + \varepsilon_{k,ij,t,l} \tag{9}$$

Figure 14 presents the results for the estimated pass-through. For almost all years and all specifications - with a few exceptions in the early years - the estimated pass-through is significant and positive. Further, the estimated pass-through does not strongly depend on either the instrument or the time lag.

Figure 14: Estimated $\alpha^{\nu}_{t,l}$



Note: [GDP] stands for GDP price level, [I] stands for investment price level,
[OM] for the price evolution in other markets, [LP] for lagged price

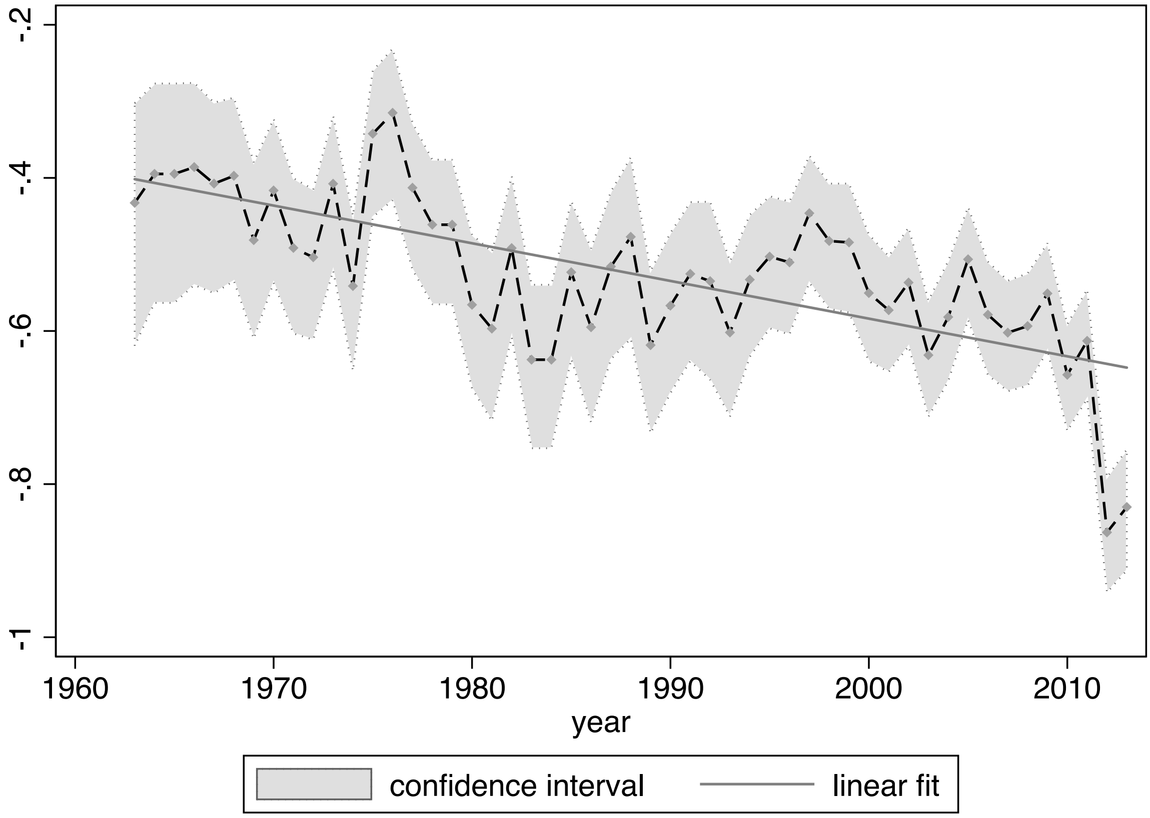
The predicted unit values in t, $\hat{p}_{k,ij,t}$ are constructed by augmenting the unit values observed in $t-l$ with the change in the common cost component, given the estimate of the pass-through $\hat{\alpha}^{\nu}_{t-l}$:

$$\hat{p}_{k,ij,t} \quad = \quad p_{k,ij,t-l}\left(\frac{P^{\nu}_{i,t}}{P^{\nu}_{i,t-l}}\right)^{\hat{\alpha^{\nu}_{t,l}}} \tag{10}$$

As the past unit values are a very strong predictor of current unit values, we are not worried about the strength of our instrument, notwithstanding the fact that economy-wide cost shocks explain a relatively small fraction of variation in landed sectoral prices.

We estimate (6) while replacing current unit values with the predicted unit values $\hat{p}_{k,ij,t-l}$ from (10). To keep as many years as possible, we opt for the smallest time lag ($l = 1$). Figure 15 presents the results. [49]

Figure 15: Estimated $1-\sigma$, predicted unit values



The absolute value of the estimated trade elasticity $|1-\sigma|$ has increased by 60% from 1963 to 2013. The predicted trade elasticity equals $-.65$ in 2013. This corresponds to an annual increase of .9% per year.[50] The evolution of the parameter is much more pronounced when predicted unit values are used in the estimation, suggesting that price endogeneity may indeed be affecting not only the level but also the evolution of the estimated elasticity.

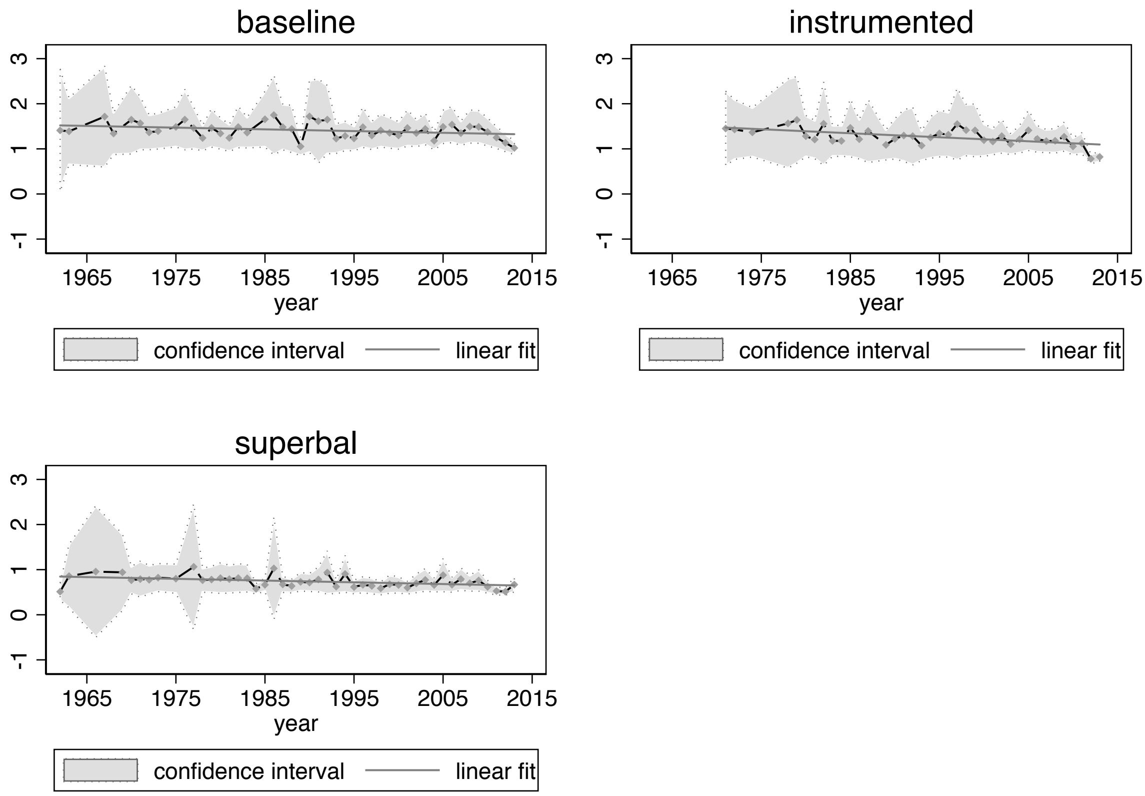
[49] Reported confidence intervals do not take into account the fact that we are using predicted unit values $\hat{p}_{k,ij,t-l}$ on the right-hand side of the equation: we do not know how to make the adjustment in that setting

[50] The estimated annualized growth rate is significant at the 1% level, the 95% confidence interval is between 0.7% and 1.2% (not taking into account though the uncertainty around yearly estimates).

3.3 Is there a distance puzzle left?

This section has provided empirical evidence on the evolution of the aggregate substitutability parameter for world trade from 1963 to 2013. In the Armington framework, this substitutability parameter $(1-\sigma)$ corresponds to the aggregate trade elasticity ϵ. We find that this parameter has increased by 22% between 1963 and 2009 in the baseline estimation, and by 60% when prices are instrumented. Section 1 has shown that the distance elasticity of trade (δ)has increased by 4% over the same period in the baseline estimate. A naïve combination of these two results suggests, there is no distance puzzle left in the framework of the Armington model in as much as the elasticity of trade costs to distance (ρ) has decreased by 15% in 1963-2013. More formally, we run Monte Carlo estimation of the mean and standard deviation of ρ defined as $\frac{\delta}{1-\sigma}$. Figure 16 gives the results for the baseline estimation, the instrumented estimation and the superbalanced sample. Years where the estimated standard deviation is too large or the estimation of σ did not converge are dropped out. The resulting precision of the measure of ρ is not what we would like. That is partially because of the use of the Huber-White method to correct standard errors in the gravity equation. Still, in all cases, the point estimates suggest a decrease of ρ by 13% in the baseline estimate, 27% using the instrumented method and 23% in the superbalanced sample. Increasing perceived substitutability of country-specific composite goods contributes to the non-decreasing distance elasticity of trade.

Figure 16: Estimated ρ



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.[51]

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 composition of world trade, leading to a change in the substitutability parameter measured on aggregate data. However, at first approximation, the rising importance of manufactures compared to primary products in world trade should have reduced substituability.

[51] (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.

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 et Méjean 2013), 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 et Khandelwal 2013) as well as with increased vertical specialization of countries within sectors documented by (Fontagné, Gaulier, et Zignago 2008).

4 Conclusion

The estimated effect of distance in gravity equations has not decreased in the past fifty years despite substantial innovation in transportation and communication that should have reduced the relative effect of distance: this is the 'distance puzzle'. Using COMTRADE 4-digit bilateral trade data in 1962-2013, 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 4% from 1962 to 2013. This result holds when we correct for changes in the sample of trading partners and the composition of world trade.

Second, the paper suggests a method to measure structural heterogeneity in the Armington framework. 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 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, which can be approximated by studying the relations between the prices and the market share of importers in destination markets.

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. Depending on the method of estimation, the estimated elasticity increases by between 22% and 60% between 1963 and 2013.

Combining the estimation of the distance coefficient and the elasticity of trade to trade costs, we can then compute the elasticity of trade costs to distance. Alas, that estimate is bound to be very measured with uncertainty, because it the the ratio of two uncertain results. Still, the point estimates are our best guess at the estimated values and are compatible with a decrease of the elasticity of trade costs to distance of between 13% and 23% between 1963 and 2013.

References

Allen, J., C. Arkolakis, and Takahashi, Y. (2019). Universal Gravity. *Journal of Political Economy (forthcoming).*

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

Andrews, D., C. Criscuolo, and P. N. Gal (2016, December). The Best versus the Rest: The Global Productivity Slowdown, Divergence across Firms and the Role of Public Policy. OECD Productivity Working Papers 5.

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

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2017, May). The Fall of the Labor Share and the Rise of Superstar Firms. Technical Report w23396, National Bureau of Economic Research, Cambridge, MA.

Baldwin, R. and J. Harrigan (2011). Zeros, quality and space: Trade theory and trade evidence. *American Economic Journal: Microeconomics 3*(2), 60–88.

Bas, M., T. Mayer, and M. Thoenig (2017). From micro to macro: Demand, supply, and heterogeneity in the trade elasticity. *Journal of International Economics 108*, 1–19.

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 (2014). Applying the GLM Variance Assumption to Overcome the Scale Dependence of the Negative Binomial QGPML Estimator. *Econometric Reviews 33*(7).

Bosquet, C. and H. Boulhol (2015). What is really puzzling about the “distance puzzle”? *Review of World Economics 151*(1), 1–21.

Broda, C., J. Greenfield, and D. Weinstein (2006). From Groundnuts to Globalization: A Structural Estimate of Trade and Growth. *NBER Working Paper Series 12512*.

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

Brun, J.-F., C. Carrère, P. Guillaumont, and J. de 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.

Cairncross, F. (1997). *The Death of Distance: How the Communications Revolution Will Change Our Lives*. London: Orion Business Books.

Caliendo, L., R. C. Feenstra, J. Romalis, and A. M. Taylor (2015). Tariff Reductions, Entry, and Welfare: Theory and Evidence for the Last Two Decades. *National Bureau of Economic Research Working Paper Series No. 21768*.

Caliendo, L. and F. Parro (2015, January). Estimates of the Trade and Welfare Effects of NAFTA. *The Review of Economic Studies 82*(1), 1–44.

Chaney, T. (2008). Distorted gravity: The Intensive and Extensive Margins of International Trade. *American Economic Review 98*(4), 1707–1721.

Chaney, T. (2018). The Gravity Equation in International Trade: An Explanation. *Journal of Political Economy 126*(1), 150–177.

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. *The Review of Economic Studies 79*(2), 581–608.

Costinot, A. and A. Rodríguez-Clare (2014). Trade Theory with Numbers: Quantifying the Consequences of Globalization. In *Handbook of International Economics*, Volume 4, pp. 197–261. Elsevier.

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.

Daudin, G., C. Rifflart, and D. Schweisguth (2011). Who Produces for Whom in the World Economy. *Canadian Journal of Economics 44*(4), 1403–1437.

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

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

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.

Erkel-Rousse, H. and D. Mirza (2002). Import price elasticities: Reconsidering the evidence. *The Canadian Journal of Economics 35*(2).

Fally, T. (2015). Structural Gravity and Fixed Effects. *Journal of International Economics 97*(1).

Feenstra, R. (1994). New Product Varieties and the Measurement of International Prices. *The American Economic Review 84*(1), 157–177.

Feenstra, R. C. (2018, January). Restoring the product variety and pro-competitive gains from trade with heterogeneous firms and bounded productivity. *Journal of International Economics 110*, 16–27.

Feenstra, R. C., R. Inklaar, and M. Timmer (2015). The next generation of the Penn World Table. *American Economic Review 105*(10), 3150–3182.

Feenstra, R. C., P. Luck, M. Obstfeld, and K. N. Russ (2018, March). In Search of the Armington Elasticity. *The Review of Economics and Statistics 100*(1), 135–150.

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

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

Gabaix, X., J.-M. Lasry, P.-L. Lions, and B. Moll (2016). The Dynamics of Inequality. *Econometrica 84*(6), 2071–2111.

Head, K. and T. Mayer (2013). What separates us? Sources of resistance to globalization. *Canadian Journal of Economics 46*(4), 1196–1231.

Head, K. and T. Mayer (2014). Handbook of international economics. Volume 4. Elsevier.

Head, K. and J. Ries (2001). Increasing Returns versus National Product Differentiation as an Explanation for the Pattern of US-Canada Trade. *American Economic Review 91*(4), 858–876.

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

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 (2013). Time as a Trade Barrier. *American Economic Review 103*, 1–27.

Imbs, J. and I. Méjean (2015). Elasticity Optimism. *American Economic Journal: Macroeconomics 7*(3), 43–83.

Krautheim, S. (2012). Heterogeneous firms, exporter networks, and the effect of distance on international trade. *Journal of International Economics 87*, 27–35.

Larch, M., P.-J. Norbäck, S. Sirries, and D. M. Urban (2016). Heterogenous Firms, Globalization, and the Distance Puzzle. *The World Economy 39*(9), 1307–1338.

Levchenko, A. A. and J. Zhang (2016). The Evolution of Comparative Advantage: Measurement and Welfare Implications. *Journal of Monetary Economics 78*, 96–111.

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*.

Mayer, T., V. Vicard, and S. Zignago (2019). The Cost of Non-Europe, Revisited. *Economic Policy (forthcoming)*.

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

Melitz, M. (2003). The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity. *Econometrica 71*(6), 1695–1725.

Noguera, G. (2012). Trade Costs and Gravity for Gross and Value Added Trade. *unpublished manuscript*.

Reinert, K. A. and C. R. Shiells (1991). 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.

Santos Silva, J. M. C. and S. Tenreyro (2011). Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator. *Economics Letters 112*, 220–222.

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

Simonovska, I. and M. Waugh (2014). The Elasticity of Trade: Estimates and Evidence. *Journal of International Economics 92*(1).

Soderbery, A. (2015). Estimating Import Supply and Demand Elasticities: Analysis and Implications. *Journal of International Economics 96*(1), 1–17.

Soderbery, A. (2018). Trade elasticities, Heterogeneity, and Optimal Tariffs. *Journal of International Economics 114*, 44–62.

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.

Yi, K.-M. (2010). Can Multistage Production Explain the Home-Bias in Trade? *American Economic Review 100*(1), 364–393.

A Full and superbalanced samples

The full sample contains 205 reporters (R) and 227 partners (P) listed in tables 2 and 3 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 17 shows the distribution of pairs in the full sample according to the number of years in which the pair reports a positive amount of trade.

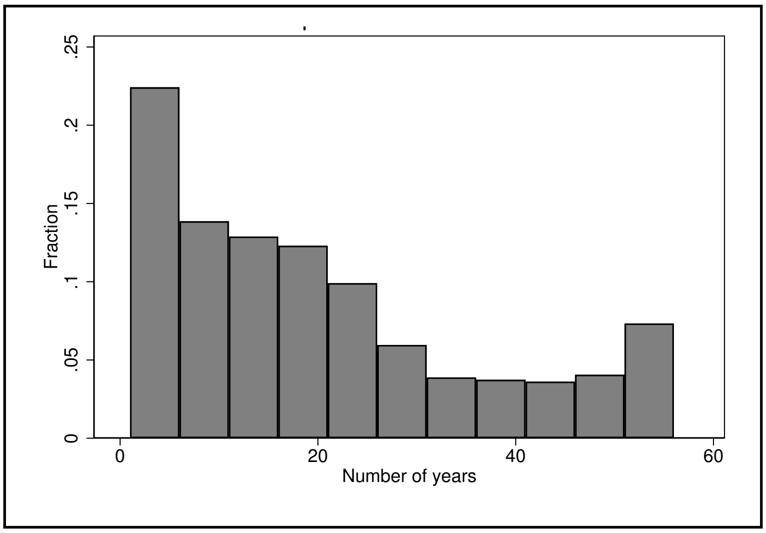


Figure 17: Number of years each pair is present in the sample

The superbalanced sample corresponds to the subsample of pairs which trade both ways in each and every year in 1962-2013. To avoid discarding pairs which fall out of the superbalanced sample because countries split up or reunite at some point in 1962-2013, we introduce several additional single country identifiers before constructing the superbalanced sample. Consequently, Germany is present in the superbalanced sample.[52]

The superbalanced sample comprises 786 trading pairs and corresponds to 32 countries. This is less than the 992 pairs which would be observed if each reporter traded both ways with every other country. Indeed, the set of countries which trade with every other country in each

[52] A single identifier is used for East, West, and reunited Germany. A single identifier is used for the Czech Republic, Slovakia, and Czechoslovakia. A single identifier is used for the USSR and the 15 countries formed after the USSR split up. The 15 countries which constituted the USSR are absent from the superbalanced sample because the USSR is never a reporter to COMTRADE. The Czech Republic and Slovakia also drop out because there is no other country in the sample with which they have two-way trade in every year.

year and which we refer to as the 'square sample' comprises just 21 countries (420 pairs). Trade coverage in the superbalanced sample decreases from 68 to 39% of total trade over 1962-2013 while it is reduced from 54 to 28% for the square sample (see Figure 18).

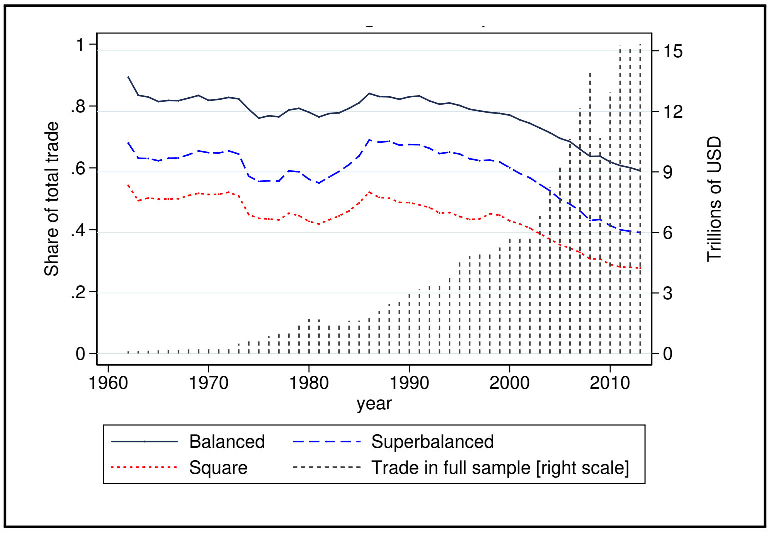


Figure 18: Trade coverage in 1962-2013

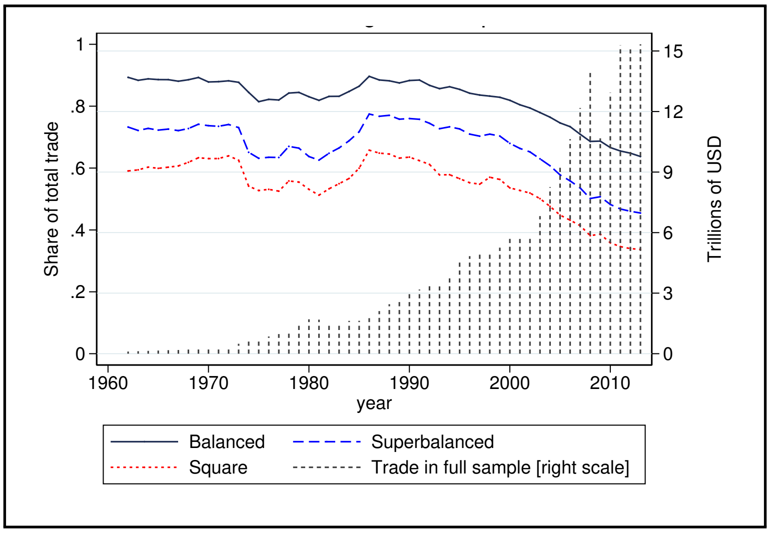


Figure 19: Trade coverage in 1965-2013

To check the sensitivity of trade coverage to the choice of the benchmark year we redefine the set of stable pairs as partners trading both ways in each year in 1965-2013.[53] The superbalanced sample for 1965-2013 contains 1286 pairs that comprise 42 countries (out of 1722 possible pairs). The square sample for 1965-2013 contains 23 countries (506 pairs).

[53] The number of reporters increases from 70 in 1962 to 86 in 1965 to 111 in 1970 and 129 in 1975. But the reduction in trade coverage between the 1970s and the 1990s is qualitatively similar if we choose 1970 or 1975.

As illustrated in Figure 19, the evolution of trade coverage is not sensitive to the choice of the starting year. Albeit from a higher level (73% in 1965), trade coverage drops to 45% by 2013. Indeed, the main reason for the reduction in trade coverage since the mid-1990s is due to the absence of China and of Central and Eastern European countries from the superbalanced sample. These countries drop out because they do not report trade to COMTRADE until the more recent period.[54]

Another way to check the sensitivity of trade coverage to the choice of the starting year is to compute the evolution of trade coverage for the sample of pairs which trade both ways in a specific year. The difference with the superbalanced sample is that we now relax the constraint that the pair have two-way trade in every year in 1962-2013. The share of total annual trade attributable to two-way pairs in 1962 and in 1965 is shown in Figure 20. As previously, the level but not the evolution of trade coverage is affected by the choice of the starting year.

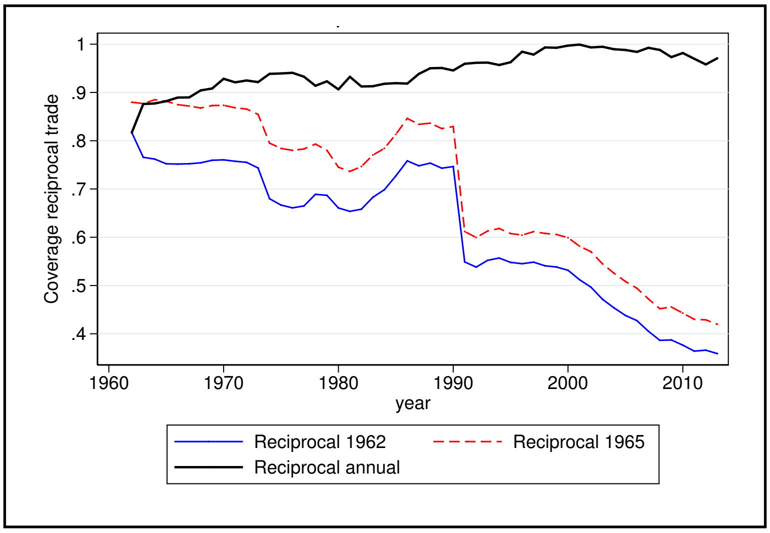


Figure 20: Trade coverage in reciprocal trade

Helpman et al. (2008) document 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 because most of the increase was driven by pairs which traded both ways in 1970. We nuance this finding by showing that since the mid-1990s new reciprocal trade relationships did contribute strongly to the growth of world trade. Figure 20 illustrates that while more than 70% (resp. 80%) of world trade in 1990 is attributable to pairs that traded both ways in 1962 (resp. 1965), less than 40% (resp. 50%) is still attributable to such pairs in 2013.

[54] For example, China reports trade to UN COMTRADE in 1987-2013.

We document that these new trade relationships were formed between countries trading both ways. To illustrate, Figure 20 also shows the share of total trade attributable to pairs which trade both ways in a given year (in black). Since the 1990s more than 95% of total annual trade takes place between pairs which trade both ways. One-way trade flows represent a marginal and decreasing share of world trade. The greater frequency of one-way trade relationships observed in the early years of the sample may be in part attributable to the fact that the number of countries reporting trade to COMTRADE was initially relatively small. Consequently, reported zeros may be at least in part statistical zeros, i.e. non-zero trade flows reported as zeros due to missing reports to COMTRADE.

Table 2: 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 (Fm Panama Cz)	*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-Luxembourg	**R;P;S**	**Hong Kong**	**R;P;S**	St. Helena	P
Belize	*R;P*	Hungary	*R;P*	St. Kitts and Nevis	*R;P*
Benin	*R;P*	**Iceland**	**R;P;S**	St. Lucia	*R;P*
Bermuda	*R;P*	India	*R;P*	St. Vincent-Grenadines	*R;P*
Bhutan	*R;P*	Indonesia	*R;P*	Samoa	*R;P*
Bolivia	*R;P*	Iran	*R;P*	San Marino	P
Bosnia-Herzeg.	*R;P*	Iraq	*R;P*	Sao Tome-Principe	*R;P*
Botswana	*R;P*	Ireland	R;P	Saudi Arabia	*R;P*
Brazil	**R;P;S**	**Israel**	**R;P;S**	Senegal	*R;P*
Br. Virgin Islands	P	**Italy**	**R;P;S**	Serbia-Montenegro	*R;P*
Brunei Darussalam	*R;P*	Jamaica	*R;P*	Seychelles	*R;P*
Bulgaria	*R;P*	**Japan**	**R;P;S**	Sierra Leone	*R;P*
Burkina Faso	*R;P*	Jordan	*R;P*	**Singapore**	**R;P;S**
Burma (Myanmar)	*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*

Table 3: List of countries in the full and superbalanced samples: Contd.

Country name	**Status**	**Country name**	**Status**	**Country name**	**Status**
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	Syria	*R;P*
Congo	*R;P*	Macau (Aomen)	*R;P*		
Dem. Rep. of Congo	*R;P*	Macedonia	*R;P*	Tajikistan	*R;P*
Cook Islands	*R;P*	Madagascar	*R;P*	Tanzania	*R;P*
Costa Rica	*R;P*	Malawi	*R;P*	**Thailand**	**R;P;S**
Croatia	*R;P*	**Malaysia**	**R;P;S**	Togo	*R;P*
Cuba	*R;P*	Maldives	*R;P*	Tokelau	P
Cyprus	*R;P*	Mali	*R;P*	Tonga	*R;P*
Czech Republic	*R;P*	Malta	*R;P*	Trinidad-Tobago	*R;P*
Czechoslovakia	*R;P*	Marshall Islands	P	**Tunisia**	**R;P;S**
Côte d'Ivoire	*R;P*	Martinique	*R;P*	**Turkey**	**R;P;S**
Denmark	**R;P;S**	Mauritania	*R;P*	Turkmenistan	*R;P*
Djibouti	*R;P*	Mauritius	*R;P*	Turks-Caicos Islands	*R;P*
Dominica	*R;P*	**Mexico**	**R;P;S**	Tuvalu	*R;P*
Dominican Republic	*R;P*	Micronesia	P	Uganda	*R;P*
East Germany (DDR)	R;P	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 (Fm Vietnam Rp)	*R;P*
Falkland Islands	P	**Netherlands**	**R;P;S**	Wallis-Futuna	*R;P*
Palestine	*R;P*	New Caledonia	*R;P*	West Germany (FRG)	R;P
Fiji	*R;P*	New Zealand	*R;P*	Western Sahara	P
Finland	R;P	Nicaragua	*R;P*	Yemen	*R;P*
France	**R;P;S**	Niger	*R;P*	Yugoslavia (Serbia-Mont.)	*R;P*
French Guiana	*R;P*	Nigeria	*R;P*	Zambia	*R;P*
Niue	P	Norfolk Island	P	Zimbabwe	*R;P*