The Changing Magnitude of the Gains from Trade*

Preliminary: Do Not Cite

Elizaveta Archanskaia†

August 2014

Abstract

This paper nests producer cost heterogeneity within a two-tier CES preference structure and finds that this approach helps define a consistent estimation strategy to identify the Armington elasticity in cross-section. This parameter is found to have increased by 25% in 1963-2009. In this model, the Armington elasticity contributes to determining the trade elasticity in combination with the lower-tier elasticity and the extent of producer cost heterogeneity. We provide direct estimates of the trade elasticity together with estimates of lower-tier substitutability and deduce the contribution of each dimension of heterogeneity to movements in the trade elasticity. The trade elasticity is found to have been mainly determined by the Armington elasticity. Estimates of the trade elasticity are combined with information on the share of domestic expenditure to show that the magnitude of static gains from trade has remained largely unchanged, notwithstanding increased reliance on foreign partners between 1963 and 2009.

Keywords: trade elasticity, cost heterogeneity, product substitutability, Armington

JEL codes: F11, F15

*This paper has greatly benefited from helpful discussions with Thomas Chaney, Lionel Fontagné, Jacques Le Cacheux, Thierry Mayer, André Romahn, Johannes van Biesebroeck, and Frank Verboven.

†KU Leuven.

1 Introduction

Structural changes in the organization of global production networks have put into question the ability of the Armington model (Armington (1969)) to deliver accurate results in counterfactual analysis because it fails to take into account the incidence of shocks on global value chains (Johnson and Noguera (2012), Bems et al. (2011, 2010)). Although in refined CGE modelling the sensitivity of trade volumes to cost shocks remains intrinsically determined by the extent of product substitutability, shocks' dampening or amplification through the supply chain comes to play a crucial role. The degree of input substitutability becomes the key parameter for policy analysis while the benchmark Armington model singled out perceived substitutability of final goods as the main object of interest (Bems (2014)).

More disturbingly, the Armington approach may be uninformative on the sensitivity of trade flows to cost shocks in a world with producer cost heterogeneity unconditionally on allowing for or precluding trade in inputs.[1] In particular, the elasticity of substitution between goods of different origin plays no role in determining the elasticity of trade flows to trade costs in Eaton and Kortum (2002) and Chaney (2008) while the trade elasticity remains the single most important parameter for determining the magnitude of welfare gains from trade (Arkolakis et al. (2012)).[2]

Nonetheless, the debate on the magnitude of the Armington elasticity of substitution among goods of different origin remains vivid, as evidenced by the string of recent papers on this subject (Feenstra et al. (2014), Feenstra and Romalis (2014), Imbs and Méjean (2014), Soderbery (2010, 2012)). Feenstra et al. (2014) attribute this endurance to two features. Firstly, there remains 'substantial uncertainty' about the type of substitutability which matters for policy analysis. In particular, the finding that the Armington elasticity has no incidence on the magnitude of gains from trade is itself specific to models with producer heterogeneity in which the degree of substitutability between goods of different foreign origin and goods of domestic and foreign origin coincide (Costinot and Rodriguez-Clare (2014)). Secondly, there remains 'substantial uncertainty' about the true underlying value of the Armington elasticity. Unobserved product quality, finite price elasticities of supply, measurement error in international prices, and estimates' sensitivity to aggregation are prominent methodological concerns.

[1] Influential papers on producer heterogeneity are Bernard et al. (2003) and Eaton et al. (2011). See Melitz and Redding (2014) and Bernard et al. (2012) for an overview of empirical evidence and model-based analysis.

[2] The trade elasticity may not suffice to compute total gains from trade in models with producer heterogeneity because additional welfare gains are generated by endogenous entry and exit. See Melitz and Redding (2013).

The state-of-the-art approach to obtain consistent estimates of the Armington elasticity at the product level exploits information on average variation of relative prices and relative market shares across the set of suppliers over a 10- to 20-year window (Feenstra (1994)). This estimation strategy is refined in Feenstra et al. (2014) to obtain two parameter estimates for each product on the US market: the elasticity of substitution between varieties of different foreign origin (the 'micro-elasticity') and the elasticity of substitution between the bundle of domestic varieties and the bundle of foreign varieties (the 'macro-elasticity').

A theoretically grounded weighted average of product-specific elasticities estimated in Feenstra et al. (2014) delivers two Armington elasticities for assessing substitutability in aggregate trade. The micro-elasticity captures substitutability between product bundles of different foreign origin while the macro-elasticity captures substitutability between the domestic bundle and the foreign bundle.[3] Whenever these two elasticities differ, they both contribute to determining the elasticity of trade flows to trade costs. Whenever they coincide, the trade elasticity is determined by the degree of producer cost heterogeneity.

One potential limitation of this approach stems from extracting information from the time dimension in the data.[4] Given overwhelming empirical evidence on the expansion in the number and type of available product varieties in the period under study (Schott (2004, 2008)), the built-in assumption is that changes in the set of available varieties have no incidence on the degree of perceived product substitutability. While being fully consistent with the modeling of consumer preferences, this assumption may be violated in practice if entry reduces firm ability to horizontally differentiate their varieties.[5]

The main contribution of this paper is to develop a microfounded approach which relaxes this relatively restrictive assumption and delivers annual estimates of the Armington elasticity. We use the cross-sectional dimension of the widely available data on bilateral trade at the product level to construct an instrument for observed prices. The key intuition is that the introduction of producer cost heterogeneity in the model generates a one-to-one mapping betweeen the number of goods effectively delivered to the world market and fundamental exporter

[3] The formula for consistent aggregation is conceptually similar to Imbs and Méjean (2014).

[4] Time-series variation in relative prices and market shares across the set of exporters to the US market is used in Broda and Weinstein (2006) for 1972-1988 and 1990-2001. In Feenstra et al. (2014) the window is 1992-2007 and information on prices and market shares of domestic US varieties is used together with information on US imports.

[5] Salop (1979), Sutton (2007). Product space crowding is picked up through reduced variance of unobserved consumer taste heterogeneity, i.e. increasing perceived product substitutability (Ackerberg and Rysman (2005)). Romahn (2014) provides evidence in favor of the crowding model for the Swedish beer market in 1996-2002.

ability. Predicted variation in the product set across the set of exporters picks up variation in (unobserved) average product quality. The Armington elasticity is identified in the instrumented specification by estimating the responsiveness of demand to variation in fundamental exporter ability.[6] The extent of perceived substitutability of country-specific product bundles is found to have increased by 25% between 1963 and 2009.[7] In Feenstra et al. (2014) terminology the estimated parameter is the micro-elasticity for aggregate trade. But in our model this micro-elasticity is the only Armington elasticity which enters the trade elasticity formula.

Hence, we document the evolution of the trade elasticity since 1963. Empirical evidence of this kind is notoriously scarse (Head and Mayer (2013)).[8] For comparability with previous work on microfoundations of the gravity equation, we combine cost heterogeneity in the spirit of Eaton and Kortum (2002) with a two-tier CES specification of consumer preferences. The model delivers the gravity formulation in Anderson and van Wincoop (2003) if all goods are traded. It delivers the gravity formulation in Eaton and Kortum (2002) if upper- and lower-tier substitutability coincide and only a subset of goods is traded. In the most general case all three dimensions of heterogeneity contribute to determining the trade elasticity.[9] Results on the Armington elasticity are combined with estimates of lower-tier substitutability to deliver the magnitude of the trade elasticity under the assumption of unchanged producer cost heterogeneity. The trade elasticity has increased by at least 50% in 1963-2009. We cross-check this result with an alternative estimation procedure which directly delivers an estimate of the annual trade elasticity.

The final step consists in computing the annual welfare cost of autarky for each country by combining information on the trade elasticity with information on the share of domestic expenditure (Costinot and Rodriguez-Clare (2014)). Increased reliance on trade partners has not been associated with a significant increase in the welfare cost of autarky. One reason for this is the increase in the perceived substitutability of exchanged product bundles. We conclude that the Armington elasticity has played a role in determining the magnitude of (static) gains from trade between 1963 and 2009.

[6] In the spirit of Feenstra (1994) the parameter is identified from *average* variation in expenditure and *average* price variation as predicted by variation in fundamental exporter ability across the set of exporters in cross-section, with the average constructed for each exporter across the set of observed destination markets.

[7] Archanskaia and Daudin (2012) find a 13 to 33% increase.

[8] Eaton and Kortum (2002), Caliendo and Parro (2012), Simonovska and Waugh (2014a,b) provide estimates for a single year.

[9] This result recalls Costinot and Rodriguez-Clare (2014) and Feenstra et al. (2014). But our specification of consumer preferences differs while also delivering an expression for the trade elasticity bounded below by the Armington parameter and bounded above by the Ricardian parameter.

2 The model

In the Armington set-up countries provide the world market with a composite good produced by homogeneous firms. We relax the assumption of producer homogeneity and nest the production set-up of Eaton and Kortum (2002) within a two-tier CES preference structure. This approach generates a distribution of prices for country-specific sectoral output while maintaining the gravity structure of aggregate trade in Anderson and van Wincoop (2003). If all sectoral goods are traded, the elasticity of aggregate trade to variable trade costs is determined by the elasticity of substitution between country-specific composite goods. If only a subset of goods is traded, the magnitude of the trade elasticity comes to depend on all dimensions of heterogeneity in the economy. The added value of this approach is to identify an empirical counterpart of unobserved country ability. This variable constitutes a theoretically grounded instrument for observed prices and allows obtaining consistent annual estimates of bundle substitutability.

2.1 An Armington model of trade with cost heterogeneity in production

The world contains N countries with labor endowment L_i in each. Output is produced using labor which is perfectly mobile across sectors and immobile across countries. Production technology is non-proprietory within the country and non transferable across countries.

Production technology is linear in labor, with unit labor cost denoted c_i. Output can be produced using one of the production techniques for sector k available in country i. Production techniques vary in efficiency z. Techniques are drawn independently in each sector from a common distribution. For consistency with the assumption of non-proprietory technology, varieties of good k produced within the same country are taken to be perfect substitutes. Constant returns to scale and within-sectoral product homogeneity entail that the best available technique is used in production of each sector within the country. Nonetheless, techniques differ across sectors within the country and across countries for any given sector.

Technology improvement follows the Poisson process described in Eaton and Kortum (2010) whereby at each point in time the number of techniques available for producing output in sector k with efficiency $Z > z$ follows a Poisson distribution with parameter $\lambda_i = T_i(t)z^{-\theta}$. This parameter is increasing in $T_i(t)$ which denotes the stock of technology accumulated by period t and in $1/\theta$ which denotes the extent of dispersion in technology draws[10]. This parameter is

[10] θ is the shape parameter of the Pareto distribution from which efficiency is drawn. A lower θ corresponds to a distribution with a fatter tail, e.g. a higher probability of getting a high draw (Eaton and Kortum (2002, 2010)).

key for our estimation strategy because it defines a one-to-one mapping between fundamental exporter ability and the number of goods which can be produced with efficiency higher than any given threshold z.

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

$$\Pr[Z \leq z] \quad = \quad \Pr[X = 0] = \frac{(\lambda_i)^0 \exp\{-\lambda_i\}}{0!} = \exp\{-\lambda_i\} \tag{1}$$

The probability that a technique of higher efficiency occurs is given by:

$$\Pr[Z > z] \quad = \quad 1 - \Pr[X = 0] = 1 - \exp\left\{-T_i(t) z^{-\theta}\right\} \tag{2}$$

As the process of technology upgrading takes place independently within each sector in the unit continuum, this probability distribution also characterizes the cross-sectoral distribution of best-of ideas in each country. The structure of production thus replicates Eaton and Kortum (2002) wherein techniques effectively used in production are distributed Fréchet. But the structure of preferences, to which we now turn, replicates the Armington hypothesis of product bundles differentiated by place of origin.[11]

Consumer preferences are assumed well represented by a two-tier CES utility function. At the lower-tier country-specific sectoral goods are combined into a composite product bundle. At the upper-tier composite goods of different origin are combined into an aggregate consumption good.[12] Overall utility is:

$$U \quad = \quad \sum_{i=1}^{N} \left\{ Q_i^{(\sigma-1)/\sigma} \right\}^{\sigma/(\sigma-1)} \tag{3}$$

where the source-specific composite good Q_i is:

$$Q_i \quad = \quad \left[\int_0^1 Q_i(k)^{\frac{\sigma'-1}{\sigma'}} \, \mathrm{d}k \right]^{\frac{\sigma'}{\sigma'-1}} \tag{4}$$

Parameter restrictions $1 < \sigma \leq \sigma'$ ensure that finite positive utility is attained in autarky. The welfare cost of autarky corresponds to the change in real welfare brought about by restricting consumption to the domestic composite good (Arkolakis et al. (2012)).

[11] Recall that in Eaton and Kortum (2002) countries supply homogeneous sectoral goods, and the consumer only cares about the combination of least-cost goods in the unit continuum.

[12] This is a departure from the standard set-up in which consumers first make a choice between country-specific sectoral varieties at the lower-tier, and sector-specific composite goods at the upper tier. This choice is made to mimick Anderson and van Wincoop (2003) in which composite goods are differentiated by place of origin.

Define expenditure on the country-specific sectoral good $X_i(k) = P_i(k)Q_i(k)$ and expenditure on the corresponding composite good $X_i = P_iQ_i$. The share of expenditure on each sector is:

$$\frac{X_i(k)}{X_i} = \left[\frac{P_i(k)}{P_i}\right]^{1-\sigma'} \quad (5)$$

where the price index across the unit continuum of sectors is:

$$P_i = \left[\int_0^1 P_i(k)^{1-\sigma'} \mathrm{d}k\right]^{\frac{1}{1-\sigma'}} \quad (6)$$

Alternatively, denoting F_i the price distribution in each source i, we obtain the lower-tier price index by aggregating across the distribution of realized prices:

$$P_i(p) = \left\{\int_0^\infty p^{1-\sigma'} \mathrm{d}F_i(p)\right\}^{\frac{1}{1-\sigma'}} \quad (7)$$

Efficiency is the realization of the random variable Z with independent draws for each sector from the Fréchet distribution with parameter λ. The unit cost of producing k in i is then the realization of the random variable $W = c_i/Z$. Consequently, the number of techniques which allow production of output with cost lower than some threshold w is distributed Poisson with parameter $\lambda_i = T_i(c_i/w)^{-\theta}$ (using $z = c_i/w$) where the time subscript is suppressed given our focus on expenditure allocation in cross-section. Applying (1) the probability of no technique allowing production with cost less than w arriving in a unit interval is given by $\exp\{-\lambda_i\}$. Applying (2) the probability of a lower cost draw arriving is given by $1 - \exp\{-\lambda_i\}$. The distribution of lowest costs is Weibull with parameter λ_i (Eaton and Kortum (2002)):

$$F(w) = \Pr[W \leq w] = 1 - \exp\left\{-T_ic_i^{-\theta}w^{\theta}\right\} \quad (8)$$

and the corresponding pdf is:

$$f(w) = T_ic_i^{-\theta}\theta w^{\theta-1}\exp\left\{-T_ic_i^{-\theta}w^{\theta}\right\} \quad (9)$$

The assumption of perfect competition within each sector entails that the distribution of realized prices is directly given by the distribution of least costs. The structure of preferences entails that all domestic goods survive to compose the country-specific composite good:

$$P_i(p)^{1-\sigma'} = P_i(w)^{1-\sigma'} = \int_0^\infty w^{1-\sigma'} f(w)\,\mathrm{d}w \quad (10)$$

Hence we can use Lemma 2 in Eaton and Kortum (2010) together with parameter restrictions $1 < \sigma \leq \sigma' < \theta + 1$ to compute the price of the country-specific composite good:

$$P_i = \left\{T_i c_i^{-\theta}\right\}^{-1/\theta} \{\Gamma(\gamma)\}^{1/1-\sigma'} \quad (11)$$

where $\gamma = (\theta + 1 - \sigma')/\theta$ is the parameter of the Gamma function[13].

At the upper-tier we get the set-up in Anderson and van Wincoop (2003) whereby each country produces a single country-specific composite good Q_i, and supply of this good is perfectly inelastic. Under the assumption of iceberg trade costs t_{ij}, the scaled price of the composite good delivered from i to j is:

$$\kappa P_{ij} = T_i^{-1/\theta} c_i \tau_{ij} \quad (12)$$

where $\tau_{ij} = 1 + t_{ij}$ and $\kappa = \{\Gamma(\gamma)\}^{-1/(1-\sigma')}$ is a source-invariant scalar.

Denoting total expenditure $Y_j = \sum_{i \in N} P_{ij} Q_{ij}$, the share spent on goods from i is:

$$\frac{X_{ij}}{Y_j} = \frac{(P_{ij})^{1-\sigma}}{\sum_{n=1}^{N} (P_{nj})^{1-\sigma}} \quad (13)$$

where the value of bilateral trade is obtained by maximizing (3) subject to the constraint that expenditure not exceed total income. Total income is given by the landed value of exports from j to all partners $\sum_{n \in N} P_{jn} Q_{jn}$ and is equal to total expenditure in equilibrium.

The gravity structure of aggregate bilateral trade replicates Anderson and van Wincoop (2003) whereby the magnitude of the trade elasticity is determined by the Armington elasticity σ which captures perceived substitutability of country-specific product bundles:

$$X_{ij} = \frac{Y_i Y_j}{Y_w} \left(\frac{\tau_{ij}}{\Pi_i \Phi_j}\right)^{1-\sigma} \quad (14)$$

where Y_w is world expenditure, $\Phi_j = \left[\sum_{n=1}^{N} (P_{nj})^{1-\sigma}\right]^{1/(1-\sigma)}$ is the overall price index of the importer, $\Pi_i = \left[\sum_j s_j (\tau_{ij}/\Phi_j)^{1-\sigma}\right]^{1/(1-\sigma)}$ is the multilateral trade resistance term of the exporter, and $s_j = Y_j / Y_w$ is the expenditure share of each country[14].

[13] To get (11) plug (9) into (10). Use the definition of λ to write $d\lambda = T_i c_i^{-\theta} \theta w^{\theta-1} dw$ and $(\lambda / T_i c_i^{-\theta})^{(1-\sigma')/\theta} = w^{1-\sigma'}$. Change the variable of integration and rearrange (10) to get $P_i^{1-\sigma'} = \left\{T_i c_i^{-\theta}\right\}^{-(1-\sigma')/\theta} \int_0^\infty \lambda^{(1-\sigma')/\theta} \exp\{-\lambda\}\, d\lambda$. The latter integral is equal to $\Gamma[1 + (1-\sigma')/\theta]$.

[14] Follow Anderson and van Wincoop (2003) to get the gravity equation. Use (13), and sum over i's trade partners to get income in source i: $Y_i = \sum_j X_{ij} = \sum_j (T_i^{-1/\theta} c_i \tau_{ij})^{1-\sigma} \Phi_j^{\sigma-1} Y_j$. Solve for $\left(T_i^{-1/\theta} c_i\right)^{1-\sigma} = Y_i \left[\sum_j (\tau_{ij}/\Phi_j)^{1-\sigma} Y_j\right]^{-1}$ and plug the solution back in (13) to get $X_{ij} = Y_i Y_j \left(\frac{\tau_{ij}}{\Phi_j}\right)^{1-\sigma} \left[\sum_j (\tau_{ij}/\Phi_j)^{1-\sigma} Y_j\right]^{-1}$. Multiply and divide the RHS by Y_w, replace Π_i by its value to get (14).

2.2 The incidence of the truncated product set

Only the intensive margin is operational in theory whereby higher production or trade costs leave the set of traded goods unaffected. In practice, product coverage of the world market is highly fragmented (App.A). This section shows that producer cost heterogeneity combined with a two-tier CES preference structure suffices to accomodate zeros as a statistical feature of the data.[15] We derive the lower-tier price index for the truncated product set and show that truncation increases perceived substitutability of country-specific product bundles. We examine the incidence of truncation on the gravity formulation of aggregate bilateral trade and characterize the wedge that truncation introduces between the Armington elasticity and the trade elasticity relevant for observed aggregate trade.

Assume there exists a statistical threshold $\bar{X}$ common to all countries such that the nominal value of sectoral bilateral trade is registered iff it is at least equal to this threshold. Given $\sigma' > 1$, sectors in which the least cost draw is sufficiently high carry marginal weight in expenditure on the exporter-specific composite good. To see this, define $\bar{w}$ as the maximal production cost associated with the smallest observed nominal value and apply (5) to the price of each sectoral good: $X_i(k) \geq \bar{X}$ implies $P_i(k) \leq \bar{w}$. The fraction of high cost draws determines observed bundle variety on the world market. Destination-specific characteristics and bilateral trade frictions determine bilateral variation in bundle variety.

The cost threshold $\bar{w}$ is incorporated in the lower-tier price index to obtain the landed price of the truncated product bundle $\bar{P}_{ij}$:

$$\bar{P}_{ij}(p) \quad = \quad \bar{P}_{ij}(w) = \left\{ \int_0^{\bar{w}} w^{1-\sigma'} f(w)\, \mathrm{d}w \right\}^{1/(1-\sigma')} \tag{15}$$

Recall that the number of draws which allow production with cost less than some threshold $\bar{w}$ is given by $\lambda_i = T_i(c_i/\bar{w})^{-\theta}$. Unit labor cost c_i is augmented with τ_{ij} to take into account bilateral trade frictions. The stock of technology T_i is replaced by $\tilde{z}_i^{\theta}$ to explicitly connect the number of draws to fundamental exporter productivity $\tilde{z}_i$.[16]. The expected bilateral number of draws is $\lambda_{ij} = \left(c_i \tau_{ij}/\tilde{z}_i\right)^{-\theta} \bar{w}^{\theta}$.

The process of techniques' arrival follows Eaton and Kortum (2010) whereby techniques

[15] Both sectoral and aggregate zeros are generated in this way. The alternative approach consists in modifying the production side of the economy to generate structural zeros (Helpman et al. (2008), Eaton et al. (2012)).

[16] The mean of the Fréchet distribution is $T_i^{1/\theta}\Gamma(1-1/\theta)$. Define $\tilde{z}_i$ the scale parameter of the Fréchet distribution: $\tilde{z}_i = T_i^{1/\theta}$, i.e. i's fundamental productivity (Costinot et al. (2012)).

are drawn from a Pareto distribution with parameter θ. The distribution of costs conditional on the cost threshold $\bar{w}$ is $F_c(w) = \Pr(W \leq w | W \leq \bar{w}) = (w/\bar{w})^{\theta}$. The corresponding conditional density function is $f_c(w) = \theta w^{\theta-1} \bar{w}^{-\theta}$.

The landed price of the truncated product bundle is given by the product of the expected number of draws below the cost threshold and the expected cost of such draws (Eaton and Kortum (2010)):

$$\bar{P}_{ij}(w) = \left\{ \underbrace{\lambda_{ij}}_{\text{nb draws}} \underbrace{\int_0^{\bar{w}} w^{1-\sigma'} f_c(w)\, dw}_{\text{expected cost}} \right\}^{1/(1-\sigma')} \tag{16}$$

Plugging the values of λ_{ij} and $f_c(w)$ and solving for the integral expresses the price of the truncated bundle in terms of exporter characteristics, trade frictions, and the cost threshold:

$$\bar{P}_{ij} = \left\{ \frac{\theta}{\theta - \sigma' + 1} \left(\frac{c_i \tau_{ij}}{\tilde{z}_i} \right)^{-\theta} \bar{w}^{\theta - \sigma' + 1} \right\}^{1/(1-\sigma')} \tag{17}$$

Plug (13) into (5) to write the landed value of observed sectoral bilateral trade:

$$X_{ij}(k)\Big|_{X_{ij}(k) \geq \bar{X}} = \left[\frac{P_{ij}(k)}{P_{ij}} \right]^{1-\sigma'} \left[\frac{P_{ij}}{\Phi_j} \right]^{1-\sigma} Y_j \tag{18}$$

The cost threshold is obtained by rearranging (18) to solve for the upper bound of the observed landed sectoral price.

$$P_{ij}(k) \leq \left[\frac{Y_j}{\bar{X}} \Phi_j^{\sigma-1} \right]^{1/(\sigma'-1)} P_{ij}^{\frac{\sigma'-\sigma}{\sigma'-1}} = \bar{w} \tag{19}$$

Plug (12) in (19) to visualize the four components of the cost threshold:

$$\bar{w} = \left[\kappa^{(\sigma'-\sigma)} \bar{X} \right]^{\frac{1}{(1-\sigma')}} \left[Y_j \Phi_j^{\sigma-1} \right]^{\frac{1}{(\sigma'-1)}} \left[\frac{c_i}{\tilde{z}_i} \right]^{\frac{\sigma'-\sigma}{\sigma'-1}} \tau_{ij}^{\frac{\sigma'-\sigma}{\sigma'-1}} \tag{20}$$

Define $\upsilon = \frac{\theta}{\theta - \sigma' + 1} \left[\kappa^{(\sigma'-\sigma)} \bar{X} \right]^{(\theta-\sigma'+1)/(1-\sigma')}$. Plug (20) into (17) to get the landed price of the truncated composite good:

$$\bar{P}_{ij} = \left\{ \upsilon \left[Y_j \Phi_j^{\sigma-1} \right]^{\frac{\theta-\sigma'+1}{\sigma'-1}} \left(\frac{c_i \tau_{ij}}{\tilde{z}_i} \right)^{-(\theta-\alpha)} \right\}^{1/(1-\sigma')} \tag{21}$$

The magnitude of the exponent on exporter ability is $(\theta - \alpha)$ with $\alpha = (\sigma' - \sigma)(\theta - \sigma' + 1)/(\sigma' - 1)$ and $0 \leq \alpha < \theta$.[17]

[17] Parameter restrictions $\theta + 1 > \sigma' \geq \sigma > 1$ entail $\theta > \alpha \geq 0$.

With truncation, the object of interest becomes the degree of perceived substitutability of effectively traded product bundles. Expenditure allocation at the upper tier is obtained by conditioning utility $\bar{U}_j$ to be derived from registered quantities $\bar{Q}_{ij}$ according to the truncated analog of (3). Total expenditure $\bar{Y}_j$ is set equal the sum of registered bilateral imports: $\bar{Y}_j = \sum_i \bar{X}_{ij}$ where $\bar{X}_{ij} = \sum_k X_{ij}(k) \left\{X_{ij}(k) : X_{ij}(k) \geq \bar{X}\right\}$.[18] A vector of trade deficits D_j equalizes truncated expenditure to truncated income: $\bar{Y}_j = \sum_n \bar{P}_{jn}\bar{Q}_{jn} + D_j$. This gives (13) redefined in terms of observed expenditure:

$$\frac{\bar{X}_{ij}}{\bar{Y}_j} = \frac{\left(\bar{P}_{ij}\right)^{1-\bar{\sigma}}}{\sum_{n=1}^{N}\left(\bar{P}_{nj}\right)^{1-\bar{\sigma}}} \tag{22}$$

To show that $\sigma < \bar{\sigma} < \sigma'$, consider relative truncated expenditure on the world market for some pair $\{i, i'\}$ such that $\tilde{z}_i/c_i > \tilde{z}_{i'}/c_{i'}$:

$$\frac{\bar{X}_i}{\bar{X}_{i'}} = \frac{X_i P_i^{\sigma'-1} \int_0^{\bar{w}} p^{1-\sigma'} f_i(p)\mathrm{dp}}{X_{i'} P_{i'}^{\sigma'-1} \int_0^{\bar{w}} p^{1-\sigma'} f_{i'}(p)\mathrm{dp}} \tag{23}$$

Consider the numerator on the right hand side of this expression. The last component is a monotonic transformation of the truncated price index $\bar{P}_i^{1-\sigma'}$. The second component is $P_i^{\sigma'-1} = P_i^{\sigma-1} P_i^{\sigma'-\sigma}$. Since $[X_i/X_{i'}]/(P_i/P_{i'})^{1-\sigma} = 1$, the expression simplifies to:

$$\frac{\bar{X}_i}{\bar{X}_{i'}} = \left[\frac{\bar{P}_i}{\bar{P}_{i'}}\right]^{1-\sigma'} \left[\frac{P_i}{P_{i'}}\right]^{\sigma'-\sigma} < \left[\frac{\bar{P}_i}{\bar{P}_{i'}}\right]^{1-\sigma'} \tag{24}$$

where the inequality is established by $[P_i/P_{i'}] < 1$ given $\tilde{z}_i/c_i > \tilde{z}_{i'}/c_{i'}$. Hence, truncated bundles are perceived to be less substitutable than sectoral goods which compose the bundle. To establish the lower bound, notice that $\bar{P}_i^{1-\sigma'} = \bar{P}_i^{1-\sigma}\bar{P}_i^{\sigma-\sigma'}$:

$$\frac{\bar{X}_i}{\bar{X}_{i'}} = \left[\frac{\bar{P}_i}{\bar{P}_{i'}}\right]^{1-\sigma} \left[\frac{\bar{P}_i}{\bar{P}_{i'}}\right]^{\sigma-\sigma'} \left[\frac{P_i}{P_{i'}}\right]^{\sigma'-\sigma} = \left[\frac{\bar{P}_i}{\bar{P}_{i'}}\right]^{1-\sigma} \left[\frac{\bar{P}_{i'}/P_{i'}}{\bar{P}_i/P_i}\right]^{\sigma'-\sigma} \tag{25}$$

The ratio of truncated to non-truncated prices is always greater than 1.[19] If this ratio were constant across exporters the degree of substitutability would be unchanged. But in practice the magnification factor is monotonically decreasing in scaled exporter ability $\tilde{z}_i/c_i$.[20] Hence,

[18] This simply says that the solution to the non-truncated problem directly gives expenditure allocation in the truncated problem by conditioning on some threshold $\bar{X}$.

[19] This is established by taking the partial derivative of the truncated price index: $\partial\bar{P}_i/\partial w < 0$ and observing that $\bar{P}_i \to P$ when $w \to \infty$.

[20] Focus on the bundle exported to the world market and use (12) and (21) to establish $\bar{P}_i/P_i = \alpha(\tilde{z}_i/c_i)^{-\rho}$ with $\rho = (\theta - \sigma' + 1)(\sigma - 1)/(1-\sigma')^2$ and α constant across exporters.

$[\bar{P}_{i'}/P_{i'}]\,/\,[\bar{P}_i/P_i] > 1$. Perceived substitutability is higher for truncated than for non-truncated product bundles but lower than within-bundle substitutability:

$$\left[\frac{\bar{P}_{i'}}{\bar{P}_i}\right]^{\sigma'-1} > \frac{\bar{X}_i}{\bar{X}_{i'}} > \left[\frac{\bar{P}_{i'}}{\bar{P}_i}\right]^{\sigma-1} \tag{26}$$

Consequently, truncation increases the extent of perceived bundle substitutability differently across trading pairs. Perceived price sensitivity of demand is increasing in the ability gap of the pair. Thus, truncation entails higher price sensitivity of demand for low- relatively to high-ability exporters than within the high-ability group.[21] In practice, the relative magnification factor approaches 1 for all but very small exporters whereby in the full sample $\bar{\sigma} \approx \sigma$.[22]

An additional insight is that zeros are generated in aggregate trade as a consequence of the interplay between σ, σ', and the dispersion in exporter ability. If σ is high, a relatively small share of total world expenditure is allocated to low-ability exporters. If σ' is low, a relatively small share of expenditure on goods produced by low ability exporters is allocated to their least cost goods. As a consequence, the trade flow for the least cost good of the low-ability exporter may be below the registration threshold, and this is all the more likely if the receiving country is itself a low-ability exporter. It follows that high-ability exporters are likely to export positive amounts to low-ability markets while low-ability exporters are more likely to export positive amounts to high-ability markets.

Truncation modifies the elasticity of trade flows to variable trade costs. Plugging (21) in the numerator and denominator of truncated upper-tier demand (22) leaves fundamental exporter characteristics and bilateral trade frictions raised to the power $\varepsilon = -(\theta-\alpha)\bar{\gamma}$ where $0 < \bar{\gamma} = (\bar{\sigma}-1)/(\sigma'-1) < 1$:

$$\frac{\bar{X}_{ij}}{\bar{Y}_j} = \frac{\left(c_i\tau_{ij}/\tilde{z}_i\right)^{-(\theta-\alpha)\bar{\gamma}}}{\sum_{n=1}^{N}\left(c_n\tau_{nj}/\tilde{z}_n\right)^{-(\theta-\alpha)\bar{\gamma}}} \tag{27}$$

The expression of the trade elasticity is simplified by plugging (12) and (21) in (24):

$$\frac{\bar{X}_{ij}}{\bar{X}_{i'j}} = \left[\frac{c_i\tau_{ij}/\tilde{z}_i}{c_{i'}\tau_{i'j}/\tilde{z}_{i'}}\right]^{-(\theta-\alpha)}\left[\frac{c_i\tau_{ij}/\tilde{z}_i}{c_{i'}\tau_{i'j}/\tilde{z}_{i'}}\right]^{\sigma'-\sigma} \tag{28}$$

Defining $\gamma = (\sigma-1)/(\sigma'-1) < \bar{\gamma}$ and rearranging to simplify the exponent gives:

$$\frac{\bar{X}_{ij}}{\bar{X}_{i'j}} = \left[\frac{c_i\tau_{ij}/\tilde{z}_i}{c_{i'}\tau_{i'j}/\tilde{z}_{i'}}\right]^{-\theta\gamma} \tag{29}$$

[21] Waugh (2010) finds that asymmetric trade frictions are needed to rationalize relative expenditure on exports from developed and developing countries. A complementary explanation of this trade pattern is the ability gap which determines relative truncation. The price sensitivity of demand is increasing in the truncation gap.

[22] Simulation results are obtained for a sample with 10000 products and 100 countries, with the max/min ability ratio set at 100, and the parameter range defined as $2 \leq \sigma < 4$ and $4 < \sigma' \leq 8$.

The magnitude of the trade elasticity $|\varepsilon| = \theta\gamma$ is jointly determined by the Armington elasticity, the degree of cost dispersion, and the extent of within-bundle substitutability.[23] Given a change in trade costs, θ regulates the change in the number of observed draws while σ' regulates the incidence of these marginal draws on the price of the truncated composite good. Whenever σ' is relatively high ($\theta/(\sigma'-1) \to 1$), marginal draws have little incidence on the price of the truncated bundle. This dampens the incidence of the extensive margin and magnifies the role of σ in determining $|\varepsilon|$. σ regulates the sensitivity of aggregate expenditure to price differences. If σ is relatively high ($\sigma \to \sigma'$), expenditure allocation is directly determined by the variability of sectoral draws.

The trade elasticity is higher in magnitude than in the benchmark Armington model ($\theta > \sigma' - 1$) but lower than in the benchmark Ricardian model ($\gamma < 1$). Its bounds are determined by the bounds of the Armington elasticity: $|\varepsilon|$ tends to $(\sigma - 1)$ when $\sigma \to 1$ and to θ when $\sigma \to \sigma'$. $|\varepsilon|$ is strictly increasing in the degree of substitutability of non-truncated product bundles, with the impact magnified relatively to the benchmark Armington model.

We pin down the gap in observed and unobserved bundle substitutability using $\theta\gamma = (\theta - \alpha)\bar{\gamma}$. Rearranging and simplifying gives:

$$\frac{\bar{\sigma}}{\sigma} = \frac{\theta(\sigma'-1) - (\sigma'-\sigma)(\theta-\sigma'+1)/\sigma}{\theta(\sigma'-1) - (\sigma'-\sigma)(\theta-\sigma'+1)} \qquad (30)$$

The ratio tends to 1 whenever $(\sigma'-\sigma)(\theta-\sigma'+1) \to 0$. For the magnification factor to be significantly different from 1, it must be that $\sigma \ll \sigma' \ll \theta$. But in this case $\bar{\sigma}/\sigma \ll \sigma'/\sigma$. Hence, $\bar{\sigma}$ is always a better approximation of σ than of σ'.

3 Evolution of perceived bundle substitutability

3.1 Lower-tier substitutability

3.1.1 Estimation of lower-tier substitutability

The lower-tier elasticity of substitution is needed to compute the price of the truncated composite good. We use the structure of the model to estimate exporter-sector productivity draws $z_i(k)$. We then use the fact that these draws should be inversely proportional to sectoral prices to obtain annual estimates of σ'.

In the benchmark model expenditure variation within any bundle is directly determined by the variability of productivity draws: $X_{ij}(k)/X_{ij}(k') = [z_i(k)/z_i(k')]^{\sigma'-1}$. In estimation we work

[23] The interaction of these three dimensions depends on the specification of preferences and therefore differs here from Costinot and Rodriguez-Clare (2014) and Feenstra et al. (2014).

with a flexible specification whereby some component of trade costs may be sector-specific:

$$\frac{X_{ij}(k)}{X_{ij}(k')} = \left[\frac{z_i(k)/\tau_j(k)}{z_i(k')/\tau_j(k')}\right]^{\sigma'-1} \quad (31)$$

Exporter-sector productivity draws $z_i(k)$ are identified from within-bundle expenditure variation by focusing on variation in sectoral expenditure across the set of destination markets:

$$\ln X_{ij}(k) = \eta_0 + f_i(k) + f_j(k) + \eta_{ij} + \eta_{ij}(k) \quad (32)$$

Destination-sector fixed effects $f_j(k)$ pick up systematic variation in sectoral trade costs across the set of destinations together with destination-specific sectoral costs while pair fixed effects η_{ij} pick up bilateral trade costs (Costinot et al. (2012)). Exporter-sector fixed effects $f_i(k)$ identify sectoral productivity $(\sigma'-1)\ln(z_i(k))$ relatively to a benchmark country and sector. In a second step, observed sectoral prices for each pair $P_{ij}(k)$ are regressed on estimated exporter-sector fixed effects $\hat{f}_i(k)$ while controlling for pair-specific determinants of trade with pair fixed effects β_{ij}:

$$\ln P_{ij}(k) = \beta_0 - \zeta \hat{f}_i(k) + \beta_{ij} + \beta_{ij}(k) \quad (33)$$

According to the model within-bundle variation in sectoral prices is determined by variation in sectoral productivity. It follows that $-\zeta \hat{f}_i(k) = -\hat{z}_i(k)$, and the coefficient $-\zeta = -1/(\sigma'_f - 1)$ whereby lower-tier substitutability is identified as $\sigma'_f = 1 + 1/\zeta$.

The alternative is to work with the sectoral price component predicted by sectoral productivity draws $\ln \hat{P}_i(k) = -\zeta \hat{f}_i(k)$. Bilateral sectoral expenditure is then regressed on exporter-specific sectoral prices $\ln \hat{P}_i(k)$ while controlling for pair fixed effects $\tilde{\eta}_{ij}$:

$$\ln X_{ij}(k) = \tilde{\eta}_0 - (\sigma'_s - 1)\ln \hat{P}_i(k) + \tilde{\eta}_{ij} + \tilde{\eta}_{ij}(k) \quad (34)$$

The implicit assumption in the second-stage estimation is again $-\zeta \hat{f}_i(k) = \hat{z}_i(k)$ which implies $\sigma'_f = \sigma'_s$ where the subindex f(s) refers to estimates obtained in the first (second)-stage.

3.1.2 BACI: lower-tier substitutability in 1995-2009

The estimation is first implemented on the BACI dataset. As explained by the authors (Gaulier and Zignago (2010)), BACI has the advantage of providing more complete and accurate information on unit values than the data supplied in UN COMTRADE while maintaining an extensive ($>$ 200 countries) and detailed (HS 6-digit) coverage of bilateral trade in 1995-2009.

The estimation of (32) encounters feasibility constraints because of the sheer number of fixed effects. This problem is solved by estimating (32) separately for each exporter.[24] Lower-tier substitutability is identified by estimating (33) with exporter and destination fixed effects (resp. β_i and β_j) together with the standard set of controls for bilateral trade frictions Z_{ij} (Mayer and Zignago (2011)):

$$\ln P_{ij}(k) \quad = \quad \beta_0 - \zeta \hat{f}_i(k) + \beta_i + \beta_j + Z'_{ij}\beta + \beta_{ij}(k) \tag{35}$$

Analogously, (34) becomes:

$$\ln X_{ij}(k) \quad = \quad \tilde{\eta}_0 - (\sigma' - 1)\ln \hat{P}_i(k) + \tilde{\eta}_i + \tilde{\eta}_j + Z'_{ij}\tilde{\eta} + \tilde{\eta}_{ij}(k) \tag{36}$$

The key identification assumption is that variation in sectoral expenditure within the bundle across the set of destinations identifies underlying sectoral productivity. This is more likely to hold across the set of stable trade relationships. The first trade-off is due to sample truncation. The sample is restricted to observe the bundle sufficiently frequently but not too much to avoid excessive reduction in the number of potential destinations. We evaluate how this trade-off plays out by first conducting the estimation in the balanced and then in the square panel.[25]

The second trade-off is linked to the extent of data disaggregation. To estimate the variation in sectoral expenditure within the bundle each sectoral good has to be observed sufficiently frequently but excessive aggregation may blur the difference between lower and upper-tier substitutability. We investigate the incidence of aggregation by conducting the estimation at the 4-digit (1222 goods) and 2-digit (93 goods) level.[26]

Sufficient restriction of the set of sources (square panel) combined with a higher level of data aggregation (2-digit) allows estimating the full set of dummies in one go which improves precision and consistency of estimated exporter-sector fixed effects. Hence, additional data cleaning is implemented whenever the estimation is conducted for each exporter separately.[27]

Fig.1 and fig.2 present results for the balanced panel. In both figures, point estimates are centered around the same value in the first and the second stage, and the evolution of the point

[24] Sectoral productivity is normalized relatively to the best draw for each exporter. The best draw is defined to be the good observed on the maximal number of destination markets. Exporters observed on less than 10 markets and with less than 40(10) goods at 4(2)-digit are dropped.

[25] The balanced panel is the set of 13651 pairs observed in each year (1995-2009). The square panel is the set of 50 exporters (2450 pairs) which trade with each of the other 49 exporters in each year (1995-2009).

[26] App.C motivates the hierarchical aggregation procedure for unit values within the bundle.

[27] Exporters with $< 50\%$ of significant exporter-sector dummies are dropped. To be consistent with normalizing by the best draw, the estimated distribution is truncated by keeping only negative and significant dummies. Exporters with a correlation coefficient lower than .3 between estimated and standardized dummies are dropped, and only standardized dummies are used in the estimation.

estimate is similar overtime. However, first-stage estimates are characterized by wider confidence intervals, although standard errors are clustered by exporter in both specifications. Further, the magnitude of the lower-tier elasticity is increasing in the degree of data disaggregation: it doubles from around 5 for a 93-good bundle to around 10 for a 1222-good bundle. Finally, we note that lower-tier substitutability is best described as stable between 1995-2009.

Figure 1: BACI balanced panel 4-digit

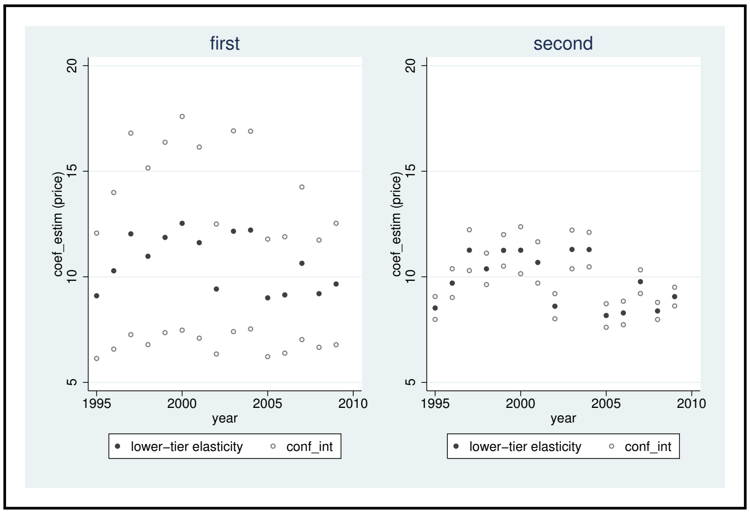


Figure 2: BACI balanced panel 2-digit

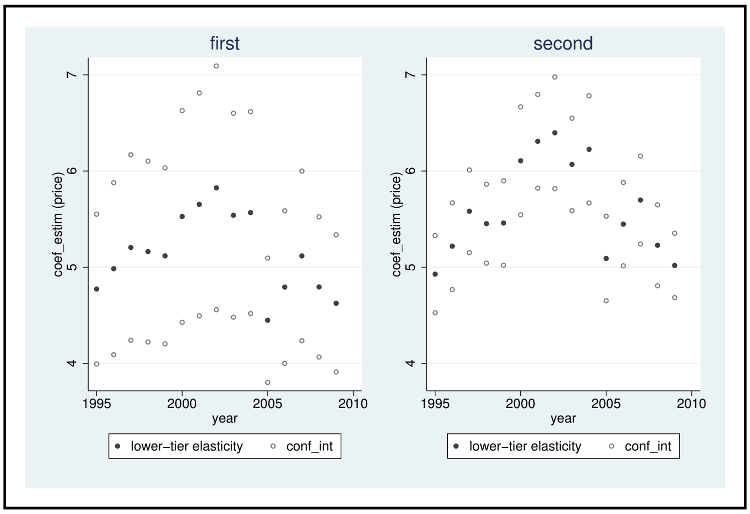
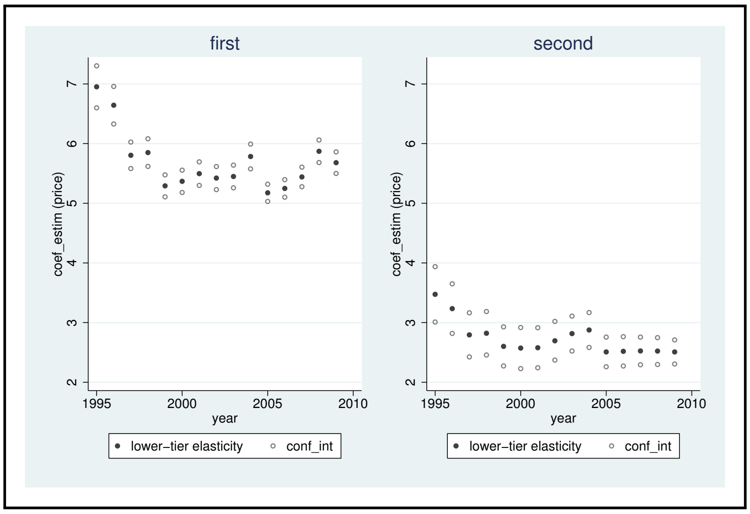


Fig.3 presents results for the square panel at the 2-digit level for which the full set of dummies is estimated simultaneously. The first stage estimate replicates results obtained in the balanced panel (5-6) while the second stage estimate is significantly lower (2-3). Sample truncation appears to result in a downward bias in the second stage estimate while still recovering a consistent estimate in the first stage. Results for the square panel at the 4-digit level for which dummies are estimated separately for each exporter are not shown because the instrument becomes too weak. We conclude that whenever sample truncation leads to a strong reduction in the number of potential destinations, data aggregation must be sufficient to allow simultaneous estimation of the full set of dummies.

Figure 3: BACI square panel 2-digit



3.1.3 UN COMTRADE: lower-tier substitutability in 1963-2009

To go further back in time we work with the UN COMTRADE dataset at the SITC 4-digit level (600-700 goods). Unit value data is arguably a worse proxy of underlying sectoral prices in this sample. The estimation procedure is conducted for the square sample at the 3-digit level (175 goods) and in the stable sample at the 2-digit level (55 goods).[28] In both cases the set of exporter-sector dummies is estimated simultaneously.[29]

Fig.4 presents results for the square panel at the 3-digit level. The first stage estimate (7-10) is situated between estimates obtained in BACI at the 2-digit and 4-digit level. The second stage estimate is again significantly lower (3-4). Further, lower-tier substitutability is found to have been reduced by about 30% in 1963-2009. This reduction occurs between 1963 and 1994, and hence does not contradict results obtained in BACI.

[28] The square sample is defined as the set of 24 sources which trade with every other source in each year between 1963 and 2009. The stable sample is defined as the set of 85 sources which in each year between 1963 and 2009 export to at least 10 destination markets.

[29] The instrument is too weak if dummies are estimated separately for each exporter.

Figure 4: UN COMTRADE square panel 3-digit

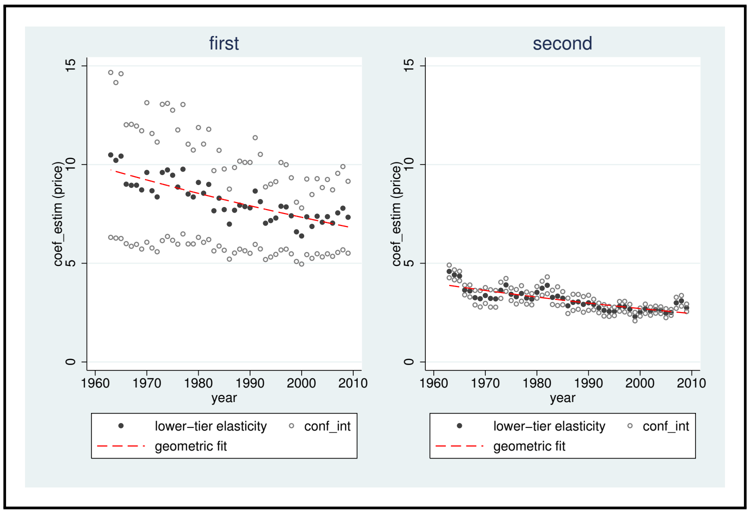
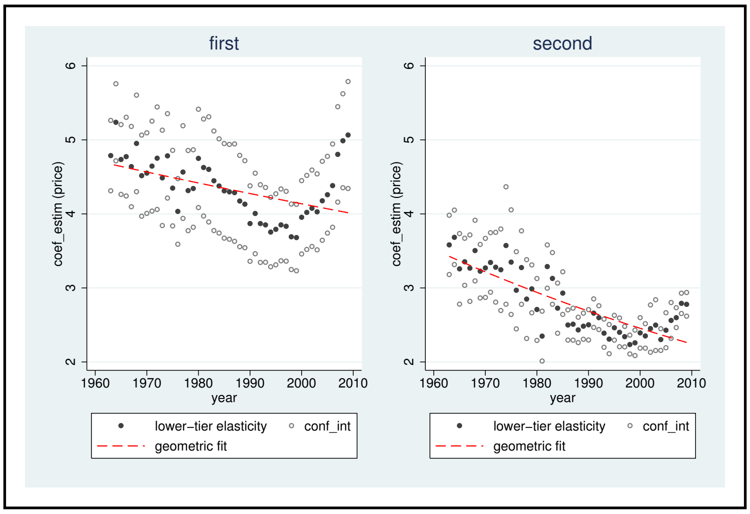


Fig.5 presents results for the stable sample at the 2-digit level. The first stage estimate (4-5) is lower than in the 3-digit sample which is consistent with the higher level of data aggregation. The reduction in lower-tier substitutability is dampened to 14% in 1963-2009 if the first stage estimate is considered while it remains at about 30% in the second stage.

Figure 5: UN COMTRADE stable sample 2-digit



We conclude that lower-tier substitutability is specific to the degree of data disaggregation but relatively stable across data samples. Further, this parameter is found to have been reduced by about 30% in 1963-2009, with the bulk of this decrease occurring before the mid-1990s.

3.2 Upper-tier substitutability

3.2.1 The price of the truncated composite good

The price of the truncated composite good is computed for each bilateral relationship by direct implementation of the formula for a CES price index (15) whereby each observed sectoral price is raised to the power $(1-\sigma')$ and the sum of these components is raised to the power $1/(1-\sigma')$.

As shown in the previous section, lower-tier substitutability is specific to the level of data disaggregation and does not generalize easily to the full sample. Hence, we focus on the two preferred specifications: the balanced BACI sample at the 4-digit level in 1995-2009, and the square UN COMTRADE sample at the 3-digit level in 1963-2009. Further, price aggregation is restricted to the set of bilateral relationships included in the effective estimation of σ'_f. Finally, the first-stage estimate σ'_f is used in price aggregation.

3.2.2 The instrument

A non-instrumented estimation of (22) runs into the classical endogeneity concern of unobserved quality which pushes up observed prices and observed expenditure and introduces a downward bias in the estimate of product bundle substitutability (Feenstra and Romalis (2014); Crozet and Erkel-Rousse (2004)). In the terminology of our model, there will be a downward bias if the true underlying prices which determine expenditure allocation at the upper tier are equal to the ratio of observed prices and unobserved fundamental exporter ability.

The mapping between fundamental exporter ability and bundle variety on the world market helps identify the quality-adjusted component of the observed price. The statistic of interest is $\lambda_{ij} = \left(c_i \tau_{ij}/\tilde{z}_i\right)^{-\theta} \bar{w}^{\theta}$ where we replace the cost threshold by its value using (20) and rearrange to get:

$$\lambda_{ij} \quad = \quad \left[\bar{X}\kappa^{\sigma'-\sigma}\right]^{-\frac{\theta}{\sigma'-1}} \left[Y_j \Phi_j^{\sigma-1}\right]^{\frac{\theta}{\sigma'-1}} \tau_{ij}^{-\theta\gamma} \left[\tilde{z}_i/c_i\right]^{\theta\gamma} \tag{37}$$

The above expression is of interest because it says that conditional on destination-specific characteristics and bilateral trade frictions, the number of goods that the exporter delivers to the world market is increasing in its fundamental ability adjusted for its unit labor cost.

The number of products in the bundle delivered to each destination market is directly observed (see App.A). In both the BACI and the UN COMTRADE dataset this number is defined at the highest available level of data disaggregation (respectively, HS 6-digit and SITC 4-digit).

In both datasets, adjusted fundamental exporter ability is identified by estimating the following relationship in the full sample in each year (time subscripts are omitted for notational convenience):

$$\lambda_{ij} = \exp\left\{\beta_0 + Z'_{ij}\beta + f_j + f_i\right\}\eta_{ij} \tag{38}$$

where β_0 is a constant, Z_{ij} is a vector of conventional bilateral trade cost controls (distance, common language...), and f_n with $n = i, j$ are country fixed effects.

We retrieve coefficients on exporter fixed effects in each year $\hat{f}_i = \theta\gamma\ln(\tilde{z}_i/c_i)$ which capture cost-scaled ability for each exporter relatively to the benchmark country (USA) for which this characteristic is normalized to 1. We combine the resulting sample with information on physical and human capital stocks provided in the Penn World Tables (Feenstra et al. (2013)).[30] While the correlation with human capital stocks is weak (<.2), the correlation with physical capital stocks is strong in both datasets (>.7). Consequently, both estimated fixed effects and logged physical capital stocks are used to instrument observed prices of truncated composite goods.

We isolate the exporter-specific component of bilateral prices by estimating the following relationship in each year:

$$\bar{P}_{ij} \quad = \quad \mu_0 - \mu_1\delta_i + Z'_{ij}\mu + f_j\mu_{ij} \tag{39}$$

where μ_0 is a constant, Z_{ij} is the vector of bilateral trade cost controls, f_j is the destination fixed effect, and $\delta_i = \left\{\hat{f}_i, \ln k_i\right\}$ is the exporter-specific regressor (the standardized exporter-specific dummy or the logged physical capital stock).[31]

Fig.6 presents the estimated $-\mu_1$ for both instruments in BACI for the balanced sample at the 4-digit level. As expected, exporter ability reduces the price of the bundle. The coefficient on bundle variety is centered at -1 which is consistent with the model.

Fig.7 shows that results are qualitatively similar in UN COMTRADE for the square sample at the 3-digit level when the instrument is the logged physical capital stock. On the other hand, the coefficient on bundle variety is neither stable nor centered at -1. This is because bundle coverage is relatively similar in the square sample of UN COMTRADE, and the instrument becomes relatively weak. Consequently, for UN COMTRADE, only logged capital stocks will be used to instrument observed prices.

[30] PWT 8.0 is available for download at http:\www.ggdc.net/pwt.

[31] If both instruments are used only the stock of physical capital is significant.

Figure 6: BACI balanced 4-digit: bundle price and underlying ability

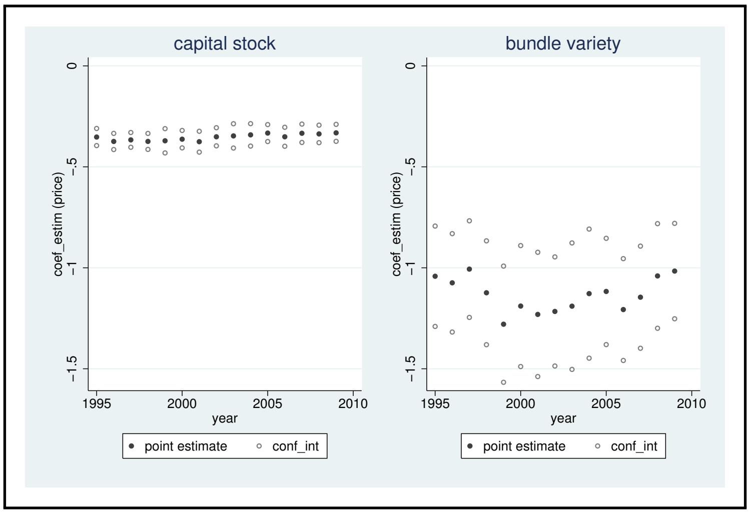
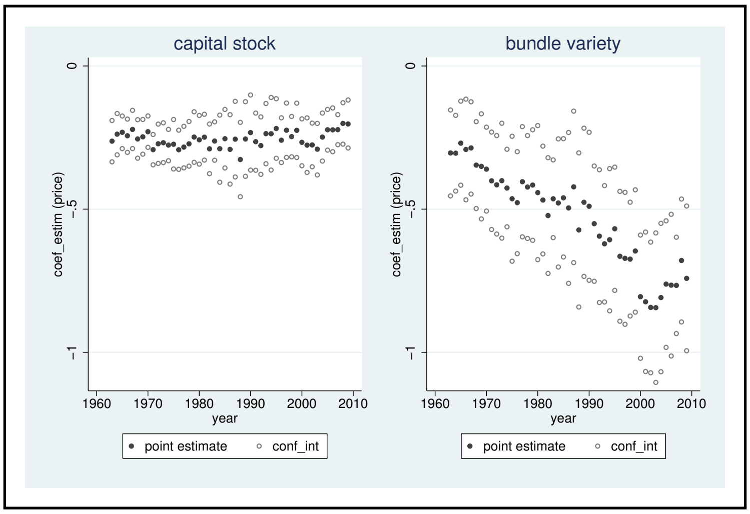


Figure 7: UNC square 3-digit: bundle price and underlying ability



3.2.3 Results on upper-tier substitutability

Upper-tier substitutability is estimated in a standard specification in which total bilateral expenditure is regressed on instrumented exporter-specific prices while controlling for destination fixed effects η_j and bilateral trade frictions:

$$X_{ij} = \exp\left\{\tilde{\eta}_0 - (\sigma - 1)\ln \hat{P}_i + \tilde{\eta}_j + Z'_{ij}\tilde{\eta}\right\}\tilde{\eta}_{ij} \tag{40}$$

The estimation is conducted in PPML which provides a consistent point estimate under unknown form of heteroskedasticity in the trade data (Santos Silva and Tenreyro (2006)).

Fig.8 presents the results for $-(\sigma - 1)$ in BACI for the balanced sample at the 4-digit level.

The magnitude of upper-tier substitutability has increased by 10-20% in 1995-2009.

Figure 8: BACI balanced 4-digit: upper-tier substitutability

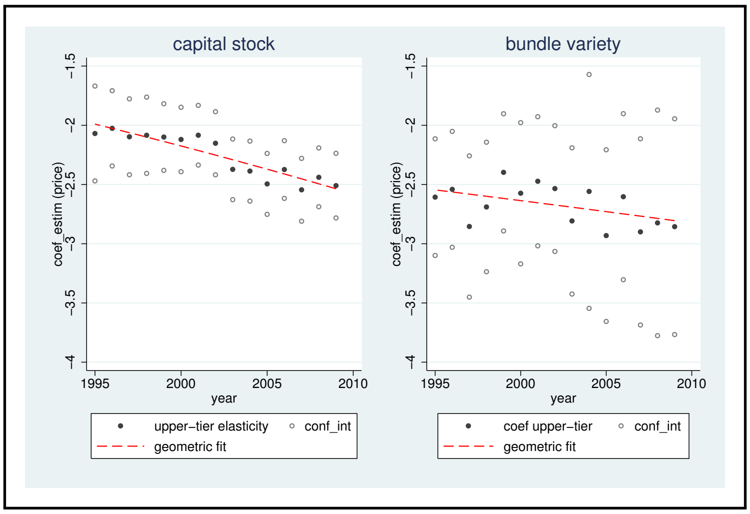
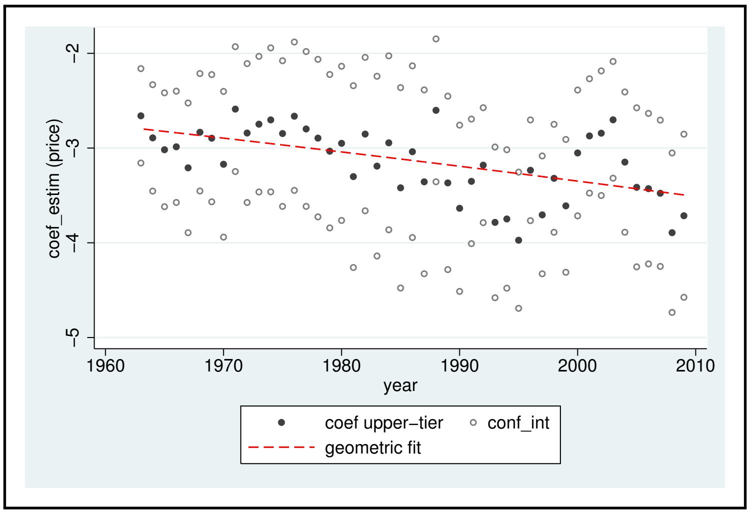


Fig.9 presents the results for $-(\sigma - 1)$ in UN COMTRADE for the square sample at the 3-digit level. The magnitude of upper-tier substitutability has increased by 25% in 1963-2009.

Figure 9: UNC square 3-digit: upper-tier substitutability



Under the assumption of a non-decreasing θ, the combined evolution of upper- and lower-tier substitutability entail a stronger increase in the magnitude of the trade elasticity $|\varepsilon|$ of at least 50% between 1963 and 2009.

4 Evolution of the trade elasticity

4.1 Estimation methodology for $|\varepsilon|$

TO BE COMPLETED

4.1.1 Variation in observed bundle variety

TO BE COMPLETED

4.1.2 Variation in observed expenditure

TO BE COMPLETED

4.2 Implied producer cost heterogeneity

TO BE COMPLETED

5 The welfare cost of autarky

TO BE COMPLETED

6 Conclusion

TO BE COMPLETED

References

Ackerberg, D. A. and M. Rysman (2005). Unobserved product differentiation in discrete-choice models: Estimating price elasticities and welfare effects. *RAND Journal of Economics 36*(4), 771–788.

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

Archanskaia, E. and G. Daudin (2012). Heterogeneity and the distance puzzle. *FIW Working Paper series* (095).

Arkolakis, C., A. Costinot, and A. Rodriguez-Clare (2012). New trade models, same old gains? *American Economic Review 102*(1), 94–130.

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

Bems, R. (2014). Intermediate inputs, external rebalancing, and relative price adjustment. *unpublished manuscript*.

Bems, R., R. C. Johnson, and K.-M. Yi (2010). Demand spillovers and the collapse of trade in the global recession. *IMF Economic Review 58*(2), 295–326.

Bems, R., R. C. Johnson, and K.-M. Yi (2011). Vertical linkages and the collapse of world trade. *American Economic Review: Papers and Proceedings 101*(3).

Bernard, A. B., J. Eaton, B. J. Jensen, and S. Kortum (2003). Plants and productivity in international trade. *American Economic Review 93*(4), 1268–1290l.

Bernard, A. B., B. J. Jensen, S. J. Redding, and P. K. Schott (2012). The empirics of firm heterogeneity and international trade. *Annual Review of Economics 4*(1), 283–313.

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

Caliendo, L. and F. Parro (2012). Estimates of the trade and welfare effects of nafta. *NBER Working Paper Series* (18508).

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

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

Costinot, A. and A. Rodriguez-Clare (2014). Trade theory with numbers: Quantifying the consequences of globalization. In G. Gopinath, E. Helpman, and K. Rogoff (Eds.), *Handbook of International Economics*, Volume 4, pp. 197–262. Amsterdam: Elsevier.

Crozet, M. and H. Erkel-Rousse (2004). Trade performances, product quality perceptions, and the estimation of trade price elasticities. *Review of International Economics 12*(1), 108–129.

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

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

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from french firms. *Econometrica 79*(5), 1453–1498.

Eaton, J., S. Kortum, and S. Sotelo (2012). International trade: Linking micro and macro. *NBER Working Paper Series* (17864).

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

Feenstra, R., R. Inklaar, and M. Timmer (2013). The next generation of the penn world table. *unpublished manuscript*.

Feenstra, R. C., M. Obstfeld, and K. N. Russ (2014). In search of the armington elasticity. *NBER working paper series* (20063).

Feenstra, R. C. and J. Romalis (2014). International prices and endogenous quality. *Quarterly Journal of Economics 129*(2), 477–527.

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

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

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

Imbs, J. and I. Méjean (2014). Elasticity optimism. *American Economic Journal: Macroeconomics (forthcoming).*

Johnson, R. C. and G. Noguera (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of International Economics 86*, 224–236.

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

Melitz, M. and S. Redding (2013). New trade models, new welfare implications. *NBER Working Paper Series* (18919).

Melitz, M. and S. Redding (2014). *Handbook of international economics*, Volume 4, Chapter Heterogeneous Firms and Trade. Elsevier.

Romahn, A. (2014). Unobserved crowding in product space: an application to the swedish beer market. *unpublished manuscript*.

Salop, S. C. (1979). Monopolistic competition with outside goods. *The Bell Journal of Economics 10*(1), 141–156.

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

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

Schott, P. (2008). The relative sophistication of chinese exports. *Economic Policy 23*(53), 5–49.

Simonovska, I. and M. Waugh (2014a). The elasticity of trade: Estimates and evidence. *Journal of International Economics 92*(1), 34–50.

Simonovska, I. and M. Waugh (2014b). Trade models, trade elasticities, and the gains from trade. *unpublished manuscript*.

Soderbery, A. (2010). Investigating the asymptotic properties of elasticity of substitution estimates. *Economic Letters 190*(2), 57–62.

Soderbery, A. (2012). Estimating import supply and demand elasticities: Analysis and implications. *unpublished manuscript*.

Sutton, J. (2007). *Handbook of Industrial Organisation*, Volume 3, Chapter Market Structure - Theory and Evidence. North Holland.

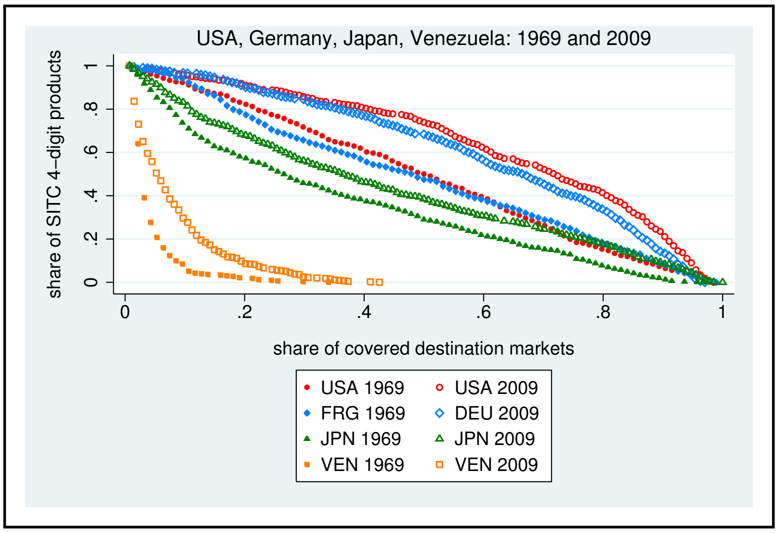
Waugh, M. (2010). International trade and income differences. *American Economic Review 100*(5), 2093–2124.

A Product coverage and product overlap

We work with the SITC 4-digit classification in UN COMTRADE because it provides the longest possible year coverage. The first stylized fact provides evidence of fragmented product coverage of world markets. For each exporter and year we compute the rate of decay in the share of SITC 4-digit goods which reach each additional destination market. The average rate of decay in the sample is 1.13: a one percentage point increase in the number of served markets is associated with a 1.13 percentage point reduction in product bundle variety. Hence, two salient features of the data are the prevalence of zeros in sectoral trade and the fact that for any given exporter only a small fraction of the observed product bundle reaches all destination markets.

The rate of decay in bundle variety is lower for exporters which have the best coverage of world variety where world variety is defined as the total number of SITC 4-digit goods imported by at least one destination. In the case of USA and Germany which have full coverage of world variety in 1969, 80% of their product bundle reaches 20% of their export markets. This figure is reduced to 60% for Japan which covers 96% of world variety, and to less than 5% for Venezuela which covers 54% of world variety in 1969. Nonetheless, product coverage improves for all exporter types as shown by the systematic rightward shift of the distribution between 1969 and 2009. The average rate of decay is reduced from 1.28 in 1963 to 1.05 in 2009. Fig.10 illustrates these patterns.

Figure 10: Evolution of product coverage in world markets



Tab.1 illustrates the reduction in the share of zero trade flows (ztf) in terms of world variety coverage in (1), and destination-specific variety coverage in (2). Col.(3) shows that the reduction in the share of ztf is proceeding at a quicker pace in the sample of stable pairs, i.e. pairs which trade both ways in each year of the sample.

Table 1: **Bundle variety and bundle similarity in 1963-2009**

depvar:	Share of ZTF			Overlap: count		Overlap: value	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
year	−0.0074♯	−0.0025♯	−0.0021♯	0.0042♯	0.0168♯	0.0048♯	0.0161♯
stable			−0.0042♯			−0.0022♯	0.0017♯
Obs	9,381	645,659	645,659	631,656	636,207	631,651	636,202
R^2	0.94	0.61	0.64	0.39	0.54	0.28	0.42
Ref.				USA	CHN	USA	CHN

Notes: The share of ZTF is computed at the SITC 4-digit level for the world market in (1), separately for each destination market in (2)-(3). Standard errors are clustered by source in (1), by pair in (2)-(9). Source fixed effects are always included. Destination fixed effects are included in (2)-(9). ♯: p<0.01. 'stable' indicates the change in the slope for pairs which trade both ways in every year.

In (4)-(7) we investigate the evolution of bilateral product overlap. Product overlap is defined in (4)-(5) as the share of product categories supplied by the source as well as by the reference country (USA in (4), China in (5)) to each destination market. Product overlap is defined in (6) (resp.(7)) as the share of bilateral trade which occurs in product categories also supplied by the USA (resp. China). We find that the degree of similarity in the composition of country-specific product bundles is increasing independently of the choice of the reference country[32].

Tab.2 is reproduced from Archanskaia and Daudin (2012) to further illustrate the one-to-one mapping between expenditure and bundle variety whereby the share of sectoral zeros is decreasing in the market share of the exporter. The predicted share of ztf is .95 in 1963 (.83 in 2009) for an exporter with .02% market share while it is .72 in 1963 (.65 in 2009) for an exporter with 10% market share. This is consistent with the model in which expenditure and

[32] Product overlap increases more quickly for pairs in the superbalanced sample whenever the reference country has sufficiently low initial coverage of world variety such as China (see (7)) or Brazil (not shown).

bundle variety are simultaneously determined by unobserved fundamental exporter ability.

Table 2: **Predicted share of ztf for exporters with different market share, 4-digit level**

year	ms=0.02%	ms=1%	ms=10%	ms=28.7%
1963	0.95	0.80	0.72	0.69
2009	0.83	0.71	0.65	0.62

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

To sum up, the two salient features of the data are the prevalence of zeros in sectoral trade and the increasing product overlap in country-specific composite goods. The complementary finding is that exporters with better coverage of total world variety export a higher fraction of this bundle to any given market and hence export more goods on average to any given market. Average bundle variety has a one-to-one mapping to unobserved exporter ability in the model.

B Frequency, duration, and reciprocity of trade relationships

B.1 The UN COMTRADE dataset: 1963-2009

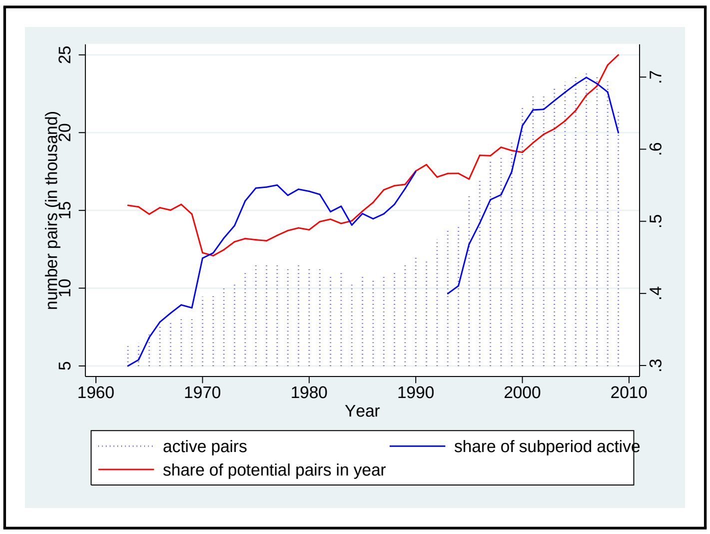
Note: Descriptive information on the UN COMTRADE dataset provided in this appendix is reproduced from Archanskaia and Daudin (2012). The reader is referred to that paper for the investigation of sample and product composition effects on the distance elasticity of trade.

The full sample contains 207 reporters (R) and 230 partners (P) which are listed in tables 3 and 4 below. 'S' indicates that the country is present in the superbalanced sample. 'SQ' indicates that the country is also present in the square 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.

Fig.11 summarizes the main features of the data. The number of active pairs increases more than fourfold in 1963-2009 (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 70% of the total number of possible trade relationships, with a clear upward trend (in red, right scale).

Figure 11: Active pairs in COMTRADE (1963-2009)



If we focus on the set of pairs which trade a positive amount in at least one year of the sample, the share of active pairs increases by 20 percentage points between 1963 and 1990 and by 20 percentage points between 1993 and 2009 (in blue, right scale)[33]. By the end of the sample about 2/3 of pairs which trade at least once between 1963 and 2009 are reporting non-zero trade.

Fig.12 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.

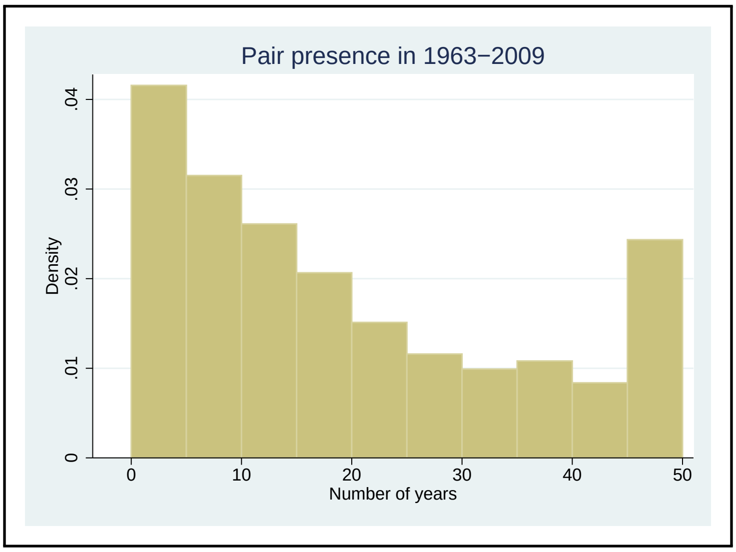


Figure 12: Number of years each pair is present in the sample

The bulk of total trade is attributable to the subsample of pairs which trade both ways in each and every year in 1963-2009. We refer to this set of stable reciprocal pairs as the 'superbalanced' sample. To avoid discarding pairs which fall out of the superbalanced sample because countries split up or reunite at some point in 1963-2009, several additional single country identifiers are introduced.[34]

The superbalanced sample comprises 1056 active pairs and corresponds to 37 countries. This is less than the 1332 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

[33] We split the sample in two subperiods, 1963-1990 and 1993-2009, to take into account country creation and disappearance in the early 1990s.

[34] A single identifier is used for for East, West, and reunited Germany. Similarly, we use a single identifier for the Czech Republic, Slovakia, and Czechoslovakia; and another single identifier for the USSR and the 15 countries which were formed after 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 they do not have reciprocal trade in all years with another country of the sample.

year and which is referred to as the ‘square sample’ comprises just 24 countries (552 pairs). Trade coverage in the superbalanced sample decreases from 70 to 50% of total trade while it is reduced from 60 to 40% for the square sample (see Fig.13).

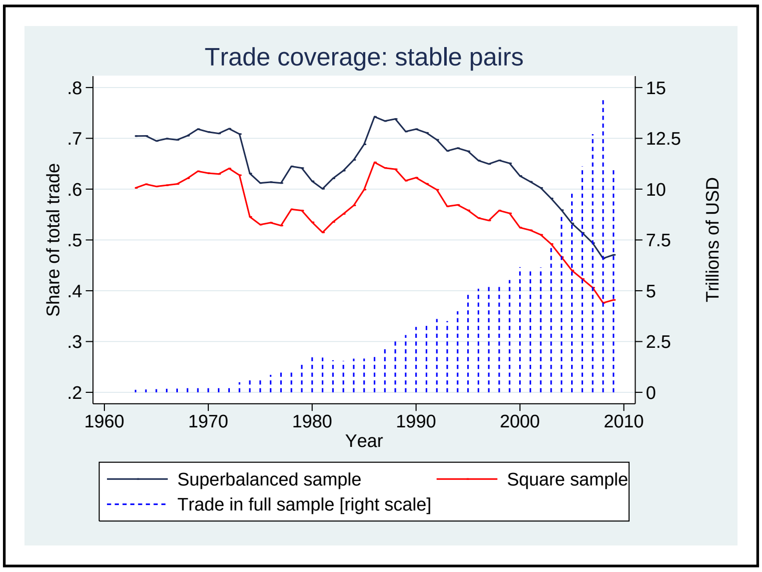


Figure 13: Trade coverage in 1963-2009

B.2 The BACI dataset: 1995-2009

TO BE COMPLETED

Table 3: Country list in UN COMTRADE: full, superbalanced, and square 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;SQ**	Panama	*R;P*
Antigua-Barbuda	*R;P*	Ghana	*R;P*	Papua New Guinea	*R;P*
Argentina	**R;P;S**	Gibraltar	P	**Paraguay**	**R;P;S**
Armenia	*R;P*	**Greece**	**R;P;SQ**	Peru	*R;P*
Aruba	*R;P*	Greenland	*R;P*	**Philippines**	**R;P;SQ**
Australia	**R;P;SQ**	Grenada	*R;P*	Pitcairn	P
Austria	**R;P;SQ**	Guadeloupe	*R;P*	Poland	*R;P*
Azerbaijan	*R;P*	Guatemala	*R;P*	**Portugal**	**R;P;SQ**
Bahamas	*R;P*	Guinea	*R;P*	Qatar	*R;P*
Bahrain	*R;P*	Guinea-Bissau	*R;P*	Reunion	*R;P*
Bangladesh	*R;P*	Guyana	*R;P*	Romania	*R;P*
Barbados	*R;P*	Haiti	*R;P*	Russian Federation	*R;P*
Belarus	*R;P*	Honduras	*R;P*	Rwanda	*R;P*
Belgium	**R;P;SQ**	**Hong Kong**	**R;P;SQ**	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;SQ**	Saudi Arabia	*R;P*
Brazil	**R;P;S**	**Israel**	**R;P;SQ**	Senegal	*R;P*
Br. Virgin Islands	P	**Italy**	**R;P;SQ**	Serbia-Montenegro	*R;P*
Brunei Darussalam	*R;P*	Jamaica	*R;P*	Seychelles	*R;P*
Bulgaria	*R;P*	**Japan**	**R;P;SQ**	Sierra Leone	*R;P*
Burkina Faso	*R;P*	Jordan	*R;P*	**Singapore**	**R;P;S**
Burma	*R;P*	Kazakstan	*R;P*	Slovakia	*R;P*
Burundi	*R;P*	Kenya	*R;P*	Slovenia	*R;P*
Cambodia	*R;P*	Kiribati	*R;P*	Solomon Islands	*R;P*
Cameroon	*R;P*	**Korea**	**R;P;S**	Somalia	*R;P*
Canada	**R;P;SQ**	DPR of Korea	P	South Africa	*R;P*

Table 4: 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;SQ**
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;SQ**
Colombia	**R;P;S**	Lithuania	*R;P*	**Switzerland**	**R;P;SQ**
Comoros	*R;P*	**Luxembourg**	**R;P;S**	Syria	*R;P*
Congo	*R;P*	Macau (Aomen)	*R;P*	Taiwan	*R;P*
Dem. Rep. of Congo	*R;P*	Macedonia	*R;P*	Tajikistan	*R;P*
Cook Islands	*R;P*	Madagascar	*R;P*	Tanzania	*R;P*
Costa Rica	*R;P*	Malawi	*R;P*	**Thailand**	**R;P;S**
Croatia	*R;P*	**Malaysia**	**R;P;SQ**	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;SQ**	Mauritania	*R;P*	Turkmenistan	*R;P*
Djibouti	*R;P*	Mauritius	*R;P*	Turks-Caicos Islands	*R;P*
Dominica	*R;P*	**Mexico**	**R;P;SQ**	Tuvalu	*R;P*
Dominican Republic	*R;P*	Micronesia	P	Uganda	*R;P*
East Germany	**R;P;S**	Moldova	*R;P*	Ukraine	*R;P*
East Timor	*R;P*	Mongolia	*R;P*	United Arab Emirates	*R;P*
Ecuador	*R;P*	Montserrat	*R;P*	**United Kingdom**	**R;P;SQ**
Egypt	*R;P*	Morocco	*R;P*	**USA**	**R;P;SQ**
El Salvador	**R;P;S**	Mozambique	*R;P*	Uruguay	*R;P*
Equatorial Guinea	P	Namibia	*R;P*	Uzbekistan	P
Eritrea	*R;P*	Nauru	P	Vanuatu	*R;P*
Estonia	*R;P*	Nepal	*R;P*	**Venezuela**	**R;P;S**
Ethiopia	*R;P*	Netherland Antilles	*R;P*	Vietnam	*R;P*
Falkland Islands	P	**Netherlands**	**R;P;SQ**	Wallis-Futuna	*R;P*
Faroe Islands	*R;P*	New Caledonia	*R;P*	**West Germany**	**R;P;S**
Fiji	*R;P*	New Zealand	*R;P*	Western Sahara	P
Finland	**R;P;SQ**	Nicaragua	*R;P*	Yemen	*R;P*
Fm Vietnam DR	*R;P*	Niger	*R;P*	Yugoslavia	*R;P*
Fm Vietnam Rp	*R;P*	Nigeria	*R;P*	Zambia	*R;P*
France	**R;P;SQ**	Niue	P	Zimbabwe	*R;P*
French Guiana	*R;P*	Norfolk Island	P		

C Unit value aggregation within exporter bundles

TO BE COMPLETED