Proximity as a Source of Comparative Advantage *

Liza Archanskaia†
SciencesPo-OFCE

June 2013

Abstract

This paper establishes that production unbundling has coincided with an inscreasing role of input costs in shaping the pattern of comparative advantage. I show that the wedge in the cost of the input bundle across countries in a multisectoral Ricardian model is given by a composite index of trade frictions incurred in sourcing inputs. As the cost share of inputs is sector-specific this wedge becomes source of comparative advantage whereby countries characterized by relatively high proximity to input suppliers specialize in sectors which use inputs more intensively. I find robust empirical evidence that the input cost channel has growing importance over 1995-2009. Nonetheless, consistently with the fundamental intuition of Ricardian models, the ranking of relative sectoral technology stocks still determines intersectoral specialization. Between 53-55% of intersectoral variation in relative sectoral exports is explained by technology while the contribution of the input cost channel increases from 3 to 8% in the full sample, and from 3 to 13% for the EU-15.

Keywords: Ricardian model, Intersectoral specialization, Trade costs
JEL codes: F10,F15

---

*I am grateful to Roberto Bonfatti, Thomas Chaney, Maggie Chen, Arnaud Costinot, Mathieu Crozet, Stefania Garetto, Christian Gormsen, James Harrigan, Jacques Le Cacheux, Philippe Martin, Thierry Mayer, and participants of the Rocky Mountain Empirical Trade Conference, the RIEF, the GEP postgraduate conference, as well as seminar participants at KU Leuven and SciencesPo Economics department for advice and suggestions for improving this paper. The usual disclaimer applies.

†Email: archanskaia@gmail.com. Affiliation: PhD candidate at SciencesPo Economics department, affiliated with the French Economic Observatory (OFCE).

Introduction

What this paper does

This paper belongs to the strand of literature which, following the seminal work by Eaton and Kortum (2002, 2010) and Costinot (2009), seeks to identify the relative importance of technology, factor endowments, and trade costs in determining the pattern of specialization on world markets in many-good many-country Ricardian models. This approach allows defining a theoretically grounded measure of revealed comparative advantage, as in Chor (2010) and Costinot et al. (2012), and equips the researcher with a flexible tool to quantify the relative importance of fundamental country characteristics in determining the pattern of intersectoral specialization.

Costinot (2009) has developed a unifying framework which delivers a strong result in terms of intersectoral specialization when the primitives of the model which are technology and factor endowments are characterized by logsupermodularity. Specifically, if countries can be ranked in terms of a single characteristic, such as the quality of their institutions, and sectors can be ranked according to a sector-specific characteristic, such as their skill intensity, then if the primitives of the model are such that high-characteristic countries are relatively more likely to be endowed with factors which are relatively more productive in high-characteristic sectors, it is possible to deduce the pattern of specialization in terms of the ranking of relative sectoral output for any pair of countries.

Consequently, Costinot et al. (2012) have shown that the seminal Eaton and Kortum (EK) model extended to a multi-sector set-up with a finite number of sectors and an infinite countable number of differentiated varieties within each sector generates the stark prediction that the ranking of relative sectoral exports for any pair of countries on world markets can be predicted from the ranking of their relative sectoral technology stocks.[1]

The contribution of this paper is to show that with multistage production and trade in inputs, the proximity of the country to world technology defined as its ability to source least cost inputs worldwide, becomes a fundamental country characteristic which co-determines its pattern of comparative advantage, together with domestic technology and labor endowments.

We incorporate multistage production in the Costinot et al. (2012) model in the simplest possible way by assuming that output in each sector is produced using a bundle of inputs and labor. We incorporate a sector-specific

[1]This prediction is obtained under the assumption that bilateral trade costs contain a pair specific component common across sectors and a sector-specific component specific to the destination and common across exporters.

feature in the production function by assuming that sectors differ in the way they combine inputs and labor in production.

We show that the only component of the cost of the input bundle which varies across countries is a composite index of trade frictions which the country faces in sourcing inputs from all possible suppliers including itself. As the cost share of inputs is sector-specific, proximity to suppliers matters relatively more in sectors which use inputs relatively more intensively.

Consequently, the interaction of a sector-specific characteristic, the weight of inputs in gross output, with the trade cost magnification channel which works through trade in inputs, generates a ranking whereby countries characterized by high proximity to least cost inputs specialize in sectors which use inputs relatively more intensively, conditional on the distribution of domestic technology and labor endowments.

The magnifying effect of trade frictions in the context of cross-border vertical production segmentation has been studied by Yi (2010). In Yi (2010), this mechanism contributes to determining the co-location of the two production stages, inputs and assembly, and the extent of vertical specialization in countries' trade. In this paper we do not learn much about vertical specialization, but we gain mileage in the ability to separately identify the contributions of domestic technology and proximity to suppliers in determining the pattern of intersectoral specialization.

The Empirical Application

In the empirical analysis, we study the pattern of revealed comparative advantage of the main trade partners of the European Union in 1995-2009. We show how to bring the model to the data to quantify the relative weight of fundamental characteristics which determine the pattern of comparative advantage: domestic sectoral technology stocks, labor endowments by skill, and proximity to world technology. This additional component of comparative advantage which we refer to as the 'proximity mechanism' plays out through differences in the relative ease with which countries can source inputs from the best possible supplier of each variety in the world, interacted with the input intensity characteristic of the sector.

The empirical investigation proceeds in four steps. First, the model is used to derive a theoretically grounded measure of proximity to suppliers for each country which, brought to the data, is found to be very persistent overtime. It establishes a ranking of countries in our sample which reveals relatively high centrality of European countries, and of Central and Eastern European countries in particular, while non-European emerging economies such as China, Brazil, and Mexico are characterized by relatively low centrality. Conceptually, in relative terms, the proximity characteristic is a summary statistic of locational comparative advantage because it captures the cost advantage conferred to the country through its ability to source the cheapest inputs worldwide, relatively to every other country in the world.

Second, we implement a fixed effects approach suggested by Costinot et al. (2012) to identify exporter-sector specific relative production costs which in the framework of our model contain four components: technology, wages, input costs, and exporter-specific trade costs which correspond to the trade restrictiveness the exporter faces on world markets.

Third, we project these relative production costs on the vectors of instrumented sectoral technology stocks and wages to identify the cost component unexplained by technology and factor endowments. In this step of the estimation we obtain the structural parameters of the model: the degree of dispersion in productivity and sectoral input intensities. Our preferred point estimates for the dispersion parameter, 6.7(.4) and 7.3(.5), are consistent with values obtained by previous studies.[2] Estimated input intensities are found to be strongly correlated with the share of expenditure on inputs in gross output computed at the sectoral level.

Fourth, we split the sample in two groups according to the proximity characteristic, and regress estimated residuals of relative sectoral production costs on relative proximity for each pair of exporters while interacting proximity with the input intensity of the sector. If the model correctly describes the pattern of production, the proximity mechanism should determine the pattern of intersectoral specialization conditional on domestic technology and labor costs.

We find robust empirical evidence that countries characterized by relatively high proximity to suppliers specialize in sectors which use inputs relatively more intensively. Further, we find that the proximity mechanism becomes a stronger predictor of relative sectoral rankings in the recent period (2002-2009).

Deardorff (2004) establishes a distinction between 'global comparative advantage' defined through relative labor requirements in production under frictionless trade and 'local comparative advantage' defined through relative labor requirements for production of landed goods under positive trade costs. Trade costs are paid in local labor inducing changes in relative labor requirements for landed goods.

An additional contribution of this paper is to check whether the pattern of specialization determined by local comparative advantage, i.e. the

[2] The preferred point estimate in Eaton and Kortum (2002) (resp. Costinot et al. (2012)) is 8.3 (resp. 6.5). Caliendo and Parro (2012) find 8.2 for manufacturing.

pattern of specialization under positive trade costs and trade in inputs, is different from the pattern which would prevail in a world with trade in inputs but without trade frictions. We decompose the variance of pairwise revealed comparative advantage rankings in the share due to domestic technology, factor endowments, and proximity. The proximity characteristic is indeed a summary statistic of trade frictions which co-determine the pattern of specialization under local comparative advantage by modifying expected sectoral production costs.

The main result is that the pattern of comparative advantage observed in a world with positive trade costs and trade in inputs conforms to the specialization pattern which would prevail at the intersectoral level in a frictionless world.[3] Consistently with the fundamental intuition of Ricardian models, the ranking of relative sectoral technology stocks determines the pattern of intersectoral specialization even under positive trade costs. Nonetheless, sector-specific cost differences induced by the proximity mechanism matter increasingly overtime.

Complementarity to Recent Studies

Our results are complementary to several recent empirical investigations of the mechanisms which shape the pattern of intersectoral specialization.

Harrigan and Evans (2005) and Harrigan (2010) provide empirical evidence for the US market on a demand-side mechanism which shapes countries' specialization on specific destination markets. In their framework, products can be ranked in terms of consumer preference for timely delivery and countries can be ranked in terms of distance to destination, with partners situated closeby characterized by their ability to provide timely delivery (or, alternatively, to provide it at a relatively lower cost). The model predicts that local partners will specialize in sectors where timely delivery is valued relatively more by consumers. In this paper we investigate a different but potentially complementary mechanism of intersectoral specialization driven by proximity to suppliers.

[3] Eaton and Kortum (2002) find that in 1990 the world was on a brink of a transition from a situation in which geography played a determining role in defining countries' specialization to the situation in which specialization would be driven by technology. These authors work with a one-sector economy, and define specialization as the labor share in manufacturing. In this paper, we describe specialization patterns within manufacturing. In conformity with the intuition of these authors, we find that specialization across manufacturing sectors is driven by technology even though trade frictions continue to play a non-negligible role.

Johnson and Noguera (2012a,c) document that production linkages in conjunction with proximity play an important role in shaping the pattern of bilateral trade. The authors find that the intensity of international production sharing in bilateral relationships, measured as the fraction of value added in gross exports, is increasing in proximity between source and destination. However, the authors do not investigate whether the extent of production sharing constributes to determining the pattern of countries' intersectoral specialization on world markets. Consequently, they do not check whether the extent of production sharing in each bilateral relationship can be summarized by a synthetic index of trade frictions incurred in sourcing inputs. The analysis conducted in this paper is complementary in that we provide a characterization of the aggregate effect of all bilateral production sharing relationships on the cost of the input bundle in each country.

Chor (2010) works in the framework of a multi-sector Ricardian model to quantify the relative importance of the channels which shape the pattern of intersectoral specialization by determining relative sectoral technology stocks. Consistently with Costinot (2009), the author specifies a functional form which determines sectoral technology stocks as a function of several complementarity mechanisms between country and sector characteristics.[4] In the empirical analysis, Chor (2010) identifies the relative contribution of these different dimensions of complementarity to determining the pattern of specialization. In this paper, we conduct a complementary decomposition exercise in that we provide evidence on the relative contribution of domestic technology and proximity to world technology in determining specialization without opening the black box of what technology is.[5]

Caliendo and Parro (2012) develop a multisector Ricardian model with multistage production and trade in inputs to identify the impact of tariff reductions due to NAFTA on changes in trade patterns and welfare of the US, Canada, and Mexico. Caliendo and Parro (2012) underscore the importance of trade in inputs and intersectoral input-output linkages in magnifying the gains from trade. Indeed, one of the key results of the paper is that welfare gains are 40% lower if this magnification mechanism is unaccounted for.

[4]Examples are: the interaction of institutional quality and skilled labor endowment with the technological complexity of the sector; the degree of development of financial markets and the degree of sectoral reliance on external financing.

[5]Chor (2010) looks at the impact of inputs on comparative advantage through the lens of incomplete contracts' theory. He finds that countries with high quality legal systems have a comparative advantage in sectors which use relationship specific inputs relatively more intensively with the idea that such sectors are more dependent on law enforcement efficiency. This mechanism is very different from the input intensity mechanism documented in this paper.

Further, the authors show that differences in sectoral input intensity and the degree of intersectoral linkages in production are crucial for understanding the differential impact of a given tariff reduction across sectors.

In this paper, we follow in the steps of Caliendo and Parro (2012) in pointing out the empirical relevance of explicitly accounting for multistage production and trade in inputs in studying a world characterized by production segmentation across borders and complex intersectoral production linkages. But instead of quantifying the magnification of the gains from trade following trade liberalization, here we focus on quantifying the contribution of sector-specific differences in the cost of the input bundle to determining the pattern of countries' specialization on world markets. Specifically, by making a simplifying assumption on the input-output structure, we show that the only component which drives a wedge in the cost of the input bundle across countries is the country-specific proximity characteristic.[6] It is this proximity characteristic which co-determines the ranking of relative sectoral exports because the wedge in the cost of inputs matters relatively more in sectors which use inputs more intensively. The empirical analysis we conduct in this paper thus focuses on a different dimension for which trade in inputs and intersectoral production linkages may matter, and is complementary to the analysis conducted by Caliendo and Parro (2012).

The paper is structured as follows. Section 1 outlines the model and derives the measure of proximity to suppliers while section 2 goes over the estimation procedure used in the empirical application. Section 3 gives details on the data we use and on results obtained in the estimation of model parameters. In section 4, we show how to bring the theoretically grounded measure of proximity to the data, discuss countries' ranking according to the proximity characteristic, and report results on the contribution of the proximity mechanism to determining intersectoral specialization. Section 5 conducts a variance decomposition of revealed comparative advantage (RCA) rankings across technology, labor endowments, and proximity to quantify the relative contribution of fundamental country characteristics to determining the pattern of comparative advantage. Section 6 concludes.

1 Stylized model

In substance this paper studies two questions. First, we ask under what circumstances the structure of trade costs combined with a sequential production process may constitute a source of comparative advantage. Second,

[6] The assumption is that the unit cost of the bundle of inputs is the same in all sectors while the cost share of inputs is sector specific.

we ask to what extent trade frictions contribute to determining the pattern of countries' intersectoral specialization on world markets.

To organize ideas, we use a many-good many-country Ricardian model developed by Costinot et al. (2012) and modified in this paper to incorporate sector-specific production features. We allow for a multistage production process and trade in both inputs and final goods. The main purpose of the model is to derive the microfounded proximity characteristic of each country and to show that this characteristic, interacted with sector-specific input intensity, co-determines the pattern of comparative advantage.

1.1 Model set-up

We follow Costinot et al. (2012) in setting up a multisectoral Ricardian economy with differentiated varieties within each sector. There is a finite number of sectors k, and within each sector there is an infinite countable number of differentiated varieties $\alpha \in A \equiv \{1, ..., \infty\}$.

The production function of each variety is Cobb-Douglas in intermediate inputs and labor. By analogy with the seminal Eaton and Kortum (2002) set-up, it is assumed that inputs from all sectors are combined to produce output, with the production function reproducing exactly the features of the expenditure function so that the cost of the input bundle is given by the overall price index (see below).

Define the sectoral production cost component common to all varieties ω_i^k. The input intensity of the sector is captured by exponent ζ^k, with ν_i the wage, P_i the price of the input bundle, and ϵ^k the Cobb-Douglas constant.[7]

$$\omega_i^k \quad = \quad \nu_i^{1-\zeta^k} P_i^{\zeta^k} \epsilon^k$$

The only sector-specific feature incorporated in the production function is the assumption that sectors differ in the way they combine inputs and labor. As the factor share of inputs ζ^k is sector-specific, cross-country differences in input costs matter relatively more in sectors which use inputs more intensively. Similarly, cross-country differences in labor costs matter relatively less in sectors which use inputs more intensively.[8]

Bilateral trade costs are modelled as containing a bilateral symmetric component common across sectors τ_{ij} and an exporter-sector specific component $\tau_i^{E,k}$ common across destination markets.[9] We think of the pair-specific

[7] $\epsilon^k = \left(\zeta^k\right)^{-\zeta^k} \left(1 - \zeta^k\right)^{-(1-\zeta^k)}$.

[8] Notice that the wage mechanism may reinforce or dampen the input cost mechanism depending on the sign of the correlation between relative wages and relative input costs.

[9] Waugh (2010) argues that this specification fits the data better than destination-specific components of trade frictions.

component as measuring trade costs independent of trade policy such as transport, coordination, information, and other transaction costs. We think of the sector-specific trade friction as determined by tariff and non-tariff barriers. It captures the level of trade restrictiveness this exporter faces in getting her products to world markets.[10]

The key feature of the model is the assumption that the number of units of variety α which can be produced with one unit of labor, $z_i^k(\alpha)$, is drawn from the Fréchet distribution with country-sector specific scale parameters z_i^k and a common shape parameter θ:

$$F_i^k(z) \quad = \quad \exp\left[-\left(z/z_i^k\right)^{-\theta}\right]$$

Thus, each country can produce the full set of differentiated varieties within each sector but at different cost. The unit cost $c_{ij}^k(\alpha)$ of producing a variety in source i and getting it to destination j is:

$$c_{ij}^k(\alpha) \quad = \quad \frac{\omega_i^k \tau_{ij}^k}{z_i^k(\alpha)}$$

Consequently, there is perfect competition in production whereby the least cost producer of each variety supplies the market:

$$p_j^k(\alpha) = \min_i \left[c_{ij}^k(\alpha)\right]$$

This production structure appears well adapted to the world economy in the sense that a given product is identified by its brand rather than by the location of its production. For example, Oreo cookies are perceived as differing from any other brand of cookies. At the same time the consumer is generally unaware of differences within the brand 'Oreo cookies' which would stem from being produced in the USA rather than in another country.

The Fréchet distribution is consistent with this market structure because it models the probability of arrival of a production technique which quality beats all previous techniques. And only the distribution of the most productive ideas is of interest for deriving aggregate outcomes in perfect competition.

As shown by Eaton and Kortum (2010), it is sufficient to assume that productivity draws $z_i^k(\alpha)$ are independent and identically distributed across

[10] In theory, the most favored nation principle should impede such differences across exporters, but in practice the complex structure of trade policy, with multiple Preferential Trade Agreements (PTA) at different stages of implementation, and Generalized System of Preferences (GSP) tariffs granted to certain developing economies results in trade barrier variability across partners, in particular in the European Union.

varieties, sectors, and countries to separate out the stochastic component from the fundamental cost component of the sector.

Define this fundamental cost component $\left[c_{ij}^k\right]^{-\theta} = \left[\omega_i^k \tau_{ij}^k / z_i^k\right]^{-\theta}$ where z_i^k is the expected sectoral productivity, up to a country-sector invariant constant.[11] This component governs the price distribution of varieties. Specifically, for a given cost c, the number of techniques providing cost less or equal to c in source i and sector k is distributed Poisson with parameter $\left[c_{ij}^k\right]^{-\theta} c^\theta$. This number is increasing in the fundamental productivity component z_i^k and decreasing in the production cost ω_i^k.

The iid assumption for productivity draws entails that the price of each variety is iid across varieties, sectors, and countries. Similarly, the realization of least cost varieties within each sector across the set of potential suppliers is iid. Working with these distributional assumptions the sectoral market share of supplier i in destination j across the set of potential sources $i' \in I$ is given by

$$\pi_{ij}^k = \frac{\left[c_{ij}^k\right]^{-\theta}}{\sum_{i' \in I} \left[c_{i'j}^k\right]^{-\theta}} \tag{1}$$

where π_{ij}^k is the probability that source i has the lowest price in destination j across the set of varieties in sector k (see App.G for details).

The final building block is the choice of a two-tier functional form for the utility (production) function. The lower tier function is CES. Recalling that $p_j^k(\alpha)$ is the price of the effectively bought variety, the sectoral price index in destination j is:

$$P_j^k = \left\{ \sum_{\alpha \in A} \left[p_j^k(\alpha)\right]^{1-\sigma} \right\}^{1/(1-\sigma)}$$

where σ is the elasticity of substitution across varieties.

The choice of a CES aggregator together with the assumption of perfect competition whereby only the least cost variety survives in the market implies that the sectoral price index is given by the $(1-\sigma)$ moment of the distribution of least cost varieties in j (Lemma 2 in Eaton and Kortum (2010)):

$$P_j^k = E\left[p_j^k(\alpha)^{1-\sigma}\right]^{1/(1-\sigma)} \quad = \quad \Gamma^{1/(1-\sigma)} \left[\Phi_j^k\right]^{-1/\theta} \tag{2}$$

where $\Phi_j^k = \sum_{i' \in I} \left[c_{i'j}^k\right]^{-\theta}$ is the price distribution parameter in the destination and Γ is the Gamma function with the argument $[(\theta + 1 - \sigma)/\theta]$. Consequently, the price index is well defined for $\theta > 1 - \sigma$.

[11] The expected sectoral productivity is $E\left[z_i^k(\alpha)\right] = \Gamma\left(1 - 1/\theta\right) z_i^k$.

The upper tier function is Cobb-Douglas where γ^s is the share of sector s in j's total expenditure.[12] As consumer and producer choices are guided by the same functional form, the share of the sector in final demand and in intermediates' demand is also its share of total expenditure. The overall price index - and cost of the input bundle - is:

$$P_j = \prod_{s=1}^{S} \left[P_j^s\right]^{\gamma^s} \tag{3}$$

As in Costinot et al. (2012) we can work with the expression of bilateral sectoral exports to pin down the determinants of countries' intersectoral specialization. Consider relative sectoral exports for a pair of exporters to some destination:

$$\frac{X_{ij}^k}{X_{i'j}^k} = \left[\frac{\nu_i^{1-\zeta^k} P_i^{\zeta^k} \tau_{ij}^k / z_i^k}{\nu_{i'}^{1-\zeta^k} P_{i'}^{\zeta^k} \tau_{i'j}^k / z_{i'}^k}\right]^{-\theta} \frac{Y_i}{Y_{i'}}$$

In log terms, rescaling by the productivity heterogeneity parameter, and for a specific set of input intensity characteristics ζ^k, the ranking of relative sectoral exports is given by a linear combination of four vectors: relative sectoral technology stocks, relative sectoral wages, relative sectoral input costs, and relative trade restrictiveness in exporters' access to world markets. Indeed, for any two sectors and across destinations, we have:

$$\frac{1}{\theta}\left\{\ln\left[\frac{X_{ij}^k/X_{i'j}^k}{X_{ij}^s/X_{i'j}^s}\right]\right\} = \underbrace{\left\{\ln\frac{z_i^k}{z_{i'}^k} - \ln\frac{z_i^s}{z_{i'}^s}\right\}}_{TFP} + \underbrace{\ln\left[\left(\frac{P_i}{P_{i'}}\right)^{\zeta^s-\zeta^k}\right]}_{INPUTS}$$
$$+\underbrace{\ln\left[\left(\frac{\nu_i}{\nu_{i'}}\right)^{\zeta^k-\zeta^s}\right]}_{WAGES} + \underbrace{\ln\frac{\tau_i^{E,s}/\tau_{i'}^{E,s}}{\tau_i^{E,k}/\tau_{i'}^{E,k}}}_{EXPORT-COSTS}$$

The wedge in the cost of inputs may play a role in co-determining the pattern of revealed comparative advantage. To illustrate, suppose the input bundle is relatively cheap in i. This cost advantage is increasing in relative input intensity $\zeta^k - \zeta^s$, pushing i to specialize in relatively high input intensive sectors. The objective of this paper is to assess empirically how much the

[12]The set of sectors is indexed by S while k refers to a specific sector. This notation is needed to distinguish between the synthetic index of trade frictions common to all sectors and the sector-specific input intensity parameter.

input cost channel contributes to determining the pattern of intersectoral specialization.

Indeed, the way one models sectoral costs determines which cost components enter the theoretically grounded measure of revealed comparative advantage. In particular, assuming away any sector-specific features in the production function as in Costinot et al. (2012) delivers the result that the ranking of sectoral technology stocks fully determines the ranking of relative sectoral exports.[13]

With sector-specific production functions, it is no longer immediate that the pattern of intersectoral specialization is driven by relative technology stocks. This paper disentangles the contribution of domestic technology from the contribution of input costs to check whether increased international fragmentation of production has gone hand in hand with an increasing role of input costs in determining the pattern of countries' specialization on world markets.

There are other ways of introducing inputs in the model, either by using input-output tables as in Levchenko and Zhang (2011) and Caliendo and Parro (2012) or by assuming that each sector sources inputs from itself. Both the cost of the input bundle and the index of trade frictions incurred in sourcing inputs would then be sector-specific. We deliberately choose the set-up which shuts down all sources of differences in inputs' use other than the channel of sectoral input intensity to focus on the mechanism of interest for this paper. Furthermore, if the input cost mechanism is shown to co-determine the pattern of comparative advantage in this restrictive set-up, our results would likely be providing a lower bound on the role of the input cost channel in determining the pattern of intersectoral specialization.

1.2 The input cost component of comparative advantage

In this section we work with the cost of the input bundle to get a handle on the origin of cross-country differences in input costs. We show that the input-cost driven component of comparative advantage is fully determined by the structure of trade costs of the exporter with all of its potential suppliers. We refer to this synthetic index of bilateral trade frictions as the proximity characteristic of the exporter.

Recall that the sectoral market share equation is a probability measure

[13] These authors assume that sector-specific components of trade costs are destination-specific. Consequently, they wash out in relative terms leaving sectoral technology stocks as the only exporter-sector specific cost component.

π_{ij}^k which states the probability that country i is the least cost producer of varieties in sector k for country j:

$$\pi_{ij}^k = \frac{\left[\omega_i^k \tau_{ij} \tau_i^{E,k} / z_i^k\right]^{-\theta}}{\Phi_j^k}$$

Bring the bilateral trade cost component to the left hand side and sum across all suppliers $i' \in I$ including domestic consumption of domestic varieties:

$$\sum_{i'=1}^{I} \tau_{i'j}^{\theta} \pi_{i'j}^k = \frac{\sum_{i'=1}^{I} \left[\omega_{i'}^k \tau_{i'}^{E,k} / z_{i'}^k\right]^{-\theta}}{\Phi_j^k}$$

Define $\overline{\Phi}^k$ the realized least cost distribution of varieties in sector k common across countries. It summarizes the price distribution of world best practice across varieties within sector k, inclusive of exporter-specific barriers linked to trade policy.

$$\overline{\Phi}^k = \sum_{i'=1}^{I} \left[\omega_{i'}^k \tau_{i'}^{E,k} / z_{i'}^k\right]^{-\theta}$$

The country-specific distribution of least cost varieties can be written as a rescaled world distribution of least cost varieties:

$$\Phi_j^k = \left\{\sum_{i'=1}^{I} \tau_{i'j}^{\theta} \pi_{i'j}^k\right\}^{-1} \overline{\Phi}^k$$

Recall that sectoral price indices are given by:

$$P_j^k = \kappa \left[\Phi_j^k\right]^{-1/\theta} \tag{4}$$

where $\kappa = \left[\Gamma\left(\frac{\theta+1-\sigma}{\theta}\right)\right]^{1/(1-\sigma)}$ (see eqn.2).

The sectoral price index has three components:

$$P_j^k = \kappa \left[\overline{\Phi}^k\right]^{-1/\theta} \left\{\sum_{i'=1}^{I} \tau_{i'j}^{\theta} \pi_{i'j}^k\right\}^{1/\theta} \tag{5}$$

The only country-specific component of the sectoral price index is the weighted index of trade frictions which is an indicator of the ease with which country j gets access to the world distribution of least cost varieties in sector k.

The overall price index is a Cobb-Douglas aggregator of sectoral price indices. Plugging (5) in (3), the price index can be written as a product of a country-specific index of trade frictions and of two components common to all countries: the product of sectoral price distribution parameters weighted by the share of each sector in total expenditure and the constant κ.

$$P_j = \underbrace{\left\{\prod_{s=1}^{S}\left[\sum_{i'=1}^{I}\tau_{i'j}^{\theta}\pi_{i'j}^{s}\right]^{\gamma^s/\theta}\right\}}_{SPECIFIC}\underbrace{\left\{\prod_{s=1}^{S}\left[\overline{\Phi}^{s}\right]^{-\gamma^s/\theta}\right\}\kappa}_{COMMON} \tag{6}$$

Recall that for $i = j$ the price index in j gives the cost of the input bundle in i. In other words the composite index of sectoral trade frictions $\left\{\prod_{s=1}^{S}\left[\sum_{i'=1}^{I}\tau_{i'j}^{\theta}\pi_{i'j}^{s}\right]^{\gamma^s/\theta}\right\}$ for $i = j$ captures how difficult it is for exporter i to get access to the best world technology in sourcing inputs. Switching sides consider production costs in country i. For $i = j$, the reciprocal of the composite index of trade frictions in j corresponds to the proximity characteristic of exporter i:

$$\overline{PROX}_i^M = \left\{\prod_{s=1}^{S}\left[\sum_{i'=1}^{I}\pi_{i'j}^{s}\tau_{i'j}^{\theta}\right]^{\gamma^s/\theta}\right\}^{-1} \tag{7}$$

The microfounded proximity indicator is a weighted l^{θ}-norm of the vector of bilateral trade frictions, with weights given by the probability that each supplier is least cost across the spectrum of sectoral varieties. It is then aggregated across sectors with exponents given by sectoral expenditure shares.

Plugging (6) in the expression of exporter-sector specific production cost ω_i^k gives:

$$\omega_i^k = \underbrace{\epsilon^k\kappa^{\zeta^k}\left\{\prod_{s=1}^{S}\left[\overline{\Phi}^{s}\right]^{-\gamma^s/\theta}\right\}^{\zeta^k}}_{COMMON}\underbrace{\left\{\nu_i\right\}^{1-\zeta^k}\left\{\prod_{s=1}^{S}\left[\sum_{i'=1}^{I}\tau_{i'i}^{\theta}\pi_{i'i}^{s}\right]^{\gamma^s/\theta}\right\}^{\zeta^k}}_{SPECIFIC} \tag{8}$$

Plugging this expression in the equation of relative sectoral exports for a pair of exporters we observe that the only component of the cost of the input bundle which contributes to determining the pattern of comparative advantage is given by exporters' relative proximity to world technology. This indicator is sector-specific whenever sectors differ in the cost share of inputs.

$$\frac{1}{\theta}\left\{\ln\left[\frac{X_{ij}^k/X_{i'j}^k}{X_{ij}^s/X_{i'j}^s}\right]\right\} = \underbrace{\left\{\ln\frac{z_i^k}{z_{i'}^k} - \ln\frac{z_i^s}{z_{i'}^s}\right\}}_{TFP} + \underbrace{\ln\left[\left(\frac{PROX_i^M}{PROX_{i'}^M}\right)^{\zeta^k-\zeta^s}\right]}_{INPUTS}$$
$$+\underbrace{\ln\left[\left(\frac{\nu_i}{\nu_{i'}}\right)^{\zeta^k-\zeta^s}\right]}_{WAGES} + \underbrace{\ln\frac{\tau_i^{E,s}/\tau_{i'}^{E,s}}{\tau_i^{E,k}/\tau_{i'}^{E,k}}}_{EXPORT-COSTS} \tag{9}$$

Notice that for any two countries, relative proximity is also a summary statistic of the relative cost of living. The intuition is straightforward: the closer the country is to world's best practice, and the lower is its cost of living relatively to other countries. Consequently, relative real wages can be computed by adjusting the ratio of nominal wages by relative proximity without constructing actual price indices.

Proximity is clearly an endogenous object. Even if bilateral components of trade frictions may be considered exogenous in that they are determined by slow-moving characteristics of the trade network (infrastructure, costs in the transport sector, coordination costs,...), market shares are contingent on a specific trade equilibrium. We address this issue in the empirical analysis.

2 Estimation procedure

In the empirical analysis we investigate whether the cost advantage conferred by the ability to source inputs at relatively lower cost leads to specialization of high proximity countries in sectors which use inputs relatively more intensively.

The estimation procedure is based on the gravity structure of trade in the formulation of the EK model at the sectoral level. The unit of analysis is sectoral bilateral trade:

$$X_{ij}^k = \frac{\left[\omega_i^k \tau_{ij}^k / z_i^k\right]^{-\theta}}{\Phi_j^k} X_j^k \tag{10}$$

where $X_j^k = \gamma^k Y_j$ is expenditure on sector k in country j.

We isolate exporter-sector components of production costs to identify the contribution of the input cost channel to shaping the pattern of specialization separately from other fundamental country characteristics.

The estimation procedure consists of three steps. First, we work in cross section, with $t \in T = \{1995, ..., 2009\}$. Bilateral sectoral exports are regressed on pair, destination-sector, and exporter-sector fixed effects to isolate exporter-sector specific components of production costs relatively to a benchmark country (the US) and industry (processed foods and beverages).

$$X_{ij,t}^k = \exp\left\{fe_{ij,t} + fe_{j,t}^k + fe_{i,t}^k + \xi_{ij,t}^k\right\} \tag{11}$$

where $fe_{ij,t}$, $fe_{j,t}^k$, $fe_{i,t}^k$ are respectively pair, destination-sector, and exporter-sector fixed effects, and $\xi_{ij,t}^k$ is the error term.

This step is identical to the estimation conducted in Costinot et al. (2012) to retrieve relative sectoral productivities. Indeed, under the assumption that the only exporter-sector cost component is the expected productivity, this regression would retrieve sectoral technology stocks $z_{i,t}^k$ relatively to a benchmark country and industry.[14]

However, if sectoral trade costs contained an exporter-sector specific component $\tau_{i,t}^{k,E}$, this approach would retrieve relative fundamental sectoral productivity scaled by export-side trade costs $z_{i,t}^k/\tau_{i,t}^{k,E}$. And if labor and inputs were combined differently across sectors in the production process, exporter-sector dummies would capture sectoral components of wages and input costs contained in $\omega_{i,t}^k$.

This paper puts forward the hypothesis that with this first step we may be picking up the comprehensive exporter-sector unit cost of production.[15] The exporter-sector dummy corresponds to a combination of technology, wage, input, and trade cost components, relatively to the benchmark country and industry for which sectoral production costs had been normalized to one.[16]

$$\widehat{fe}_{i,t}^k = \theta\ln(z_{i,t}^k) - \theta(1-\zeta^k)\ln\nu_{i,t}^k - \theta\zeta^k\ln(P_{i,t}) - \theta\ln(\tau_{i,t}^{E,k}) + \xi_{i,t}^k \tag{12}$$

A complementary objective of the paper is to measure sectoral technology stocks in the data, and to assess to what extent the pattern of specialization is effectively driven by the ranking of relative sectoral technology stocks. In particular, if other exporter-sector cost components are largely determined by domestic technology, the specification of a sector-specific production function would be redundant.

[14]The exponentiated source-sector dummy $e^{fe_{i,t}^k}$ raised to the exponent $1/\theta$ would correspond to $\left[\frac{z_{i,t}^k}{z_{i,t}^s}/\frac{z_{i',t}^k}{z_{i',t}^s}\right]$ where $z_{i,t}^s$, $z_{i',t}^k$, and $z_{i',t}^s$ had been normalized to 1.

[15]App.C reports descriptive statistics on estimated relative production costs for countries of our sample, both in cross-section and overtime.

[16]The residual $\xi_{i,t}^k$ picks up the error component due to sampling uncertainty of the estimated dummy.

Consequently, in the second step we pool data on estimated exporter-sector dummies $\widehat{fe}_{i,t}^{k}$ and regress them on instrumented sectoral wages $(\widehat{\nu}_{i,t}^{k})$ and instrumented sectoral technology stocks $(\widehat{z}_{i,t}^{k})$, controlling for the benchmark country component with year fixed effects fe_t:

$$\widehat{fe}_{i,t}^{k} \;=\; \theta \left[\ln \widehat{z}_{i,t}^{k} - (1-\zeta^{k}) \ln \widehat{\nu}_{i,t}^{k}\right] + fe_t + \lambda_{it}^{k} \tag{13}$$

Sec.3 reports details on the data used to compute technology stocks and wages and discusses the instrumenting procedure. Here we simply acknowledge that wages have an intrinsic sectoral component in the data and henceforth index wages $\nu_{i,t}^{k}$.

The residual of the second step equation contains the $\xi_{i,t}^{k}$ error component due to sampling uncertainty of the exporter-sector dummy. However, using an estimated regresor as dependent variable will not lead to inconsistency of parameter estimates as long as the dummy is consistently estimated in the first step. Consequently, the second step equation is estimated in OLS with heteroskedasticity-consistent standard errors.[17]

This specification allows retrieving structural parameters θ and ζ^{k} in a way consistent with the underlying model: the heterogeneity parameter of the productivity distribution is assumed constant across sectors and overtime, while input intensity characteristics are assumed sector-specific, but common across countries and overtime. We check that model parameters are precisely estimated, stable across variants of the instrumenting procedure, and consistent with previous studies. Further, we check that estimated ζ^{k} parameters are strongly positively correlated with observed input intensity in our dataset.

The residual of the second step equation isolates all channels through which trade costs may play a role in the pattern of intersectoral specialization. On the exports side, the residual contains the trade policy determined cost component $\tau_{i,t}^{E,k}$ which captures how costly it is for the country to ship products to world markets relatively to other exporters. On the imports side, the residual contains a vector of $i' \in I$ pair-specific bilateral trade cost components $\tau_{i'i}$ which enter in the expression of the sectoral price index and which capture in a complex way how costly it is for country i to get access to inputs produced in all sources i' including itself.

The estimated residual $\widehat{\lambda}_{i,t}^{k}$ also contains an exporter-year specific component $fe_{i,t}$ which corresponds to the production cost in the benchmark sector. Furthermore, it contains a residual component $\eta_{i,t}^{k}$ which captures the error component due to sampling uncertainty of the dependent variable as well

[17] This is the approach advised in Hausman (2001) and Lewis and Linzer (2005) for cases in which sampling uncertainty is likely to be small.

as production costs which were not picked up by instrumented technology and wages. The latter are assumed statistically independent from export or import side trade costs.[18]

$$e^{\widehat{\lambda}_{i,t}^{k}} = \left\{ \prod_{s=1}^{S} \left[\sum_{n=1}^{N} \tau_{nj}^{\theta} \pi_{nj}^{s} \right]^{\gamma^s/\theta} \right\}^{-\theta\zeta^k} \left[\tau_{i,t}^{E,k} \right]^{-\theta} e^{\theta f e_{i,t}} e^{\theta \eta^k{}_{i,t}} \qquad (14)$$

The residual is thus a combination of two sector-specific trade cost characteristics of the country: the barriers it overcomes in getting access to world best practice on the supply side, and the trade cost it pays to get domestically produced varieties to world markets. The crucial feature of this residual for our identification strategy is that the impact of export side trade costs is not magnified in input intensive sectors. As in the seminal EK model, a given gap in relative export side trade costs has a proportional effect on relative exports in all sectors, and does not constitute a source of comparative advantage. On the other hand, the impact of import side trade costs is magnified in input intensive sectors. Consequently, if sectoral input intensity interacted with an index of trade costs determines the ranking of residual relative sectoral exports, it unambiguously identifies the role of the input cost channel in codetermining the pattern of comparative advantage.

The role of this mechanism in determining the pattern of specialization is spelled out by Limão and Venables (2002) in a three sector economy in which one sector produces intermediates and the two remaining sectors produce final goods with different input intensity. The authors rank final good sectors in terms of 'transport intensity' which in our model corresponds to their ranking in terms of input intensity. These authors show that under the assumption of a uniform distribution of factor endowments across locations, the input cost channel determines the pattern of specialization, whereby locations close to input suppliers specialize in transport intensive goods.

Limão and Venables (2002) fix the location specialized in inputs' production to pin down the contributing role of the input cost channel under different capital-labor allocations. In particular, they show that specific distributions of factor endowments across locations overturn the pattern of specialization defined by the input cost channel. We use a more flexible model to check for the presence of the input cost channel in the data and to quantify its importance relatively to the two other fundamental characteristics which are factor endowments and technology stocks.

[18]In practice, we relax this assumption because we instrument the proximity indicator.

Relative bilateral sectoral exports to market j for exporters i and i' are:

$$\ln\left\{X^k_{ij,t}/X^k_{i'j,t}\right\} \quad = \quad \theta\left[\ln\frac{z^k_{i,t}}{z^k_{i',t}} - (1-\zeta^k)\ln\frac{\nu^k_{i,t}}{\nu^k_{i',t}} - \ln\frac{\tau_{ij}\tau^{E,k}_{i,t}}{\tau_{i'j}\tau^{E,k}_{i',t}} + \zeta^k\ln\left\{\frac{\overline{PROX}^M_{i,t}}{\overline{PROX}^M_{i',t}}\right\}\right]$$

Exporter-sector dummies $fe^k_{i,t}$ obtained in the first step capture the sectoral cost component of the exporter relatively to a benchmark sector and country for which cost components are normalized to one. Thus, relative exporter-sector dummies for any pair of exporters capture relative sectoral cost components for this pair of exporters, up to an exporter-year component which corresponds to the production cost of each exporter in the benchmark sector. We control for it in the estimation with exporter-year fixed effects $fe_{n,t}$ for $n = i, i'$.

$$fe^k_{i,t} - fe^k_{i',t} = \theta\left[\ln\frac{z^k_{i,t}}{z^k_{i',t}} - (1-\zeta^k)\ln\frac{\nu^k_{i,t}}{\nu^k_{i',t}} - \ln\frac{\tau^{E,k}_{i,t}}{\tau^{E,k}_{i',t}} + \zeta^k\ln\frac{\overline{PROX}^M_{i,t}}{\overline{PROX}^M_{i',t}}\right]$$
$$+\theta fe_{i,t} - \theta fe_{i',t} + \xi^k_{ii',t} \qquad (15)$$

In the empirical analysis, we use data on sectoral exports to each of EU-15 markets. According to our modelling of trade costs, the export side component $\tau^{E,k}_{i,t}$ contains tariff and non tariff barriers which are exporter-sector specific and common across destination markets. This hypothesis adequately describes the underlying trade cost structure in the EU-15 because the EU is characterized by a unique external trade policy and a multiplicity of exporter-specific trade agreements, in particular with emerging economies. Thus, bilateral components capture plausibly symmetric barriers to trade such as transport costs while exporter-specific components capture source-specific trade restrictiveness linked to trade policy.

As noted above, the impact of export-side trade costs is not magnified in input-intensive sectors. Consequently, they could impede identification only if it were the case that in 1995-2009 relatively high-proximity countries faced systematically lower trade policy barriers in input-intensive sectors. We posit that this is not the case.[19]

More formally, we assume that sectoral trade costs $\tau^{E,k}_{i,t}$ are well approximated for each exporter by the component $\tau^{E}_{i,t}$ which is time-varying and common across sectors. In relative terms, this common component is a summary statistic of the relative trade restrictiveness faced by a pair of exporters on destination markets. Notice that any synthetic index of trade costs on the exports' side, such as 'proximity to clients', would be absorbed by this

[19] We will check this assertion using actual tariffs' data in future work.

common component and picked up by pair fixed effects in the first step of the estimation. As there is no magnification mechanism on the exports' side linked to input intensity, relatively high proximity to clients has no incidence on the pattern of intersectoral specialization.

Assume that this relative common component is multiplied by a stochastic component $\iota^k_{i,t}/\iota^k_{i',t}$, distributed lognormal with mean 1. If this stochastic component is statistically independent of the regressors in (15) we can rewrite relative pairwise RCA rankings as a function of three complementary components: relative technology stocks, relative sectoral wages, and relative proximity. These three components fully account for the microfounded measure of revealed comparative advantage for a given pair of exporters on world markets, up to a relative exporter-year fixed effect and a stochastic component captured by the residual $\xi^k_{ii',t}$ which comprises the stochastic component $\ln(\iota^k_{i,t}/\iota^k_{i',t})$.

$$fe^k_{i,t} - fe^k_{i',t} = \theta \left[\ln \frac{z^k_{i,t}}{z^k_{i',t}} - (1-\zeta^k) \ln \frac{\nu^k_{i,t}}{\nu^k_{i',t}} + \zeta^k \ln \frac{\overline{PROX}^M_{i,t}}{\overline{PROX}^M_{i',t}} \right] + \theta fe_{i,t} - \theta fe_{i',t} + \xi^k_{ii',t} \tag{16}$$

The residual component of RCA rankings illustrates that conditional on the distribution of technology and wages, intersectoral specialization within a pair is determined by the relative proximity characteristic interacted with the input intensity of the sector. The proximity index is a summary statistic of the input component of the cost advantage conferred to the country by the ease of its access to best technology worldwide in sourcing inputs. It becomes a source of comparative advantage at the intersectoral level because the input component of production costs matters relatively more in sectors which use inputs intensively.

Consequently, in the third step of the estimation procedure, we use the proximity ranking of countries to test for the role of relative input costs in determining the pattern of comparative advantage. We split the sample in two groups according to the proximity characteristic of the country (details on the proximity ranking are provided in sec.4). We rescale estimated residuals $\widehat{\lambda}^k_{i,t}$ by the estimated heterogeneity parameter $\widehat{\theta}$, and compute all pairwise combinations of sectoral annual residuals $(1/\widehat{\theta})(\widehat{\lambda}^k_{i,t} - \widehat{\lambda}^k_{i',t})$ where $i \in H$ are countries of the high proximity group, and $i' \in L$ are countries of the low proximity group.

The indicator of sectoral relative proximity is computed as the log of the relative proximity characteristic for each pair, and is instrumented with an indicator of relative proximity endowment.[20] Instrumented relative proximity is then interacted with the estimated input intensity characteristic of the sector $\widehat{\zeta}^k$: $\widehat{\zeta}^k \ln(\widehat{PROX}^M_{i,t}/\widehat{PROX}^M_{i',t})$.

$$\frac{1}{\widehat{\theta}}\left[\widehat{\lambda}^k_{i,t} - \widehat{\lambda}^k_{i',t}\right] = \beta_0 + \beta_1 \ln\left\{\left(\frac{\widehat{PROX}^M_{i,t}}{\widehat{PROX}^M_{i',t}}\right)^{\widehat{\zeta}^k}\right\} + fe_{i,t} - fe_{i',t} + \eta^k_{ii',t} \tag{17}$$

We estimate (17) on data pooled for all years. As in the second step of the estimation, we report heteroskedasticity-robust standard errors to take into account sampling uncertainty of the dependent variable.[21] Exporter-year fixed effects are included to control for characteristics of the benchmark sector for each exporter and year. The coefficient of interest is β_1: according to the model, β_1 should be positive and close to 1.

3 Data and Estimation of Model Parameters

3.1 The Data

3.1.1 Exporter-sector dummies

To obtain the ranking of relative sectoral exports on EU-15 markets (step 1 of the estimation), we use the COMEXT database. COMEXT provides exhaustive information on bilateral trade flows for each country of the EU-15 with each other country in the world at the 8-digit level (CN classification). We use data on total imports to identify the set of EU-15 main trading partners in 1995-2010, defined as the set of countries which make up at least 1% of total EU-15 imports in more than one year in the period under study.[22]

As the model is silent about countries' endowments of primary goods, we restrict attention to categories classified as manufacturing. We use the CN8-BEC correspondence to drop inputs produced from raw gas, petroleum, coal, and nuclear fuel. We construct a correspondence from the CN8 to the 4- and 2-digit NACE 1.1 and ISIC Rev.3 classifications where manufacturing

[20] We discuss the computation of proximity indicators and the motivation for instrumenting the microfounded proximity indicator in sec.4.

[21] Using bootstrap to compute standard errors leads to a relatively small improvement in efficiency.These results are available upon request.

[22] See tab.10. In practice, we include all members of the European Union, excluding Cyprus and Malta but including Croatia.

corresponds to sectors $15 - 36$ at the 2-digit level.[23] In the estimation, we exclude energy products (sector 23) to be consistent with dropping energy inputs, and tobacco products (sector 16) for which data is patchy. This leaves 20 sectors at the 2-digit level (see tab.13).

App.C provides descriptive statistics on exporter-sector dummies estimated at the 2-digit level in 1995-2010. It underlines the persistence in country-specific relative sectoral rankings. It discusses changes in the pattern of revealed comparative advantage at the bilateral level and by partner type. In Costinot et al. (2012), these rankings would correspond to the ranking of fundamental sectoral productivities while in this paper the ranking results from technology, factor, input, and export-side trade cost components specific to the sector and exporter.

3.1.2 TFP and wages

To estimate the parameters of the model (step 2 of the estimation), we need information on technology stocks and labor costs. We construct these components using the World Input Output Database (see Timmer (2012)) which provides harmonized information on gross output, workforce, hourly wages, expenditure on inputs and labor, nominal investment and real capital stocks by sector for all but six countries of our sample.[24]

Sectoral total factor productivity is constructed by fitting a Cobb-Douglas production function while allowing factor shares to vary by country and sector.[25] In logs, TFP is given by the residual of real gross sectoral output Y_i^k from which we subtract the contribution of three production factors which are inputs I, labor H, and capital K, weighted by their respective income shares $\beta_{f,i}^k$, with $f = \{I, H, K\}$:

$$\ln(\overline{z}_i^k) \quad = \quad \ln Y_i^k - \beta_{I,i}^k \ln I_i^k - \beta_{H,i}^k \ln H_i^k - \beta_{K,i}^k \ln K_i^k \qquad (18)$$

Real gross sectoral output and real expenditure on inputs are obtained by deflating the corresponding nominal values by output and input deflators

[23]There are 121 active 4-digit codes in ISIC Rev.3, and a bit more in NACE 1.1. There are minor discrepancies between NACE 1.1 and ISIC Rev.3 at the 4-digit level, mainly because NACE 1.1 is a more detailed classification. There are no discrepancies at 2-digit in the sense that CN8 products are classified within the same 2-digit category in both classifications.

[24]Croatia, Norway, Switzerland, Malaysia, Singapore, and Thailand are the six countries absent from the WIOD. We have not been able to find an alternative data source for these countries.

[25]Sectoral factor shares are not constrained to be common across countries because we seek to construct sectoral TFP using the full set of information in the data.

provided in the WIOD. Labor use is taken directly from WIOD as the total number of hours worked in the sector. Obtaining real capital expenditure is more tricky. WIOD provides information on nominal capital expenditure, nominal investment, and a deflator for nominal investment. We approximate the use of capital in the production process by predicting nominal capital expenditure by deflated sectoral investment.[26] Income shares of production factors are computed as ratios of nominal expenditure on inputs, labor costs, and capital to the nominal value of sectoral output.

We use two measures of sectoral wages. The first is the hourly wage in the sector ν_i^k obtained by dividing total labor compensation by total hours worked. The second measure is obtained by adjusting observed labor costs for human capital accumulated through education.

The adjustment consists in rescaling hourly skill-specific wages by a proxy of worker efficiency as in Eaton and Kortum (2002). Skill-specific wages $\nu_{edu,i}^k$ are given by the average hourly wage rescaled by the ratio of the cost share of the skill in total costs $\omega_{edu,i}^k$ to the time share of the skill in total hours worked $\overline{\omega}_{edu,i}^k$:

$$\nu_{edu,i}^k = \frac{\omega_{edu,i}^k}{\overline{\omega}_{edu,i}^k}\nu_i^k$$

The efficiency adjustement is implemented by multiplying skill-specific wages by an exponential function which argument is the average number of years of schooling for the skill S_{edu} multiplied by the return to education $g = .06$.[27]

$$\overline{\nu}_{edu,i}^k = \nu_{edu,i}^k e^{-gS_{edu}}$$

In WIOD, hourly wages are reported for low (l), medium (m), and high (h) skilled workers. We use the International Standard Classification of Education correspondence of skills to educational attainment to define $S = \{8, 13, 18\}$ for $edu = \{l, m, h\}$, respectively.[28]

The sectoral efficiency-adjusted hourly wage corresponds to a weighted average of skill-specific efficiency-adjusted wages. In practice, we adjust the number of hours worked for each skill, and compute the efficiency-adjusted wage $\overline{\nu}_i^k$ as the ratio of expenditure on labor to the efficiency adjusted number of hours.

[26] Real GFCF explains 67% of variation in nominal capital expenditure.

[27] We follow Eaton and Kortum (2002) in using 6% as the return to education. This estimate is reported as conservative in Bils and Klenow (2000).

[28] UNESCO, ISCED 1997, reedited 2006.

3.1.3 Level of aggregation

WIOD reports data for 13 manufacturing sectors instead of the 20 sectors obtained at the 2-digit level (compare tab.10 and tab.13). Since we only have data on measured TFP and hourly wages for 13 sectors, we reestimate exporter-sector dummies at this level of aggregation, and work with 13 manufacturing sectors in the second and third steps of the estimation.

In four cases, this higher level of aggregation corresponds to pooling data on production of processed inputs and final output for a specific industry. This is the case in the textile (sectors 17 and 18), paper (sectors 21 and 22), metal products (sectors 27 and 28), and transport industries (sectors 34 and 35). For these four industries the higher level of aggregation may actually improve consistency with the production structure considered in the model.

For two industries, this higher level of aggregation introduces a discrepancy between trade and production data. Thus, the WIOD pools together food manufacturing with the tobacco industry while the latter is dropped from trade data because of mediocre data quality. The second discrepancy is due to aggregation of miscellaneous manufacturing (36) with the recycling industry (37), the latter being absent from trade data. We assume this discrepancy to be relatively minor because the common component is likely to be representative of gross sectoral output.

The problematic aspect of data aggregation in WIOD is the pooling of data on computer manufacturing, electrical and audiovisual equipment, and medical-optical precision equipment (sectors $30-33$) into a single industry. Tab.13 shows that these four sectors vary significantly in the share of value added (VA) in gross output.[29] The precision equipment industry has the highest VA share in manufacturing (.41) while the computer and office machinery is relatively input intensive with the share of VA at just over .25 of gross output.[30]

The level of aggregation may impact the ranking of sectors in terms of input intensity. It may also play against our assumption that inputs' share in gross output is a sector-specific characteristic common across countries. Indeed, even if the underlying production functions have common factor shares at a relatively fine level of disaggregation, the sectoral mix of subsectors' input intensity is likely to be country specific. In particular, measured sec-

[29] The table reports input intensity in manufacturing at the 2-digit level for the main economies of the EU-15. The indicator is computed as $1 - VA/PROD$ using UNIDO INDSTAT.

[30] This example illustrates that the intensity of inputs' use in production is not equivalent to the ranking of sectors according to technological complexity. Similarly, tab.12 shows that there is no one-to-one mapping from the share of inputs in production to the share of inputs in total imports which we refer to as sectoral input intensity in trade.

toral input intensity at the WIOD level becomes an endogenous object if high(low) proximity countries tend to specialize in high(low) proximity subsectors within each industry. If this is the case, the pattern of intersectoral specialization is relatively less determined by the proximity mechanism. Consequently, working at a higher aggregation level is likely to make it more difficult to pick up the working of the proximity mechanism at the intersectoral level.

In App.B, we use production and value added at the 4-digit level in ISIC Rev.3 reported in UNIDO INDSTAT4 to gauge the sensitivity of measured input intensities to the level of data aggregation. For the main economies of the EU-15, the assumption of common factor shares is best borne out in the data at the 4-digit level. But if we focus on the ordinal ranking of sectors as a function of inputs' weight in the production function, this ranking is found to be relatively stable across countries even at higher aggregation levels. Similarly, in the WIOD, the ranking of measured sectoral input intensities is strongly correlated across countries. Consequently, the assumption of common sectoral input intensities used in the estimation of the model is consistent with the data in as much as it captures an ordinal ranking of sectors as a function of input intensity.

3.2 Estimation of the model

3.2.1 Motivation of an instrumental variables estimator

The second step of the estimation is the crucial point of the analysis in which we obtain the structural parameters of the model and construct the residual component of exporter-sector specific production costs. This residual is used to evaluate the role of the proximity mechanism in co-determining the pattern of comparative advantage in the final step of the estimation.

The first reason for implementing an instrumental variables estimator is to ensure that the vector of estimated residuals is orthogonal to variation in measured TFP and hourly wages attributable to domestic technology and labor endowments. Indeed, as shown in (16), we need to clean the exporter-sector dummy of these two fundamental country-sector characteristics to isolate the input cost channel.

The second reason is the need to obtain consistent estimates of model parameters which may be hindered by errors-in-variables in measured TFP and hourly wages. Furthermore, joint determination of sectoral exports with non-instrumented TFP and wages cannot be excluded. Isolating the variation in measured TFP and hourly wages determined by fundamental country characteristics helps stymie both sources of bias.

In principle, model parameters could be estimated directly in the equation of bilateral sectoral exports. We do not do this for two reasons. First, we want a single set of parameters for the full set of years. Second, we want to focus on the explanatory power of our instruments on the exporter-sector component of bilateral exports. Consequently, we would like to avoid predicting sectoral TFP and wages with the full set of pair and destination-sector dummies included in the first step of the estimation.[31]

The estimation is conducted in two stages. In the first stage sectoral TFP and wages are regressed on a common set of instruments to identify the variation in measured TFP and hourly wages explained by domestic technology and labor endowments. Both characteristics are sufficiently slow-moving to be considered exogenous to a given trade equilibrium.

In the second stage we project estimated exporter-sector dummies on the space defined by the vectors obtained in the first stage which are instrumented sectoral wages ($\widehat{\nu}^k_{i,t}$) and instrumented sectoral TFP ($\widehat{z}^k_{i,t}$) while allowing the coefficient on wages to be sector-specific. This is done to identify the component of RCA rankings which is orthogonal to variation in TFP and sectoral wages picked up by technology stocks and labor endowments:

$$\widehat{fe}^k_{i,t} \;=\; \theta \ln \widehat{z}^k_{i,t} - \theta(1-\zeta^k)\ln \widehat{\nu}^k_{i,t} + fe_t + \lambda^k_{it} \tag{19}$$

Consequently, it is the residual of the reduced form model specified in (19) which should be used in the third step of the estimation to identify the contribution of the proximity mechanism to determining the component of RCA rankings unexplained by technology stocks and labor endowments.

The productivity dispersion parameter θ is directly given by the coefficient on instrumented TFP, while sector-specific input intensities ζ^k are computed from the coefficient on instrumented sectoral wages using estimated θ.[32]

A potential caveat of this procedure is the omission of capital in the estimation of structural parameters of the production function. Conceptually we could modify the model to allow for sector-specific capital shares as we do in computing sectoral TFP. We would then need to include an empirical

[31] We stay away from the alternative which consists in estimating model parameters in the first step while also including the proximity index. This alternative requires a constraint on estimated coefficients on wages and proximity. We avoid relying on the proximity index in estimating the structural parameters of the model because we do not have data on actual trade frictions.

[32] In the data we observe empirical counterparts of structural input intensities as the country-year specific sectoral income shares of inputs in gross output $\beta^{k,I}_{i,t}$. Our choice of estimating a single sectoral input intensity for the full set of years simultaneously with the dispersion parameter achieves greater internal consistency than computing some type of weighted average across observed sectoral parameters.

counterpart to the cost of capital in the second step regression.[33] Empirically, the problem stems from the negative correlation between input and capital intensity in the data. By omitting capital we introduce a bias in the estimated input intensity. But since this bias can only blur the proximity mechanism in the sense that we underestimate input intensity in input intensive sectors and overestimate it in labor intensive sectors, our results are likely to provide a lower bound on the underlying role of the proximity mechanism.[34]

3.2.2 Which instruments?

Sectoral workforce L_i^k constitutes a logical instrument for hourly wages because sectoral wages are decreasing in labor endowment.[35] The information on the number of persons engaged in the sector is directly provided in WIOD.

Efficiency adjusted wages $\overline{\nu}_i^k$ are instrumented with efficiency-adjusted sectoral workforce $\overline{L}_i^k$. We compute the number of persons engaged by skill $L_{edu,i}^k$, and adjust skill-specific labor by the human capital of the worker:

$$\overline{L}_{edu,i}^k = L_{edu,i}^k e^{gS_{edu}}$$

The adjusted labor force is the sum of efficiency-adjusted labor by skill.

Sectoral technology stocks are modelled as a function of capital accumulation and R&D activity.[36] Accordingly, we use two sets of instruments for measured TFP. In the first specification, sectoral TFP is instrumented with real sectoral capital stocks and R&D personnel. Data on real capital stocks is provided by the WIOD in 1995-2007. Data on the full time equivalent number of persons employed to conduct R&D activity is reported in ANBERD (see below). The caveat is the restriction of the estimation window to 1995-2007. The advantage is the ability to implement standard tests on instrument validity given that the equation is overidentified.

In the second specification we use nominal R&D expenditure as the indicator of R&D activity. We consider that R&D expenses are mostly incurred to finance investment and employment of R&D personnel. Consequently, we first deflate sectoral expenditure on R&D by regressing it on real investment

[33] One possibility is to infer the return to capital from market-clearing conditions as in Levchenko and Zhang (2011).

[34] In future work I will check results' robustness by using a firm-level dataset such as Orbis to estimate the parameters of the production function. I thank Maggie Chen for this suggestion.

[35] Labor endowments by skill in each sector are considered predetermined by making the hypothesis that sector-specific human capital impedes labor movement across sectors. The sector-specific mix of skills is taken as given.

[36] See Eaton and Kortum (1999, 2002).

and R&D personnel.[37] Measured TFP is instrumented with predicted R&D expenditure. In this specification the estimation window is extended to 2009 because real investment data is reported in WIOD in 1995-2009.

The bottleneck is the availability of data on R&D activity (see App.D). Time series data on R&D personnel and nominal R&D expenditure for all developed and a subset of emerging economies are taken from the 2011 edition of OECD ANBERD.[38] For China, we compiled sectoral data on R&D personnel and nominal R&D expenditure in 1995-2009 using the Yearbook Database of China Data Online.[39] Bulgaria, Brazil, India, Indonesia, Lithuania, Latvia, and Russia were dropped because of lacking data on R&D expenditure and personnel.[40] This leaves 26 countries in the second step of the estimation.

3.2.3 Estimated parameters

To estimate this model we need instrumented sectoral hourly wages and instrumented TFP. Consequently, in the first stage we run 13 regressions in which measured TFP and hourly sectoral wages are regressed on a common set of instruments which include R&D personnel and real capital stocks in (I) and (II) (deflated R&D expenditure in (III) and (IV)) together with the workforce of each of the 12 sectors. In (I) and (III) sectoral workforce is efficiency-adjusted. In (II) and (IV) we use raw data on hourly wages and number of persons engaged in the sector.

Tab.1 reports results of the second stage while results of the first stage are reported in App.D.2. Reported values of Kleibergen-Paap rk LM and Cragg Donald Wald F statistics attest that instruments pass respectively the underidentification and weak identification tests across specifications.[41] As the equation is overidentified in the first two specifications, we report the result of the test of overidentifying restrictions (Hansen J statistic). The joint null that instruments are uncorrelated with the error term and correctly excluded from the estimation is not rejected at conventional significance levels.

The parameters of the model are precisely estimated across the four specifications. The range of point estimates for the heterogeneity parameter is

[37]The estimated coefficient on R&D personnel is .92(.009), and .23(.01) on real investment. The two variables explain 87% of observed variation in nominal R&D expenditure.

[38]Downloaded in July 2012 from OECD ANBERD.

[39]See Yearbook Database. The data is reported in html and pdf formats in China Statistical Yearbook on Science and Technology (1996-2008), China Statistics Yearbook on High Technology Industry (2002, 2003, 2007), and in the chapter 'Education, Science, and Technology' of China Statistical Yearbook (2007-2011).

[40]Only data on nominal R&D expenditure is available for Russia in ANBERD.

[41]The underidentification test rejects the null that the matrix of reduced form coefficients is not full rank.

Table 1: Second stage: Estimated parameters

	(I)	(II)	(III)	(IV)
TFP	7.258***	6.718***	7.842***	7.280***
	(0.506)	(0.431)	(0.524)	(0.448)
$WAGE$	-1.343***	-1.388***	-1.610***	-1.583***
	(0.212)	(0.145)	(0.211)	(0.149)
WAGE 19	1.090***	0.558***	1.226***	0.640***
	(0.292)	(0.138)	(0.274)	(0.131)
WAGE 20	-1.265***	-0.793***	-1.136***	-0.727***
	(0.178)	(0.101)	(0.163)	(0.095)
WAGE $21-22$	-1.471***	-0.959***	-1.365***	-0.910***
	(0.156)	(0.091)	(0.143)	(0.085)
WAGE 24	-0.522***	-0.354***	-0.339**	-0.250***
	(0.158)	(0.092)	(0.153)	(0.091)
WAGE 25	-0.520***	-0.332***	-0.410***	-0.274***
	(0.154)	(0.089)	(0.144)	(0.085)
WAGE 26	-0.840***	-0.527***	-0.767***	-0.498***
	(0.142)	(0.083)	(0.131)	(0.078)
WAGE $27-28$	-0.240	-0.142	-0.078	-0.054
	(0.156)	(0.091)	(0.149)	(0.089)
WAGE 29	-1.447***	-0.924***	-1.351***	-0.882***
	(0.142)	(0.083)	(0.131)	(0.078)
WAGE $30-33$	-1.158***	-0.750***	-1.058***	-0.702***
	(0.151)	(0.089)	(0.141)	(0.085)
WAGE $34-35$	-0.466***	-0.339***	-0.261	-0.219**
	(0.179)	(0.099)	(0.169)	(0.093)
WAGE $36-37$	-1.392***	-0.836***	-1.270***	-0.778***
	(0.177)	(0.099)	(0.162)	(0.093)
Obs	4196	4196	4833	4833
Hansen J	0.711	1.167		
Hansen J p-val	0.399	0.280		
Kleibergen-Paap rk LM	363.5	519.8	396.4	526.1
Cragg Donald Wald F	51.96	66.63	53.72	65.89

2-step GMM estimation. Depvar is estimated exporter-sector dummy: $\widehat{fe}^{k}_{i,t}$.
Regressors are logs of instrumented TFP and sectoral wages. Wages are efficiency adjusted in (II);(IV).
The coefficient on WAGE corresponds to elasticity for sector $17-18$.
For every other sector: elasticity given by sum of coef. WAGE and coef. of sector.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included.

$\theta = \{6.7, 7.8\}$ with a standard error of about 0.5. The assumption of sector-specific coefficients on hourly wages is not rejected by the data.

In App.D.2 we report first- and second-stage results obtained under the assumption that the production function is common across sectors. In this case, the point estimate for θ is 4.5 which is consistent with the intuition that we pick up a higher dispersion in efficiency when we do not control for sector-specific characteristics.

Tab.2 reports implied sector specific factor shares together with the mean

Table 2: Sectoral factor share of inputs

	DATA	(I)	(II)	(III)	(IV)
17-18	0.68	0.82	0.79	0.79	0.78
19	0.72	0.97	0.88	0.95	0.87
20	0.67	0.64	0.68	0.65	0.68
21-22	0.63	0.61	0.65	0.62	0.66
24	0.69	0.74	0.74	0.75	0.75
25	0.65	0.74	0.74	0.74	0.74
26	0.62	0.70	0.71	0.70	0.71
27-28	0.66	0.78	0.77	0.78	0.78
29	0.64	0.62	0.66	0.62	0.66
30-33	0.66	0.66	0.68	0.66	0.69
34-35	0.76	0.75	0.74	0.76	0.75
36-37	0.65	0.62	0.67	0.63	0.68

Col. "D" reports income share of inputs for the EU-15 in WIOD (mean in 1995-2009).
Col. (I)-(IV) report factor shares of inputs computed from estimated coefficients.

value of sectoral income shares observed in the data for the EU-15. There are clear discrepancies between the two sets of parameters. In particular, the variance is much higher in estimated parameters. Nonetheless, the positive correlation between the data and the values implied from estimation is strong.

4 The proximity mechanism and the pattern of intersectoral specialization

4.1 The proximity characteristic

To test for the presence of the proximity mechanism in the data, we first define an empirical counterpart to the indicator of proximity to suppliers. We then test whether locational comparative advantage, defined as the log difference in instrumented proximity interacted with the input intensity characteristic of the sector, contributes to determining intersectoral specialization in the residual component of RCA rankings.

As shown in sec.1.2, the microfounded proximity indicator is a weighted l^{θ}-norm of the vector of bilateral trade frictions in each sector, aggregated across sectors according to the Cobb-Douglas price index with exponents given by sectoral expenditure shares:

$$\left[\overline{PROX}_{i,t}^{M}\right]^{-1} = \prod_{s=1}^{S}\left\{\sum_{n=1}^{N}\pi_{ni,t}^{s}\tau_{ni}^{\theta}\right\}^{\gamma^{s}/\theta} \quad (20)$$

The proximity characteristic is constituted by four components: bilateral trade frictions, bilateral sectoral market shares, expenditure shares in each sector, and a parameter measuring the dispersion of productivity.

According to our modelling of trade costs, bilateral trade frictions pick up impediments to trade linked to physical features of the trade network, such as information and transport costs. As is common in the literature, we consider that a satisfactory approximation to this symmetric cost component is bilateral distance $dist_{ij}$: $\tau_{ij} = dist_{ij}^{\rho}$, with $\rho = 1$.[42]

As benchmark for θ, we use point estimates obtained in specifications (I) and (II): respectively 7.26 and 6.72 (see sec.3). Results are not sensitive to taking alternative values in the range of conventional values for this parameter: $\theta \in [4.5, 8.5]$.

Sectoral expenditure shares $\gamma_{j,t}^{k}$ are constructed using data on total output, exports, and imports. Sectoral expenditure is $X_{j,t}^{k} = PROD_{j,t}^{k} - EXP_{j,t}^{k} + IMP_{j,t}^{k}$, with EXP total exports, and IMP total imports. Expenditure shares are given by $\gamma_{j,t}^{k} = X_{j,t}^{k} / \sum_{s=1}^{S} X_{j,t}^{s}$. Output data is taken from the WIOD database (see sec. 3). Total sectoral exports and imports are obtained from the COMTRADE database where we take information on world exports and imports for each country of our sample at the ISIC Rev.3 nomenclature at the 4-digit level, and we aggregate this data to the level of 13 manufacturing sectors to be consistent with output data provided by the WIOD. In practice, we use sectoral expenditure shares for the EU-12 in each year, and we check that results are not sensitive to the assumption of common sectoral expenditure shares across countries: $\gamma_{j,t}^{k} = \gamma_{t}^{k}$ where γ_{t}^{k} is the expenditure share on sector k in year t for the EU-12.[43]

Sectoral bilateral market shares $\pi_{ij,t}^{k} = X_{ij,t}^{k} / X_{j,t}^{k}$ are constructed using bilateral imports data at the sectoral level together with data on sectoral output.[44]

The main difficulty consists in obtaining plausible measures of domestic expenditure on domestic production $X_{jj,t}^{k} = PROD_{j,t}^{k} - EXP_{j,t}^{k}$. For 20% of observations, consumption of domestic varieties is negative. Data on domestic market share in adjacent years is used to adjust observations with negative values of domestic consumption. If domestic consumption is also negative in adjacent years, we use the median value of domestic market

[42]The ranking of countries according to the proximity characteristic is not sensitive to picking an alternative value of ρ.

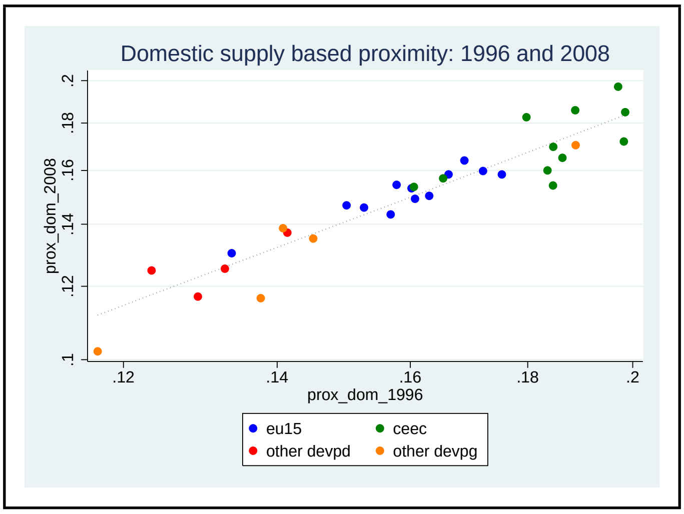
[43]The EU-12 are the EU-15 with the exclusion of Luxembourg, Belgium, and Netherlands for which output data is inconsistent with data on total exports and imports. Expenditure shares for the EU-12 are persistent overtime, and results are not sensitive to assuming $\gamma_{t}^{k} = \gamma^{k}$ in 1995-2009.

[44]Data on bilateral imports from each of their trading partners for countries in the sample is taken from the COMTRADE database.

share across the years for which domestic market share is positive. Data adjustment is done in a way which leaves the original data on output and total imports unchanged. We redefine domestic market share $\widetilde{\pi}_{jj,t} = \pi_{jj,t'}$ using information in the adjacent year (or the median). We then redefine domestic consumption as $\widetilde{X}^k_{jj,t} = \widetilde{\pi}_{jj,t} * X^k_{j,t}$. Finally, we adjust the value of total sectoral exports to be consistent with adjusted domestic market shares given data on output and total imports.

$$\widetilde{EXP}^k_{j,t} = \frac{PROD^k_{j,t} - \widetilde{\pi}_{jj}(PROD^k_{j,t} + IMP^k_{j,t})}{(1-\widetilde{\pi}_{jj})}$$

Figure 1: Proximity characteristic in 1996 and 2008



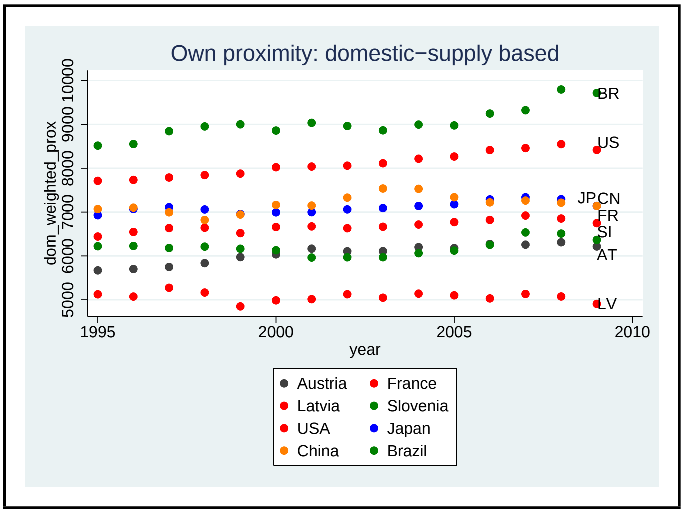
The proximity characteristic cannot be constructed for 10 countries of our sample due to lacking or inconsistent output data (see table 11). Fig.1 plots the proximity characteristic for the remaining set of countries in 1996 and 2008.[45] There is substantial variability across countries in proximity to suppliers. The sample is split in four subgroups according to the ranking of proximity: the EU-15, the CEECs, non-European developed, and non-European developing economies.[46]

Fig.2 plots the indicator of distance to suppliers measured in kilometers, i.e. the reciprocal of the proximity characteristic, for two countries of each

[45] Proximity indicators are rescaled by 10^3 in this graph.

[46] Two countries behave differently from their subgroup: Russia is characterized by relatively high and Great Britain by relatively low proximity.

Figure 2: Microfounded proximity (subset of countries)



subgroup in 1995-2009.[47] For each of the four subgroups we plot the measure of distance to suppliers for the most and the least distant country from least cost inputs.[48]

The persistence of the proximity characteristic is not surprising. Bilateral trade costs are time invariant by construction. This conforms to the assumption that bilateral components of trade costs correspond to persistent characteristics of the trade network such as transport and information costs, as captured by bilateral distance. Moreover, the weighting function is defined by sectoral bilateral market shares. The market share is the probability that a specific source is revealed least cost in the sector. The fundamental country characteristic which determines this probability is the expected sectoral productivity. Levchenko and Zhang (2011) show that in half a century the distribution of sectoral technology stocks undergoes substantial changes. It is however likely that in the short time span studied in this paper this distribution is relatively stable.

[47]See app.E for exhaustive information on each subgroup overtime.

[48]There are two exceptions to this rule. In the EU-12 group we do not plot Great Britain as the most distant country because France is more representative of countries in the group. In the non-EU developing group we do not plot Russia as the least distant country because China is more representative of countries in the group.

4.2 Motivation and construction of an instrument

The proximity characteristic is clearly an endogenous object contingent on a specific trade equilibrium. But it is not for this reason that it should be instrumented in the third step of the estimation. Indeed, the distribution of sectoral market shares is uniquely determined in a given trade equilibrium and consequently uniquely determines the ranking of countries in terms of effective proximity to suppliers in a given cross section of the data. It follows that this distribution of market shares provides the weighting system to be used in constructing the proximity index required to test empirically the predictions of the structural model.

What motivates our instrument is measurement error in the distribution of market shares. This measurement error stems from the assumption that destinations do not differ in their restrictiveness to foreign supply. As shown formally in App.E, distance to suppliers is underestimated by more for countries which are relatively closed to foreign supply whenever there is trade restrictiveness variability. This is because both domestic market share and the indicator of proximity need to be rescaled by the trade restrictiveness index (TRI) of the destination. TRI variability is likely to be pervasive in the data over the period under study. To illustrate, notice that in 2008 trade restrictiveness in manufacturing in Brazil was almost seven times higher than in the European Union (manufacturing TRI in 2008 constructed by Kee et al. (2009) are reproduced in App.E for a subsample of countries).

The second reason for instrumenting the proximity characteristic is due to the fact that the error term in the third step of the estimation may still contain exporter-sector specific cost components linked to domestic technology.[49] As proximity indices depend on the distribution of sectoral market shares, they necessarily depend on domestic sectoral technology stocks. We need to instrument the proximity characteristic with an indicator of proximity independent of the distribution of market shares to avoid violating the assumption of statistical independence of the residuals. We think of this instrument as an index of proximity endowment because it is not contingent on a specific trade equilibrium.[50]

It is not immediate how to construct the empirical counterpart of proximity endowment. As we seek to capture impediments to trade linked to physical features of the trade network such as transport or information costs, we consider that the fundamental component of proximity endowment is given

[49] The residual contains all cost components not picked up by instrumented TFP and wages as shown in (14).

[50] In App.E.1 we put forward that proximity endowment would be the exact measure of proximity to suppliers in a world without bilateral trade frictions.

by the physical location of the country relatively to all of its potential trade partners.

If geographical location is key, then one way to measure this endowment is to compute the length of the bilateral distance vector for this country with all of its potential suppliers including the country itself. This provides a time-invariant indicator of countries' centrality under the assumption that bilateral distance is a sufficiently good proxy of trade impediments other than trade policy. This assumption is difficult to verify without data on transport costs, but it is widely used in empirical applications.

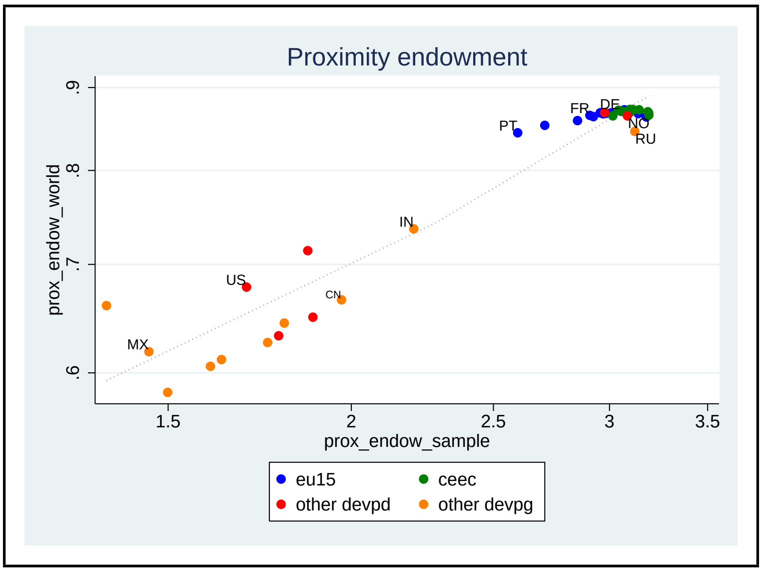
$$\left[PROX_i^M\right]^{-1} = \left[\sum_{n=1}^{N} dist_{in}^2\right]^{0.5}$$

The caveat of such an instrument is that by definition relative proximity endowment captures proximity to suppliers as well as proximity to clients. However, this is not a problem for identification. First, we show below that this instrument correctly captures the ordinal ranking of countries in terms of supplier-based microfounded proximity. Second, we have shown in sec.2 that conceptually relative proximity to clients corresponds to something different: it is an exporter-specific trade barrier incurred in getting goods to world markets. And third, even if we believe that the norm of the distance vector constitutes a good proxy for the export-side trade cost index, the crux of our identification relies on an interaction effect which is specific to import-side trade costs. As shown in (15), a given gap in export side trade costs is not magnified in input intensive sectors. Consequently, as long as we are getting the ranking of sectors in terms of input intensity and the ranking of countries in terms of proximity to suppliers right, we will be identifying the impact of the input cost channel on the pattern of specialization.[51]

We compute two measures of proximity endowment. First, we measure proximity endowment as the reciprocal of the l^2-norm of the distance vector while restricting the number of potential input suppliers to the 42 countries of the sample. Second, we measure proximity endowment as the reciprocal of the l^2-norm of the distance vector with the 224 countries in the world on which internal and bilateral distance data are available (Mayer and Zignago (2011)). Fig.3 illustrates that countries' ranking is not sensitive to restricting the number of potential trade partners to the 42 countries of our sample.

[51] A robustness check which we leave for future work consists in using the input-output matrix to calculate sector-specific proximity indices instead of assuming that all sectors source inputs in the same way. Arguably, the sectoral distribution of import shares combined with the sector-specific structure of input sourcing would be decoupled from proximity to clients.

Figure 3: Proximity endowment



Finally, we check whether this time-invariant indicator is a good predictor of the proximity characteristic computed in the previous subsection. As illustrated in fig.4, the two proximity measures are strongly correlated. The measure of proximity endowment obtained when the number of potential suppliers is restricted to the 42 countries of our sample explains 70% of the variation in weighted proximity.[52] The caveat, as is clear from the graph, is that the instrument does a relatively poor job in capturing variation in proximity within subgroups.

The ranking of countries in terms of relative proximity is invariant to using the indicator of proximity endowment or the time-varying proximity characteristic. As shown in fig.5 and 6, both measures allow splitting the sample in the high-proximity group which includes the EU-15 and the CEECs, and a low-proximity group which includes non-European emerging and developed economies.

However, the magnitude of the proximity gap differs across the two indicators for a subset of countries. In particular, China and India appear relatively more distant from world technology when the indicator of proximity endowment is used. This discrepancy may be driven by the overestimation of the time-varying proximity characteristic in markets which are relatively closed

[52]Proximity endowment computed using the full set of potential trade partners captures less than 60% of variation in the time-varying proximity characteristic.

Figure 4: Proximity endowment and microfounded proximity

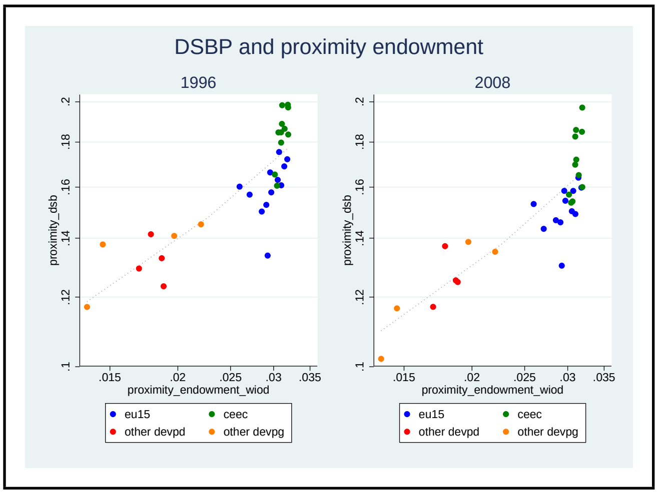
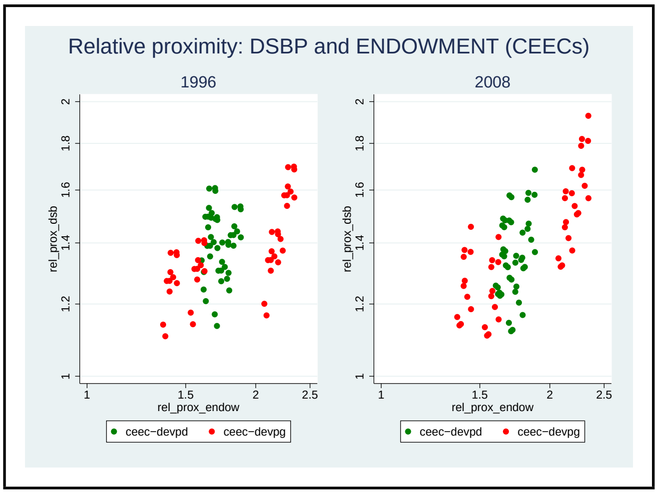


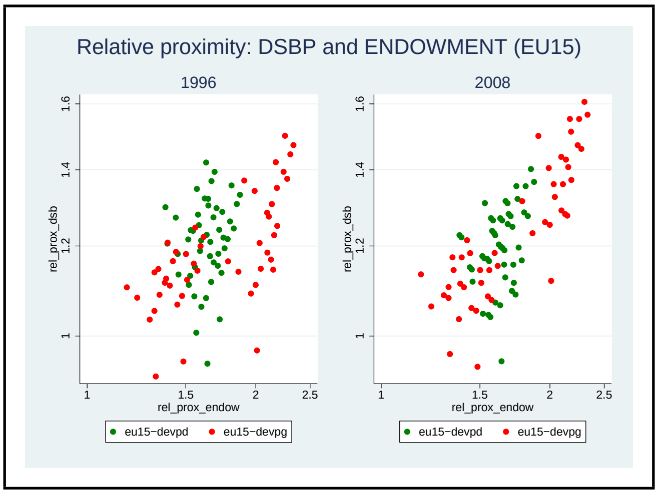
Figure 5: Relative proximity for the CEECs



to foreign supply (see app.E for the formal presentation of this argument).

We conclude that relative proximity, however measured, is expected to generate persistent differences in relative producer costs in our sample. Our model predicts that these cost differences should be source of comparative advantage because the wedge in the cost of the input bundle is enhanced in sectors where the weight of inputs in total costs is relatively high. We check whether the proximity mechanism is active in the data.

Figure 6: Relative proximity for the EU15



4.3 Proximity in intersectoral specialization

The proximity mechanism captures sector-specific differences in production costs which arise because countries differ in their ability to source inputs from the best technology worldwide. In this subsection, we check whether the cost advantage conferred by the ability to source inputs at relatively lower cost leads to specialization of high proximity countries in sectors which use inputs relatively more intensively.

The sample is split in two groups according to the proximity characteristic, with the EU15 and the CEECs in the 'high' and the non-European developed and developing countries in the 'low' proximity group. As we have already controlled for differences in relative technology stocks and relative wages, the level of development of the country should not be a source of within-group heterogeneity. However, differences in domestic market openness to foreign supply may bias the microfounded measure of relative proximity in cross-section and overtime (see app.E.4). Consequently, in certain specifications we also split each country group in two subgroups by crossing the proximity criterion with the criterion of domestic market openness.

Estimated residuals of the second step specification $\widehat{\lambda}_{i,t}^{k}$ are rescaled by the estimated heterogeneity parameter $\widehat{\theta}$. We then compute all pairwise combinations of sectoral annual residuals $(1/\widehat{\theta})(\widehat{\lambda}_{i,t}^{k} - \widehat{\lambda}_{i',t}^{k})$ where $i \in H$ are countries of the high proximity group, and $i' \in L$ are countries of the low proximity group.

We estimate (21) on data pooled for all years. Exporter-year fixed effects

$\{fe_{i,t} - fe_{i',t}\}$ are included to control for characteristics of the benchmark sector for each exporter and year.

$$\frac{1}{\widehat{\theta}}\left[\widehat{\lambda}^k_{i,t} - \widehat{\lambda}^k_{i',t}\right] = \beta_0 + \beta_1 \ln \left\{ \left(\frac{\widehat{\overline{PROX}}^M_{i,t}}{\widehat{\overline{PROX}}^M_{i',t}} \right)^{\widehat{\zeta^k}} \right\}$$
$$+ fe_{i,t} - fe_{i',t} + \eta^k_{ii',t} \tag{21}$$

The coefficient of interest is β_1: according to the model, β_1 should be positive and close to 1. In practice, there are potentially multiple sources of measurement error in the data which we have seeked to eliminate via instrumenting. Nonetheless, it is prudent to focus on the sign rather than on the absolute value of the estimated coefficient.

Table 3 shows that the proximity mechanism contributes to determining intersectoral specialization in the residual component of RCA rankings in the way predicted by the model, with high proximity countries producing relatively more for world markets in sectors which use inputs relatively more intensively: β_1 is positive and significant across specifications. In col.(1)-(4), this result is obtained when countries are grouped in 2 bins according to the proximity characteristic while in col.(5)-(6) we additionally split low proximity countries in two groups according to the initial trade restrictiveness of their domestic markets.

Col.(1) and (3) show that results are robust to the set of instruments used in the second step of the estimation. In col.(1)-(2) the set of instruments is R&D personnel, real capital stocks, and efficiency-adjusted sectoral workforce while in col.(3)-(4) it is deflated R&D expenditure and sectoral workforce unadjusted for efficiency. Results are qualitatively similar in (II) and (III) (not shown). In col.(5) and (6) it is shown that the proximity mechanism plays out strongly relatively to countries with low centrality and low levels of trade restrictiveness. Evidence in favor of the proximity mechanism is weaker relatively to countries with low centrality and high trade restrictiveness. The proximity mechanism becomes a stronger predictor of the ranking of sectoral exports in the recent period as shown in col.(2) and (4).

Table 4 shows more evidence on the proximity mechanism by splitting the sample of countries in four subgroups. The mechanism is present in the data in the majority of specifications.

This subsection has shown that once we control for technology stocks and factor endowments, the proximity mechanism determines the ranking of residual relative sectoral exports across country pairs. Locational comparative advantage thus contributes to shaping the pattern of intersectoral

Table 3: Proximity mechanism in the residual component of RCA rankings

	all (I)	*all* (I)	*all* (IV)	*all* (IV)	*both-to-devpd* (I)	*both-to-devpg* (I)
$relprox * inpint$	0.689***	0.375***	1.255***	0.658***	1.288***	0.176**
	(0.064)	(0.093)	(0.100)	(0.152)	(0.101)	(0.078)
recent		0.585***		1.033***		
		(0.126)		(0.200)		
Obs	17748	17748	20097	20097	8883	8865
R^2	0.674	0.674	0.665	0.665	0.541	0.776
Recent FE		YES		YES		

Depvar is rescaled relative residual component of the exporter-sector dummy: $1/\theta \left[\widehat{\lambda}^k_{i,t} - \widehat{\lambda}^k_{i',t}\right]$.
In (I) instruments are R&D personnel, real capital stocks, efficiency-adjusted workforce.
In (IV) instruments are deflated R&D expenditure and raw data on workforce.
'$relprox * inpint$' is log of relative proximity interacted with sectoral input intensity.
'recent' is interaction between proximity and the 2001-2007(9) subperiod.
Exporter-year fixed effects are included in each specification.
col.1-4: countries split in two groups according to proximity ranking.
col.5-6: EU15 and CEECs to resp. developed and developing.

Table 4: Proximity mechanism in the residual component by subgroup

	(I)	**(II)**	**(III)**	**(IV)**
eu15-to-devpd	1.379***	2.359***	1.344***	2.263***
std-error	(0.134)	(0.224)	(0.125)	(0.207)
nb-obs	5541	5541	6399	6399
R^2	.485	.483	.477	.476
ceec-to-devpd	1.151***	2.242***	0.890***	1.712***
std-error	(0.156)	(0.259)	(0.142)	(0.233)
nb-obs	3342	3342	3894	3894
R^2	0.517	0.518	0.507	0.506
eu15-to-devpg	0.165	0.356**	0.254**	0.520***
std-error	(0.105)	(0.177)	(0.103)	(0.171)
nb-obs	5529	5529	6100	6100
R^2	0.742	0.740	0.741	0.738
ceec-to-devpg	0.191*	0.623***	0.127	0.489***
std-error	(0.113)	(0.188)	(0.108)	(0.178)
nb-obs	3336	3336	3704	3704
R^2	0.782	0.779	0.784	0.780

(I)-(IV) differ in the set of instruments for TFP and hourly wages.
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV).
Exporter-year fixed effects are included in each specification.
Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

specialization. But how much? We use the structure of the model to quantify the contribution of this mechanism relatively to the contribution of domestic technology and factor endowments. In particular, we check whether the pattern of specialization predicted by relative sectoral technology stocks is modified by the proximity mechanism.

5 Decomposition of comparative advantage

5.1 Does proximity matter?

We pool data on residuals of exporter-sector dummies obtained in the second step of the estimation and construct pairwise relative sectoral residuals for high relatively to low proximity countries. We regress these relative residuals on sectoral indicators of relative proximity to check which share of the variance in the residual component of sectoral exports is explained by the proximity mechanism.[53] The proximity mechanism explains 18-20% of the variance in the residual component of relative sectoral exports (see tab.5).

Table 5: Fraction of variance attributable to the proximity mechanism

	all (I)	*all* (II)	*all* (III)	*all* (IV)
$relprox * inpint$	2.777*** (0.282)	3.381*** (0.336)	2.583*** (0.255)	3.043*** (0.297)
R^2	**0.178**	**0.200**	**0.181**	**0.196**
Obs	17,748	17,748	20,097	20,097

Depvar is rescaled relative residual component of the exporter-sector dummy.
'$relprox * inpint$' is log of relative proximity interacted with sectoral input intensity.
(I)-(IV) differ in the set of instruments for TFP and hourly wages.
We report the coefficient on relative proximity and the fraction of explained variance.

We check that the fraction of variance attributable to proximity is not explained by the intercorrelation of proximity with technology and wages by computing the coefficient of partial determination between proximity and the ranking of relative exports. This statistic measures the fraction of the variance in the residual component of relative sectoral exports attributable to the component of the proximity vector unexplained by technology and wages.[54].

The proximity vector is nearly orthogonal to vectors of instrumented TFP and wages. Consequently, the variation in the ranking of relative sectoral ex-

[53] Throughout this section the time-varying proximity characteristic is instrumented with proximity endowment.

[54] For some variable y, the coefficient of partial determination measures the fraction of variance in the residuals of y wrt x_i for $i \neq j$ explained by residuals of x_j wrt x_i.

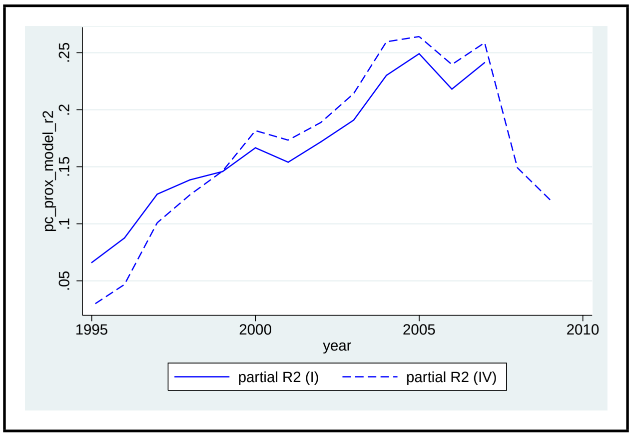
ports attributable to differences in trade frictions incurred in sourcing inputs is not reconducible to the two other characteristics which according to the model should determine the pattern of specialization (domestic technology and factor endowments). The coefficient of partial determination between proximity and relative sectoral exports is 15-17% in the full sample (see tab.6).

Table 6: Coefficient of partial determination (proximity, all years)

	all (I)	*all* (II)	*all* (III)	*all* (IV)
$resid - relprox$	2.601*** (0.305)	3.180*** (0.363)	2.446*** (0.283)	2.907*** (0.330)
R^2	**0.154**	**0.173**	**0.154**	**0.169**
Obs	17,748	17,748	20,097	20,097

Depvar is rescaled relative residual component of the exporter-sector dummy.
'$resid - relprox$' is the component of the vector of relative sectoral proximity orthogonal to instrumented TFP and wages.
(I)-(IV) differ in the set of instruments for TFP and hourly wages.

Figure 7: Coefficient of partial determination (proximity, annual)



We also check whether the contribution of proximity changes overtime by computing the coefficient of partial determination in cross section. Fig.7 documents that the fraction of unexplained variation in the ranking of relative sectoral exports attributable to the proximity mechanism increases fourfold in 10 years, exceeding 20% in 2000-2007. This result is not sensitive to the set of instruments used in the second step of the estimation (see solid line for specification (I) and dashed line for specification (IV) in fig.7).

The proportion of variance attributable to proximity plunges in 2008-2009. The reduction in the weight of the proximity mechanism is consistent with evidence that trade networks linked to production fragmentation across borders were severely hit in the aftermath of the financial crisis. Eaton et al. (2011) find that more than 80% of the decline in world trade in this period is due to the reduction in demand for manufactures (durable goods). If production of these goods is fragmented across borders, the decrease in demand leads to a reduction in inputs' trade which magnifies the reduction in trade relatively to the reduction in GDP. This chain of events reduces world trade as a share of GDP and concomitantly reduces the importance of proximity to suppliers in determining the pattern of specialization.

5.2 Variance decomposition of RCA rankings: the intersectoral component

We have documented that the proximity mechanism plays a non negligible role in explaining the variation in relative sectoral exports. However, it could be that relative proximity succeeds in capturing the fraction of variance linked to features which vary by the exporter pair but does not contribute to the pattern of intersectoral specialization. We check this hypothesis by decomposing the variance of the intersectoral component of revealed comparative advantage (RCA) rankings across technology, factor endowments, and proximity. We control for characteristics of the exporter which do not vary at the intersectoral level by including exporter-year fixed effects in the specification.

Relative exporter-sector dummies capture the components of sectoral costs which determine the pattern of comparative advantage together with the cost component of each exporter in the benchmark sector. We use estimated exporter-sector dummies $\widehat{fe}^{k}_{i,t}$ and productivity heterogeneity parameters $\widehat{\theta}$ to compute rescaled relative exporter-sector dummies for high proximity countries i relatively to low-proximity countries i': $1/\widehat{\theta}\left(\widehat{fe}^{k}_{i,t} - \widehat{fe}^{k} i', t\right)$, $\forall k, i, i'$. Estimated parameters and instrumented components of TFP, wages, and proximity are used to compute the contribution of these three characteristics to determining the pattern of specialization.

$$\frac{1}{\widehat{\theta}}\left(\widehat{fe}^k_{i,t}-\widehat{fe}^k_{i',t}\right)=\alpha_0+\alpha_1\ln\left[\frac{\widehat{z}^k_{i,t}}{\widehat{z}^k_{i',t}}\right]+\alpha_2\ln\left\{\left[\frac{\widehat{\nu}^k_{i,t}}{\widehat{\nu}^k_{i',t}}\right]^{-(1-\widehat{\zeta^k})}\right\}+$$

$$\alpha_3\ln\left\{\left[\frac{\widehat{PROX}^M_{i,t}}{\widehat{PROX}^M_{i',t}}\right]^{\widehat{\zeta^k}}\right\}+fe_{i,t}+fe_{i',t}+\xi^k_{ii',t} \tag{22}$$

According to the model, a regression of rescaled relative exporter-sector dummies on technology, wages, proximity, and exporter-year fixed effects should produce $\alpha_1=\alpha_2=\alpha_3=1$. In tab.7 we report the results of estimating (22) in the four main specifications. Consistently with the underlying model, the null hypothesis of coefficients' equality cannot be rejected in most specifications. But estimated coefficients are statistically different from 1. Col.(5) and (6) report the standardized regression coefficients for specifications (I) and (IV) respectively. To underline the relative importance of TFP and proximity, notice that one standard deviation in TFP increases relative exports by 2.5 (resp.1.9) standard deviations while 10(resp.5) standard deviations of proximity are needed to produce the same result.

Table 7: The intersectoral component of RCA rankings

	all (I)	*all* (II)	*all* (III)	*all* (IV)	*β-coef* (I)	*β-coef* (IV)
tfp	2.143***	2.105***	2.124***	1.994***	2.50	1.94
	(0.110)	(0.107)	(0.111)	(0.107)		
wage	1.981***	1.919***	2.291***	2.178***	2.32	2.08
	(0.112)	(0.109)	(0.120)	(0.117)		
proximity	1.668***	2.964***	1.642***	2.861***	0.24	0.40
	(0.160)	(0.274)	(0.156)	(0.265)		
R^2	0.731	0.731	0.731	0.726		
Obs	17,748	17,748	20,097	20,097		

Depvar is rescaled relative exporter-sector dummy: $1/\theta\left[fe^k_{i,t}-fe^k_{i',t}\right]$.
(I)-(IV) refer to alternative instrumenting procedures for technology and wages.
Col.(5)-(6) report standardized regression coefficients. Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Exporter-year fixed effects are included in each specification.
Standard errors are clustered by pair.*** p<0.01, ** p<0.05, * p<0.1

Tab.8 reports the coefficient of partial determination between technology, wages, proximity and relative sectoral exports for the full sample and by subgroup in the specification with exporter-year fixed effects. Technology largely outweighs the proximity mechanism in explaining the residual variation in relative sectoral exports at the intersectoral level. The relatively minor contribution of proximity may be in part an artefact of the simplifying assumption that all sectors use the same input bundle.

Table 8: Coefficient of partial determination: full sample and subgroups

	tfp	**wage**	**proximity**
all	**0.28**	**0.24**	**0.04**
both-to-devpd	**0.25**	**0.19**	**0.02**
eu15-devpd	0.26	0.17	0.03
ceec-devpd	0.26	0.23	0.00
both-to-devpg	**0.32**	**0.30**	**0.04**
eu15-devpg	0.34	0.33	0.06
ceec-devpg	0.27	0.26	0.02

Our objective is to quantify the fraction of the variance attributable to proximity out of total explained variance at the intersectoral level. We compute this as the ratio of semipartial r^2 of proximity to the sum of semipartial r^2 for instrumented TFP, wages, and proximity. This sum defines the fraction of the variance in relative sectoral exports unexplained by exporter-year fixed effects and uniquely associated with these three variables.[55] By construction, the fraction of variance unexplained by exporter-year fixed effects but associated with more than one of these three regressors is excluded. Total intersectoral variance is defined in this restrictive way to avoid doublecounting.

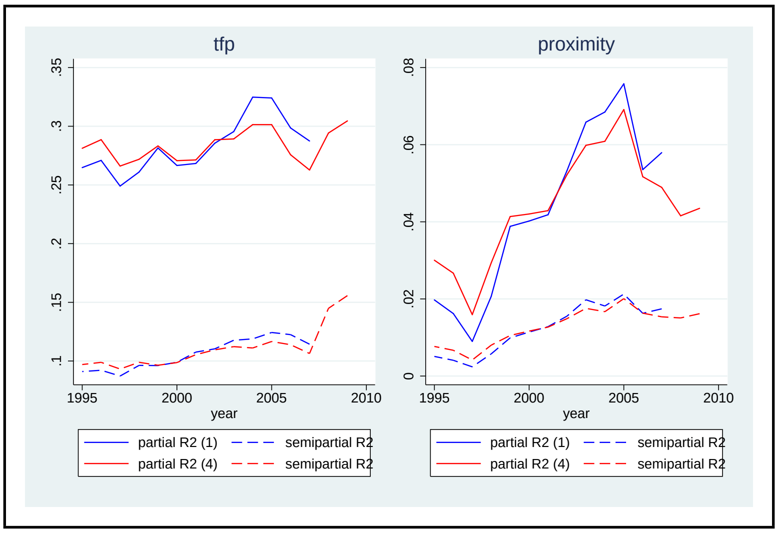
When we pool data for all years in the full sample, total explained variance at the intersectoral level uniquely associated with instrumented TFP, wages, and proximity is .20 in (I) (.23 in (IV)), i.e. about 30% of total explained variance. Domestic technology corresponds to 53% (resp. 47%) of this total while just 6% (resp.5%) is reconducible to proximity.

When we look at the sequence of cross sections the proximity mechanism is found to play an increasing role in the intersectoral component of RCA rankings. Fig.8 reports the squared partial and semipartial correlation for domestic technology and proximity for specifications (I) and (IV). Both specifications are reported to check the sensitivity of results to the set of instruments for TFP and wages used in the second step of the estimation.

The coefficient of partial determination illustrates that the fraction of the residual variance attributable to technology remains stable at slightly

[55]The semipartial r^2 reports the fraction of total variance in relative sectoral exports explained by each regressor residualized with respect to all the other regressors.

Figure 8: Partial and semipartial r^2 in cross section: full sample



less than 30% in 1995-2007 while the fraction attributable to the proximity mechanism increases from 2 to about 7%.

If we restrict attention to total explained variance at the intersectoral level, the fraction attributable to the proximity mechanism increases by 5 percentage points in 1995-2007 in (I), from 3 to 8%, while the fraction attributable to technology increases from 53 to 55%.[56]

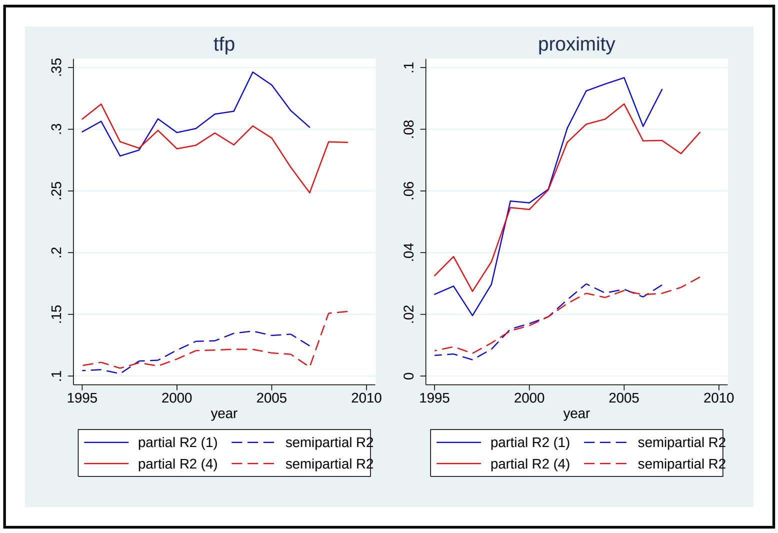
We nuance these findings by conducting the variance decomposition in cross section by subgroups. Fig.9 reports the squared partial and semipartial correlation for domestic technology and proximity for specifications (I) and (IV) for the EU-15 relatively to low proximity countries. Results for this subsample broadly replicate our findings for the full sample. Total explained variance at the intersectoral level increases from .20 to .23 in 1995-2007 in (I). The fraction of this total attributable to technology increases by 2 percentage points from 53 to 55%. Over the same period the contribution of proximity to explaining the intersectoral component of RCA rankings increases by 10 percentage points, from 3 to 13%.[57]

Fig.10 graphs the contribution of technology to explaining the variation in relative sectoral exports separately for the EU-15 and for the emerging

[56]The sum of semipartial r^2 for instrumented TFP, wages, and proximity increases from .17 to .21 in (I), and from .20 to .22 in (IV). In (IV) the contribution of proximity increases from 4 to 7%.

[57]In (IV) the contribution of proximity increases from 4 to 12% in 1995-2007.

Figure 9: Partial and semipartial r^2: EU15

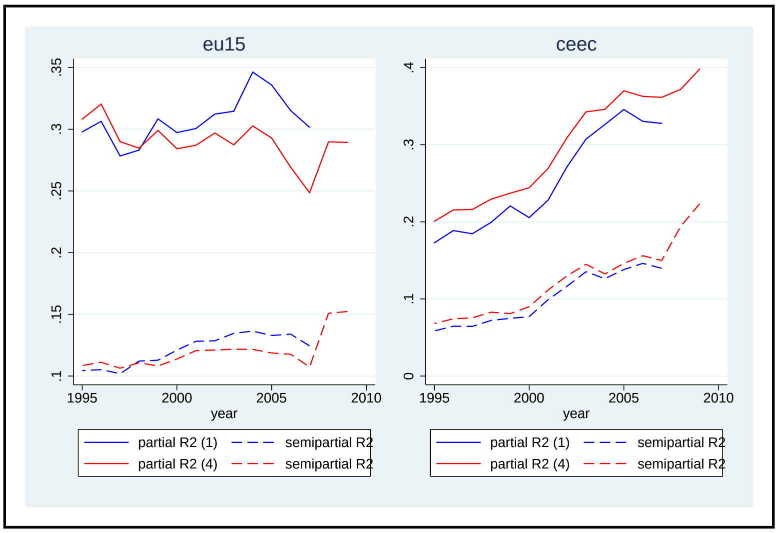


economies of Central and Eastern Europe relatively to low proximity countries. The fraction of residual variance attributable to technology is just .17 for the CEECs in 1995, much lower than the corresponding statistic for the EU-15. However, by 2009 almost 40% of the residual variation in relative sectoral exports is attributable to the intersectoral variation in technology for the CEECs. The corresponding statistic for the EU-15 is almost ten percentage points lower.

Over the same period the fraction of total explained variance at the intersectoral level increases from .11 to .26 for the CEECs relatively to low proximity countries. This means that the share of the variance uniquely associated with intersectoral determinants of comparative advantage increases from 15 to 36% of total explained variance. The contribution of technology to explaining the intersectoral component of RCA rankings is roughly stable at 52-53% of the total while the share attributable to proximity is reduced from 4 to 2%. These findings indicate that differential technology upgrading across sectors is likely to have increasingly shaped CEECs' speciailzation pattern on world markets.

As a final check we split the sample of low proximity countries in two subgroups according to the level of trade restrictiveness of their domestic markets. For the EU-15 the contribution of the proximity mechanism increases in both subsamples: from 5 to 10% (resp. from 2 to 10%) in the

Figure 10: Partial and semipartial r^2: technology



subsample of relatively open (resp. closed) economies.[58]

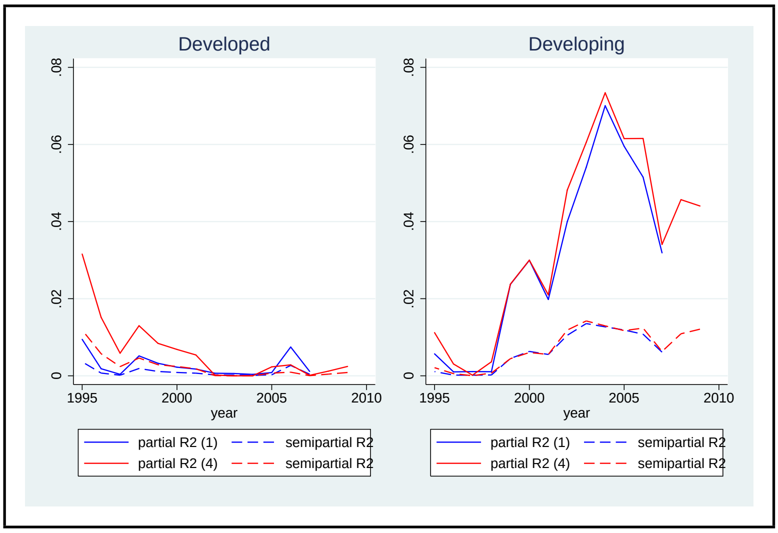
Fig.11 graphs the results for the CEECs. The fraction of the variance uniquely associated with proximity is about nil for the CEECs relatively to non-European economies open to foreign supply. On the other hand, the proximity mechanism plays an increasing role in explaining the intersectoral variation in RCA rankings of the CEECs relatively to non-European economies which are relatively closed to foreign supply. Total variance uniquely associated with instrumented TFP, wages, and proximity increases from .07 to .20 in this subsample over 1995-2004. The fraction attributable to proximity increases from 1.5 to 6.4%, but is subsequently reduced to 2.8% by 2007.[59]

This section has documented that the proximity mechanism explains a small share of the variation in relative sectoral exports at the intersectoral level relatively to the variation attributable to technology and labor endowments. We think of this result as establishing a lower bound on the impact of the input cost channel as a consequence of our simplifying assumption that all sectors use the same input bundle in production.

[58]Low proximity and low domestic restrictiveness economies correspond to non-European developed countries. Low proximity and high restrictiveness economies correspond to non-European emerging economies.

[59]The reduction in sample size is reflected in results' sensitivity to the instrumenting procedure. In specification (IV) the contribution of proximity is stable at 2.3-2.6% while the contribution of technology is stable at 49% in 1995-2007.

Figure 11: Partial and semipartial r^2: proximity (CEECs)



We find robust empirical evidence that the importance of the proximity mechanism is increasing overtime. This result is consistent with a rapidly growing empirical literature which, starting with the seminal paper by Hummels et al. (2001), has documented increasing internationalization of the production process.[60] Indeed, we expect the input cost component of comparative advantage to matter relatively more in shaping the pattern of intersectoral specialization when the pattern of production and trade costs makes it optimal for producers to increasingly segment production across borders.

Finally, we find that the fit of the model to the data is satisfactory. The full model explains about 3/4 of total variation in relative sectoral exports. Roughly a third of this total is uniquely associated with the intersectoral variation of the three explanatory variables which according to the model should determine the pattern of intersectoral specialization. Domestic technology explains between 53-55% of the variation in the intersectoral component of RCA rankings while the cost advantage conferred to high-proximity countries by the ability to source inputs relatively more cheaply explains between 3 and 8% of the intersectoral variation. For countries of the EU-15 the contribution of the proximity mechanism increases by 10 percentage points in 1995-2007, from 3 to 13% of total explained variance at the intersectoral level.

[60] The non exhaustive list is Daudin et al. (2011), Johnson and Noguera (2012c,a,b), Stehrer (2012) and Borowiecki et al. (2012).

5.3 Does technology determine the pattern of intersectoral specialization under positive trade costs?

Finally we ask whether the input cost channel modifies the pattern of comparative advantage relatively to the ranking of relative exports which would prevail if this pattern were uniquely determined by the ranking of relative domestic sectoral technology.

In col.(1) of table 9 we report the sign of the relationship between overall RCA rankings and the proximity mechanism. In col.(2) we report the sign of the relationship between the intersectoral component of RCA rankings predicted by technology, wages, and proximity with the proximity mechanism. Col.(3)-(4) report the sign of the relationship with the proximity mechanism for (resp.) rankings of relative technology $\widehat{z}^k_{i,t,}/\widehat{z}^k_{i',t}$ and relative costs $\widehat{\omega}^k_{i,t}/\widehat{\omega}^k_{i',t}$. Finally, col.(5) recalls the sign of the relationship between the residual intersectoral component of RCA rankings and the proximity mechanism obtained in the third step of the estimation.

We document that the proximity mechanism contributes to determining the pattern of comparative advantage through the sectoral cost component $\omega^k_{i,t}$ in most specifications. Furthermore, the proximity mechanism is always picked up in the residual intersectoral component of RCA rankings orthogonal to instrumented TFP and wages.

Table 9: The intersectoral component of RCA rankings

	overall RCA (1)	predicted RCA (2)	relative TFP (3)	relative cost (4)	relative residual (5)
all	−	−	−	+	+
both-to-devpd	+	+	+	?	+
both-to-devpg	−	−	−	+	+
eu15-to-devpd	+	+	+	−	+
ceec-to-devpd	+	+	+	+	+
eu15-to-devpg	−	−	−	+	+
ceec-to-devpg	−	−	−	+	+

In (1) depvar is rescaled relative exporter sector dummy: $1/\widehat{\theta}\left(\widehat{fe}^k_{it} - \widehat{fe}^k_{i't}\right)$.
In (2) depvar is the ranking of relative exports predicted by instrumented TFP, wages, and proximity.
In (3) depvar is relative instrumented TFP: $\ln(\widehat{z}^k_{it}/\widehat{z}^k_{i't})$.
In (4) depvar is relative sector-specific production cost: $\ln(\widehat{\omega}^k_{it}/\widehat{\omega}^k_{i't})$.
In (5) depvar is rescaled relative residual of sectoral exports: $1/\widehat{\theta}\left(\widehat{\lambda}^k_{it}/\widehat{\lambda}^k_{i't}\right)$.

Nonetheless, the proximity mechanism does not inflect the pattern of intersectoral specialization driven by domestic technology. Indeed, whenever relative domestic technology rankings covary positively with proximity, the intersectoral component of RCA rankings picks up a positive link with the proximity mechanism. This is the case in col.(2) for the pattern of specialization of European countries relatively to non-European developed economies. However, whenever this is not verified, the intersectoral component of RCA rankings negatively covaries with relative sectoral proximity. In particular, this is the case in col.(2) for the specialization pattern of European countries relatively to non-European emerging economies.

We conclude that in any of the cross-sections we have looked at, the proximity mechanism shapes the pattern of intersectoral specialization conditional on a given distribution of technology and labor endowments. However, it does not inflect the ranking of relative sectoral exports predicted by relative technology stocks. Consistently with the fundamental intuition of Ricardian models, the pattern of comparative advantage is determined by the ranking of relative sectoral technology stocks, even under positive trade costs and trade in inputs.

This conclusion has to be qualified in two ways. First, our results are likely providing a lower bound on the importance of the proximity mechanism in the data. Second, even though the main determinant of specialization is domestic technology, we find that the importance of proximity to world technology is increasing overtime.

6 Conclusion

It has become common knowledge that the process of production is increasingly fragmented internationally. In this paper we have investigated two questions: whether production unbundling has become a new source of comparative advantage and whether it has modified the determination of countries' specialization pattern on global markets. We answer 'yes' to the first, and 'no' to the second question.

Our main result is that production unbundling has coincided with an inscreasing role of input costs in shaping the pattern of comparative advantage in 1995-2009. But we also find that in our sample of 36 developed and emerging economies the ranking of relative sectoral technology stocks continues to determine the overall pattern of revealed comparative advantage just as in the benchmark multisectoral Ricardian world with bilateral trade frictions but no sector specific production characteristics.

We have shown that the only component in the cost of the input bundle which varies across countries is given by a composite index of trade frictions incurred in sourcing inputs from all potential suppliers. Conceptually, in relative terms, the proximity characteristic is a summary statistic of locational comparative advantage because it captures the cost advantage conferred to the country through its ability to source the cheapest inputs worldwide, relatively to every other country in the world.

Relative proximity is also a summary statistic of the relative cost of living for any pair of countries. Indeed, we show that the overall price index can be decomposed in an index which captures the realized distribution of least cost technology in the world and an index of trade frictions which is country specific. Consequently, the closer the country is to the best world technology, and the lower is its cost of living relatively to other countries. A complementary result of the paper is that relative real wages can be computed by adjusting the ratio of nominal wages by relative proximity while circumventing the problem of constructing actual price indices.

As the cost share of inputs is sector-specific the wedge in the cost of inputs becomes source of comparative advantage. The model predicts that once we control for domestic technology and labor endowments, we should find that countries characterized by relatively high proximity to suppliers specialize in sectors which use inputs relatively more intensively. We present robust empirical evidence confirming this prediction in the data.

The input cost channel explains 15-20% of the residual variation in relative sectoral exports, but just 6% of total explained variation in the intersectoral component of RCA rankings if data is pooled in 1995-2009. In annual cross sections between 53-55% of the total variation in the intersectoral component of relative sectoral exports is attributable to technology while the contribution of the input cost channel increases from 3 to 8% in the full sample, and from 3 to 13% for EU-15 countries. The input cost channel is not only active at the intersectoral level, but acquires increasing importance overtime.

This line of research can be pursued in two directions. First, it would be ineresting to investigate which type of shocks to the distribution of technology or to the structure of trade costs would be needed to inflect the pattern of comparative advantage given our result that the characteristics which determine intersectoral specialization are very slow moving. Second, it would be interesting to improve the mapping of the model to the actual sectoral structure of input sourcing to compute sector-specific theory-based indices of trade frictions. This more realistic production structure will help test our assertion that the results presented in this paper correspond to the lower bound of the true contribution of the input cost channel to the pattern of countries' specialization on world markets.

References

Anderson, J. E. and P. J. Neary (1996). A new approach to evaluating trade policy. The Review of Economic Studies 63(1), 107–125.

Bils, M. and P. Klenow (2000). Does schooling cause growth? American Economic Review 90, 1160–1183.

Borowiecki, M., B. Dachs, D. Hanzl-Weiss, S. Kinkel, J. Poschl, M. Sass, T. C. Schmall, R. Stehrer, and A. Szalavetz (2012). Global value chains and the eu industry. WIIW Research Report (383).

Bouet, A., Y. Decreux, L. Fontagné, S. Jean, and D. Laborde (2008). Assessing applied protection across the world. Review of International Economics 16(5), 850–863.

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

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. Journal of International Economics 82, 152–167.

Costinot, A. (2009). An elementary theory of comparative advantage. Econometrica 77(4).

Costinot, A., D. Donaldson, and I. Komunjer (2012). What goods do countries trade? a quantitative exploration of ricardo's ideas. Review of Economic Studies 79, 581–608.

Daudin, G., C. Rifflart, and D. Schweisguth (2011). Who produces for whom in the world economy? Canadian Journal of Economics 44(4), 1403–1437.

Deardorff, A. (2004). Local comparative advantage: trade costs and the pattern of trade. University of Michigan Discussion papers (500).

Eaton, J. and S. Kortum (1999). International technology diffusion: theory and measurement. International Economic Review 40(3), 537–570.

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., B. Neiman, S. Kortum, and J. Romalis (2011). Trade and the global recession. NBER Working Paper Series (16666).

Harrigan, J. (2010). Airplanes and comparative advantage. Journal of International Economics 82, 181–194.

Harrigan, J. and C. Evans (2005). Distance, time, and specialization: Lean retailing in general equilibrium. The American Economic Review 95(1).

Hausman, J. (2001). Mismeasured variables in econometric analysis: Problems from the right and problems from the left. Journal of Economic Perspectives 15(4).

Hummels, D., J. Ishii, and K.-M. Yi (2001). The nature and growth of vertical specialization in world trade. Journal of International Economics 54(1), 75–96.

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

Johnson, R. C. and G. Noguera (2012b). Fragmentation and trade in value added over four decades. NBER Working Paper Series (18186).

Johnson, R. C. and G. Noguera (2012c). Proximity and production fragmentation. American Economic Review: Papers and Proceedings 102(3), 407–411.

Kee, H. L., A. Nicita, and M. Olarreaga (2009). Estimating trade restrictiveness indices. The Economic Journal 119(1), 172–199.

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

Lewis, J. B. and D. Linzer (2005). Estimating regression models in which the dependent variable is based on estimates. Political Analysis 13, 345–364.

Limão, N. and A. J. Venables (2002). Geographical disadvantage: a heckscher-ohlin-von thunen model of international specialization. Journal of International Economics 58, 239–263.

Mayer, T. and S. Zignago (2011). Notes on cepii's distance measures (geodist). CEPII working paper series (25).

Parro, F. (2013). Capital-skill complementarity and the skill premium in a quantitative model of trade. American Economic Journal: Macroeconomics 5(2), 72–117.

Shikher, S. (2013). Determinants of trade and specialization in the oecd countries. Economic Enquiry.

Stehrer, R. (2012). Trade in value added and the value added in trade. WIIW working paper series (81).

Timmer, M. (Ed.) (2012, April). The World Input Output Database (WIOD): Contents, Sources, and Methods. Unpublished manuscript.

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

Yi, K.-M. (2010). Can multistage production explain the home bias in trade? American Economic Review 100(1), 364–393.

A List of countries and sectors

Tab.10 contains the list of the main EU-15 trade partners in 1995-2009. In blue are the six countries dropped from the sample because they are absent from the WIOD database.

Table 10: List of main EU-15 trading partners

ID	Country	Type	ID	Country	Type
AT	Austria	intra-eu15	PL	Poland	ceec
BE	Belgium-Luxembourg	intra-eu15	RO	Romania	ceec
DK	Denmark	intra-eu15	SK	Slovakia	ceec
FI	Finland	intra-eu15	SI	Slovenia	ceec
FR	France	intra-eu15	TR	Turkey	ceec
DE	Germany	intra-eu15	CA	Canada	other devpd
GR	Greece	intra-eu15	JP	Japan	other devpd
IE	Ireland	intra-eu15	KR	Korea	other devpd
IT	Italy	intra-eu15	NO	Norway	other devpd
NL	Netherlands	intra-eu15	CH	Switzerland	other devpd
PT	Portugal	intra-eu15	US	USA	other devpd
ES	Spain	intra-eu15	BR	Brazil	other emerging
SW	Sweden	intra-eu15	CN	China	other emerging
GB	United Kingdom	intra-eu15	IN	India	other emerging
BG	Bulgaria	ceec	ID	Indonesia	other emerging
HR	Croatia	ceec	MY	Malaysia	other emerging
CZ	Czech Republic	ceec	MX	Mexico	other emerging
EE	Estonia	ceec	RU	Russia	other emerging
HU	Hungary	ceec	SG	Singapore	other emerging
LV	Latvia	ceec	TW	Taiwan	other emerging
LT	Lithuania	ceec	TH	Thailand	other emerging

The indicator of proximity endowment can be constructed for all countries because data on bilateral distance is widely available. Tab.11 provides details on the sample of countries for which we can also construct domestic supply based proximity indicators (DSBP) derived in sec.??. In black are the countries for which output data is consistent with trade data, and for which we can construct the DSBP indicator. In blue are the countries for which output data is either lacking or inconsistent with trade data. For such countries we can only construct the indicator of foreign supply based proximity which only takes into account the distribution of market shares across foreign suppliers. In red is the only country for which no bilateral imports data is available (Taiwan). In the empirical analysis we impute Taiwan's proximity characteristic using the indicator of proximity endowment and the estimated relationship of DSBP with proximity endowment. We do not construct DSBP indicators for countries in blue using this relationship because these countries drop out anyway in the second step of the estimation.[61]

[61]Either measured TFP data is of poor quality or sectoral R&D data is missing, or both.

Table 11: Proximity indicators in the sample

ID	Country	Type	ID	Country	Type
AT	Austria	intra-eu15	PL	Poland	ceec
BE	*Belgium-Luxembourg*	intra-eu15	RO	Romania	ceec
DK	Denmark	intra-eu15	SK	Slovakia	ceec
FI	Finland	intra-eu15	SI	Slovenia	ceec
FR	France	intra-eu15	TR	Turkey	ceec
DE	Germany	intra-eu15	CA	Canada	other devpd
GR	Greece	intra-eu15	JP	Japan	other devpd
IE	Ireland	intra-eu15	KR	Korea	other devpd
IT	Italy	intra-eu15	NO	*Norway*	other devpd
NL	*Netherlands*	intra-eu15	CH	*Switzerland*	other devpd
PT	Portugal	intra-eu15	US	USA	other devpd
ES	Spain	intra-eu15	BR	Brazil	other emerging
SW	Sweden	intra-eu15	CN	China	other emerging
GB	United Kingdom	intra-eu15	IN	India	other emerging
BG	Bulgaria	ceec	ID	*Indonesia*	other emerging
HR	*Croatia*	ceec	MY	*Malaysia*	other emerging
CZ	Czech Republic	ceec	MX	Mexico	other emerging
EE	Estonia	ceec	RU	Russia	other emerging
HU	Hungary	ceec	SG	*Singapore*	other emerging
LV	Latvia	ceec	TW	*Taiwan*	other emerging
LT	Lithuania	ceec	TH	*Thailand*	other emerging

Table 12: Input Intensity in Production and Trade (EU-15)

ID	Desc	ii_{prod}'95	ii_{prod}'09	ii_{trade}'95	ii_{trade}'09
15 – 16	Manuf. Food-Tobacco	.75	.76	.22	.22
17 – 18	Manuf. Textile-Clothes	.66	.69	.25	.14
19	Manuf. Leather	.71	.70	.14	.05
20	Manuf. Wood (no Furniture)	.66	.70	.79	.77
21 – 22	Manuf. Paper and Publishing (Media)	.62	.65	.74	.76
24	Manuf. Chemicals-Pharmaceuticals	.66	.71	.84	.67
25	Manuf. Rubber-Plastic	.63	.68	.70	.67
26	Manuf. Non-Metallic Products	.59	.65	.78	.81
27 – 28	Manuf. Metal Products (no Machinery)	.65	.70	.90	.87
29	Manuf. Machinery-Equipment	.63	.66	.38	.44
30 – 33	Manuf. 'Other' Equipment	.64	.68	.48	.33
34 – 35	Manuf. Transport Equipment	.73	.79	.42	.41
36 – 37	Misc. Manuf. (furniture, toys,...)	.64	.67	.24	.17

ii_{trade} (ii_{prod}) is input intensity in trade (resp., production). Input intensity in production is the income share of inputs in gross output constructed for the EU-15 aggregate. Input intensity in trade is the value share of processed intermediate inputs in total imports constructed for EU-15 trade with non-EU15 (excl. intra-EU). Primary and processed inputs in the oil industry are excluded from calculation of ii_{trade}. 'Other' is manufacturing of computers, electrical, audiovisual, communications, measurement, and precision equipment.

B Use of inputs and the level of aggregation

Tab.13 provides information on the share of inputs in production (ii_{prod}) for 22 manufacturing sectors which correspond to the 2-digit level of the ISIC Rev.3 classification. We focus on the main economies of the EU-15 in 1996-2006, i.e. Germany, France, UK, Italy, Spain, Finland, and Sweden, because the UNIDO INDSTAT database provides consistent time series data on gross output ($PROD$) and value added (VA) at the 4-digit level in the ISIC rev.3

for these countries.[62]

The data is judged consistent in a given year if output data is consistent with data on total exports and imports. This is the case when consumption of varieties produced domestically is a positive fraction of total sectoral expenditure, and this fraction is relatively persistent in adjacent years. The income share of inputs in sectoral production is computed as $ii_{prod} = 1 - VA/PROD$. Country specific input intensity indicators are aggregated into an EU7 indicator according to countries' share in total sectoral output.[63]

Table 13: Inputs' share in gross output (EU7), 2-digit level (ISIC Rev.3)

ID	Description	Ranking	Share 1996	Share 2006
15	Manufacture of food products and beverages	3	.77	.78
16	Manufacture of tobacco products	3	.89	.88
17	Manufacture of textiles	2	.69	.72
18	Manufacture of wearing apparel	2	.66	.73
19	Tanning and dressing of leather	2	.71	.75
20	Manufacture of wood products, except furniture	2	.69	.71
21	Manufacture of pulp, paper, paper products	2	.65	.71
22	Publishing, printing, reproduction of media	1	.62	.65
23	Manufacture of coke, refined petroleum products, nuclear fuel	3	.90	.92
24	Manufacture of chemicals and chemical products	2	.68	.71
25	Manufacture of rubber & plastic products	1	.63	.70
26	Manufacture of other non-metallic mineral products	1	.61	.67
27	Manufacture of basic metals	3	.74	.79
28	Manufacture of fabricated metal products, except machinery	1	.62	.66
29	Manufacture of machinery & equipment	2	.65	.68
30	Manufacture of office machinery & computers	3	.75	.75
31	Manufacture of electrical machinery & apparatus	1	.64	.69
32	Manufacture of radio, TV, communication equipment	2	.69	.71
33	Manufacture of medical, precision & optical instruments	1	.59	.59
34	Manufacture of motor vehicles & trailers	3	.77	.79
35	Manufacture of other transport equipment	2	.72	.71
36	Manufacture of furniture	2	.66	.69

In column 3, '1' corresponds to ii_{prod} below iqr, '2' corresponds to ii_{prod} in the iqr, and '3' corresponds to ii_{prod} above iqr for the EU7 aggregate.

The objective of this section is to check whether the assumption of the model that input intensity is a sector specific characteristic common across countries and invariant overtime is tenable. Further, we want to verify how sensitive are the indicators of sectoral input intensity to the level of aggregation of the data. Tab.13 illustrates that input intensity in production is a persistent characteristic of the sector.[64] Persistence is also verified at the

[62] There is no data on Germany in 1996-1997, and no data on UK in 2006.

[63] There are two bottlenecks in extending the dataset to more years or countries: availability of VA data, consistency of domestic absorption measures overtime.

[64] At the 4-digit level for the EU7 as an aggregate, the correlation coefficient between input intensity measures in 1996 and 2001 is .91, and .88 with 2006. At the 2-digit level, the corresponding measures are .97 and .96.

level of individual countries, and reinforced for individual countries when we work at the 4-digit level.

This table also documents substantial intersectoral variability in the income share of inputs in production. In particular, if the indicators are computed for the EU7 as an aggregate, within-sectoral input intensity variability is lower than intersectoral variability. Within-sectoral interquartile range varies between .01 and .05 for the 21 manufacturing sectors while the standard deviation of sector specific median ii_{prod} indicators is .06.

Table 14 and fig.12 show that the share of inputs in production is relatively homogeneous across these 7 countries. In particular, it is possible to establish an ordinal ranking of sectors according to their input intensity which is relatively stable across countries and overtime. This pattern is particularly strong for the 5 biggest EU15 economies which are Germany, France, Italy, Spain, and the UK for which the correlation coefficient of sectoral ii_{prod} varies between .84-.93.

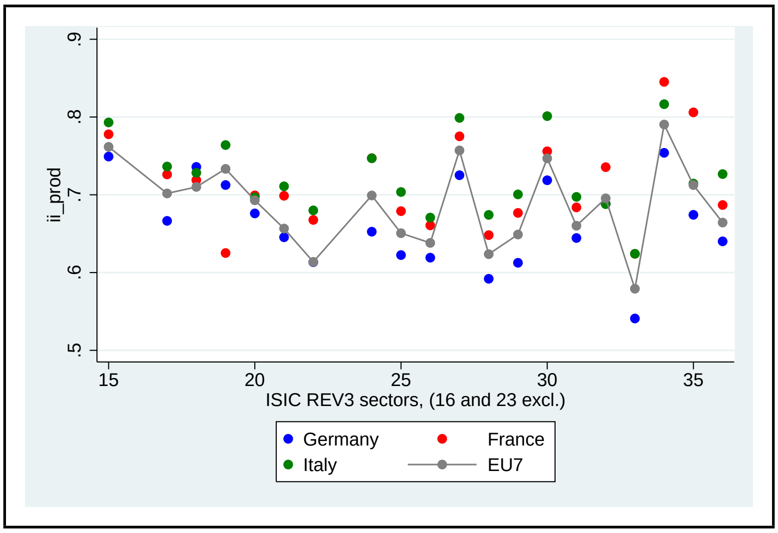
Table 14: Input intensity in production in EU7 in 1996-2006

sector	EU7	GERMANY	FRANCE	UK	ITALY	SPAIN	FINLAND	SWEDEN
15	0.76	0.75	0.78	0.69	0.79	0.79	0.74	0.75
17	0.70	0.67	0.73	0.61	0.74	0.70	0.58	0.63
18	0.71	0.74	0.72	0.58	0.73	0.68	0.62	0.69
19	0.73	0.71	0.63	0.58	0.76	0.75	0.61	0.67
20	0.69	0.68	0.70	0.61	0.70	0.70	0.76	0.74
21	0.66	0.65	0.70	0.59	0.71	0.66	0.68	0.66
22	0.61	0.61	0.67	0.53	0.68	0.61	0.59	0.67
23	0.92	0.94	0.93	0.92	0.93	0.83	0.86	0.71
24	0.70	0.65	0.75	0.62	0.75	0.71	0.67	0.54
25	0.65	0.62	0.68	0.60	0.70	0.66	0.59	0.63
26	0.64	0.62	0.66	0.57	0.67	0.64	0.61	0.64
27	0.76	0.73	0.78	0.75	0.80	0.76	0.75	0.72
28	0.62	0.59	0.65	0.57	0.67	0.65	0.62	0.61
29	0.65	0.61	0.68	0.61	0.70	0.65	0.69	0.65
30	0.75	0.72	0.76	0.76	0.80	0.78	0.80	0.67
31	0.66	0.64	0.68	0.61	0.70	0.70	0.67	0.68
32	0.70	0.69	0.74	0.66	0.69	0.74	0.65	0.77
33	0.58	0.54	0.62	0.54	0.62	0.61	0.60	0.59
34	0.79	0.75	0.85	0.77	0.82	0.81	0.63	0.77
35	0.71	0.67	0.81	0.59	0.71	0.69	0.71	0.64
36	0.66	0.64	0.69	0.59	0.73	0.66	0.61	0.66

The table gives the median value of ii_{prod} by sector and country in 1996-2006.

There are two main drawbacks in working with these ii_{prod} indicators. First, intersectoral variability in the share of processed inputs in production is blurred by inclusion of raw inputs. This caveat is present at any level of disaggregation. Second, the stability of the input intensity measure across countries is blurred by the variability of the product mix which countries

Figure 12: Input intensity in production: ordinal rankings across EU7



produce within any given 2-digit sector. For the EU7 taken as an aggregate, the problem disappears because the measure is dominated by countries such as Germany, France, and UK for which within-sectoral variability is relatively low and persistence overtime very high.

But at the individual country level, variability in within-sector product mixes quickly becomes a problem, both in terms of sectoral homogeneity and stability overtime. Indeed, even if the hypothesis of ii_{prod} being a sector-specific characteristic invariant across countries may hold in terms of ordinal sector rankings, it fits the data better at the level of 121 manufacturing sectors at the 4-digit level.

To illustrate we look at a simple indicator. At the 2-digit level, the correlation coefficient between country-specific input intensity measures in 1998 varies between .82 for Germany and Italy and .34 for Germany and Sweden. In 2005 the correlation coefficient for Germany-Italy is virtually unchanged while for Germany and Sweden it jumps up to .83.[65]

At the 4-digit level this problem is alleviated. The correlation coefficient between country-specific vectors of ii_{prod} varies between .76 (.77) for Germany-France and .64 (.69) for Germany-Sweden in 1998 (2005). Cross-country and overtime variability are strongly reduced. Furthermore, ii_{prod} indicators for Germany are more strongly correlated to France than to Italy,

[65]This example voluntarily looks at extremes: variability is much lower for the five main EU-15 economies as illustrated by the Italy-Germany measure.

contrary to the input intensity measure constructed at the 2-digit level.

We conclude that available evidence points to differences in within-sectoral specialization across EU-15 countries rather than to intrinsic differences in production functions. This may lead to difficulties in using ii_{prod} measures constructed at the 2-digit level in empirical analysis. Nonetheless, available empirical evidence at the 4-digit level is consistent with the hypothesis of input intensity in production being a sector-specific characteristic invariant across countries. This assumption is built into the model.

C Revealed Comparative Advantage

This appendix provides descriptive statistics on estimated exporter-sector dummies for 20 manufacturing sectors which correspond to the 2-digit level in the NACE 1.1 (ISIC Rev.3) classification. In Costinot et al. (2012), this would correspond to fundamental sectoral productivity estimates while in this paper the sectoral cost comprises technology, factor cost, input cost, and export-side trade cost components specific to this sector and exporter.

The fixed effects approach in the sectoral exports regression is implemented at the bilateral level as well as by partner types. The four partner types are: 'intra-eu15' which includes the 15 EU members by 1995, 'ceec' which includes 12 European emerging economies,[66] 'other emerging' which includes the main non-EU emerging economies such as Brazil, China, and India, and 'other developed' which includes the main non-EU members of the OECD such as the United States and Japan.[67]

Tables 15 and 16 give estimated values of relative sectoral exporter-sector dummies in 1995 and 2010 by trade partner group. In Costinot et al. (2012) this would correspond to fundamental sectoral productivity relatively to the processed foods' sector (15) and the group of 'other developed countries' where the exponential of the estimated dummy is normalized to 1. In the model used in this paper, the dummy contains technology, wage, proximity, and export-side trade cost components. The variance in sectoral costs has increased overtime for the group of non-European emerging economies while it has remained stable for the EU-15 and the CEECs.

Tables 17 and 18 provide estimates of relative exporter specific sectoral production costs for a subset of countries in the bilateral estimation. Cost

[66] This group is constituted by the 10 Central and Eastern European countries which have become members of the European Union in 2004 and 2007, Turkey which has a customs union with the EU since 1996, and Croatia which is set to become the 28th member of the EU on Jan.1, 2013. Cyprus and Malta are not included in the analysis.

[67] The grouping of 42 main EU-15 partners in 4 types is detailed in table (10).

Table 15: Exporter-sector production costs by partner type in 1995

Sector	ceec	non-EU emerging	intra-eu15	other devpd
15	1	1	1	1
17	1.148	1.035	0.972	1
18	1.397	1.307	1.086	1
19	1.221	1.344	1.047	1
20	1.135	1.061	0.885	1
21	0.852	0.772	0.926	1
22	0.73	0.73	0.87	1
24	0.872	0.784	0.879	1
25	0.929	0.869	0.919	1
26	1.174	0.904	1	1
27	1.089	0.934	0.931	1
28	1	0.917	0.939	1
29	0.847	0.736	0.849	1
30	0.576	0.91	0.816	1
31	0.895	0.856	0.84	1
32	0.762	0.892	0.791	1
33	0.626	0.723	0.741	1
34	0.876	0.663	0.932	1
35	0.664	0.697	0.693	1
36	0.975	1.063	0.873	1

The reported relative cost corresponds to $\left[e^{fe^k_{i,t}}\right]^{1/\theta}$.

Table 16: Exporter-sector production costs by partner type in 2010

Sector	ceec	non-EU emerging	intra-eu15	other devpd
15	1	1	1	1
17	1.217	1.274	0.985	1
18	1.471	1.619	1.201	1
19	1.211	1.606	1.194	1
20	1.149	1.102	0.917	1
21	0.919	0.878	0.921	1
22	0.923	0.963	0.949	1
24	0.776	0.813	0.84	1
25	1.036	0.998	0.924	1
26	1.102	1.097	0.968	1
27	0.915	0.922	0.87	1
28	1.061	1.066	0.924	1
29	0.934	0.912	0.843	1
30	1.093	1.154	0.939	1
31	1.036	1.034	0.877	1
32	1.034	1.1	0.839	1
33	0.73	0.814	0.748	1
34	1.109	0.819	0.962	1
35	0.651	0.792	0.671	1
36	1.113	1.246	0.927	1

The reported relative cost corresponds to $\left[e^{fe^k_{i,t}}\right]^{1/\theta}$.

is normalized to 1 in the US in all sectors, and in the food sector for all countries. There is substantial heterogeneity in RCA rankings within each partner group.

Table 17: Exporter-sector production costs for a subsample of countries in 1995

sector	CZECH	HUNGARY	CHINA	MEXICO	CANADA	JAPAN	GERMANY	ITALY
type	*devpd*	*emerging*	*ceec*	*eu15*	*ceec*	*eu15*	*devpd*	*emerging*
15	1	1	1	1	1	1	1	1
17	1.446	1.07	1.42	0.969	0.925	1.59	1.131	1.255
18	1.378	1.174	1.846	0.784	0.936	1.188	1.138	1.282
19	1.465	1.083	1.856	0.956	0.811	1.204	1.026	1.43
20	1.2	0.806	1.129	0.534	1.182	0.779	0.884	0.813
21	1.142	0.732	0.809	0.561	1.298	1.218	1.054	0.919
22	0.909	0.721	0.888	0.545	0.928	1.226	0.881	0.788
24	1.042	0.885	0.986	0.895	0.808	1.584	1.007	0.878
25	1.174	0.972	1.191	0.734	0.837	1.751	1.072	1.044
26	1.568	1.045	1.249	0.908	0.97	1.674	1.122	1.177
27	1.298	1.11	0.957	0.996	1.192	1.424	1.15	1.078
28	1.289	0.995	1.311	0.745	0.912	1.651	1.137	1.113
29	1.141	0.848	0.99	0.653	0.872	1.771	1.014	1.018
30	0.668	0.468	0.999	0.717	0.811	1.602	0.827	0.727
31	1.063	0.975	1.175	0.815	0.822	1.811	1.02	0.907
32	0.753	0.877	1.113	0.78	0.856	1.886	0.92	0.766
33	0.739	0.601	0.957	0.657	0.788	1.595	0.848	0.731
34	1.302	1.031	0.629	0.647	0.89	2.32	1.296	1.094
35	0.771	0.408	0.669	0.548	0.885	1.386	0.676	0.671
36	1.261	0.862	1.542	0.954	0.912	1.73	0.972	1.096

Table 19 provides information on the persistence of sectoral rankings overtime at the level of 2-digit sectors for each group of partners. The vector of correlation coefficients for cost rankings in 1995 with subsequent years is reported for each partner type. Cost rankings are strongly persistent in all country groups with the exception of the CEECs. On the other hand, non-European emerging economies are characterized by very persistent cost rankings.

Table 20 provides information on the persistence of sectoral cost rankings overtime at the level of 2-digit sectors for a subsample of countries. The table reports the vector of correlation coefficients for cost rankings in 1995 with subsequent years for 2 countries by partner type.

There is strong variability in relative sectoral production costs for the CEECs in 1995-2010. For the group of non European emerging economies results for China and Mexico are reported. China is a typical country in the group of non-European emerging economies in the sense that the ranking of its sectoral costs if very persistent overtime. Mexico is the only non-EU emerging economy which shows substantial changes in revealed comparative

Table 18: Exporter-sector production costs for a subsample of countries in 2010

sector	CZECH	HUNGARY	CHINA	MEXICO	CANADA	JAPAN	GERMANY	ITALY
type	*devpd*	*emerging*	*ceec*	*eu15*	*ceec*	*eu15*	*devpd*	*emerging*
15	1	1	1	1	1	1	1	1
17	1.218	1.013	1.621	0.766	0.858	1.474	1.022	1.148
18	1.289	1.024	1.972	0.927	0.986	1.116	1.121	1.265
19	1.204	0.925	1.937	0.982	0.785	1.084	1.044	1.361
20	0.909	0.927	1.202	0.471	1.055	0.741	0.89	0.758
21	1.005	0.834	0.957	0.631	0.958	1.128	0.918	0.849
22	1.023	0.763	1.099	0.632	0.887	1.186	0.889	0.812
24	0.826	0.762	0.867	0.783	0.798	1.34	0.839	0.777
25	1.145	1.032	1.204	0.746	0.828	1.574	0.969	0.947
26	1.114	0.965	1.29	0.845	0.793	1.394	0.937	0.974
27	1.05	0.869	1.06	0.765	0.931	1.431	0.985	0.958
28	1.18	0.962	1.312	0.708	0.874	1.469	0.989	0.967
29	0.999	0.936	1.085	0.778	0.846	1.559	0.868	0.891
30	1.354	1.084	1.333	0.922	0.87	1.451	0.89	0.666
31	1.129	1.067	1.261	0.943	0.846	1.542	0.937	0.862
32	1.138	1.23	1.347	1.043	0.879	1.562	0.86	0.682
33	0.747	0.757	0.895	0.863	0.784	1.361	0.733	0.646
34	1.553	1.267	1.011	0.945	0.851	2.14	1.242	1.033
35	0.643	0.527	0.829	0.5	0.902	1.18	0.647	0.637
36	1.16	0.888	1.509	0.841	0.913	1.434	0.929	0.956

Table 19: Rankings' autocorrelation by partner type

Year	**eu15**	**ceec**	**emerging**
1998	0.98	0.95	0.99
1999	0.97	0.96	0.99
2000	0.96	0.94	0.99
2001	0.96	0.93	0.99
2002	0.95	0.93	0.97
2003	0.95	0.91	0.98
2004	0.95	0.89	0.98
2005	0.94	0.85	0.99
2006	0.94	0.83	0.98
2007	0.92	0.8	0.98
2008	0.92	0.75	0.97
2009	0.9	0.72	0.95
2010	0.9	0.71	0.95

advantage on EU-15 markets at the level of 2-digit sectors. The period of study corresponds to the implementation of NAFTA, i.e. Mexico undergoes a process of regional integration with the Canadian and US economies which may impact the ranking of sectoral production costs for Mexico on world markets.

Table 20 indicates strong variability in revealed relative sectoral produc-

Table 20: Rankings' correlation for subsample of countries

sector	CANADA	CHINA	CZECH	GERMANY	HUNGARY	ITALY	JAPAN	MEXICO
type	*devpd*	*emerging*	*ceec*	*eu15*	*ceec*	*eu15*	*devpd*	*emerging*
1998	0.97	0.99	0.94	0.98	0.84	0.99	0.98	0.91
1999	0.96	0.99	0.93	0.97	0.82	0.99	0.98	0.88
2000	0.96	0.98	0.93	0.95	0.77	0.99	0.98	0.70
2001	0.96	0.98	0.91	0.94	0.78	0.98	0.98	0.68
2002	0.93	0.98	0.89	0.92	0.80	0.98	0.96	0.59
2003	0.94	0.97	0.86	0.92	0.75	0.98	0.98	0.56
2004	0.92	0.96	0.80	0.92	0.67	0.99	0.97	0.60
2005	0.91	0.96	0.71	0.91	0.60	0.99	0.98	0.67
2006	0.89	0.96	0.68	0.92	0.57	0.99	0.97	0.61
2007	0.94	0.95	0.64	0.89	0.53	0.98	0.98	0.64
2008	0.92	0.95	0.60	0.89	0.51	0.98	0.98	0.56
2009	0.87	0.94	0.56	0.89	0.54	0.98	0.97	0.55
2010	0.84	0.92	0.51	0.88	0.45	0.97	0.97	0.62

tion costs for the CEECs in 1995-2010. One explanation could be changes in data reporting thresholds for intra-EU trade which have impacted 3/4 of the CEEC group in 2004 and 1/6 of the group in 2007 upon entry in the European Union. As shown in table 21, a simple regression of cost correlation coefficients at the country level confirms there is a break in 2004 for the CEECs. However, as shown in col.3, the most robust feature of the data is the stronger variability in rankings for the CEECs relatively to other partner types over the whole period.

Tables 22 and 23 report annual ratios of revealed relative sectoral costs for the CEECs relatively to respectively non-European emerging economies and the countries of the EU-15. Sectors are ranked according to RCA in 1995. The theoretical underpinning of these tables is the prediction of the Ricardian model that the ratio of observed relative bilateral exports is increasing in the gap of fundamental sectoral costs, controlling for pair and destination-sector fixed effects.

These tables show that the pattern of comparative advantage for the CEECs on EU-15 markets varies across partner groups. To give an example, the CEECs have a persistent comparative advantage in the wearing apparel industry (18) relatively to the EU-15, but a consistent disadvantage relatively to non-European emerging economies. Another example is the motor vehicles industry (34) in which the CEECs have a persistent strong comparative advantage relatively to non-European emerging economies while this industry is not in the top 5 for the CEECs relatively to the EU-15.

Several features of the data stand out. First, there is evidence of substantial modification in cost rankings as well as in the RCA pattern for the

Table 21: **Persistence of cost rankings overtime**

depvar: Correlation coefficient	(1)	(2)	(3)
year	-0.0138***	-0.0076***	-0.0085***
	(0.0032)	(0.0017)	(0.0033)
*year*intraeu15*		0.0006	0.0003
		(0.0019)	(0.0038)
*entry*ceec*	-0.0103*		-0.000
	(0.0058)		(0.000)
*year*ceec*		-0.0104***	-0.0085**
		(0.0020)	(0.0039)
*year*emerg*		-0.0003	-0.0013
		(0.0020)	(0.0040)
*entry*intraeu15*			0.000
			(0.000)
*entry*emerg*			0.000
			(0.000)
Entry FE	YES	NO	YES
Type FE		YES	YES
Observations	192	656	656
R-squared	0.484	0.446	0.448

The dependent variable is the correlation coefficient of the sectoral production cost in 1995 with every other year.
'entry' is a dummy equal to 1 in 2004-2010. *** p<0.01, ** p<0.05, * p<0.1.
'year' picks up the time trend. In (1) the regression is on the CEEC sample.
In (2)-(3), the regression is on the full sample of EU-15 trade partners.

CEECs on EU-15 markets relatively to all partner groups, and in particular relatively to non-EU emerging economies. Second, the variance of cost ratios at the intersectoral level is reduced overtime for CEECs relatively to all partner groups while the gap in extreme values is not. The max-min gap increases for CEECs relatively to non-EU emerging economies while it is reduced relatively to the EU-15 country group.

Table 24 reports the pattern of revealed comparative advantage for the EU-15 relatively to non-EU emerging economies.

D Addenda on Estimation of the Model

D.1 R&D data: imputation of missing observations

Data on sectoral R&D activity is taken from the 2011 edition of OECD ANBERD for all countries but China.[68] For most countries, a fair amount of

[68]Seven countries are dropped from the analysis for lack of R&D data. These are Bulgaria, Brazil, India, Indonesia, Lithuania, Latvia, and Russia.

Table 22: Specialization pattern of CEECs to non-EU emerging economies

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
34	1.32	1.36	1.36	1.39	1.36	1.46	1.43	1.47	1.48	1.50	1.45	1.44	1.42	1.40	1.36	1.36
26	1.30	1.30	1.32	1.29	1.24	1.24	1.24	1.24	1.21	1.16	1.11	1.08	1.04	1.04	1.02	1.00
23	1.28	1.50	1.32	1.00	1.17	1.09	0.99	0.94	0.94	0.91	0.99	0.98	1.09	1.09	1.03	1.14
27	1.17	1.16	1.19	1.19	1.19	1.17	1.18	1.16	1.12	1.10	1.05	1.04	1.02	1.02	1.03	0.99
29	1.15	1.15	1.17	1.18	1.14	1.15	1.14	1.15	1.13	1.10	1.08	1.08	1.08	1.08	1.05	1.02
24	1.11	1.12	1.12	1.10	1.06	1.10	1.06	1.05	1.02	1.00	0.97	0.97	0.96	0.97	0.97	0.95
17	1.11	1.11	1.13	1.15	1.13	1.15	1.15	1.14	1.14	1.12	1.07	1.07	1.05	1.02	0.98	0.96
21	1.10	1.16	1.21	1.18	1.15	1.18	1.16	1.15	1.14	1.14	1.11	1.12	1.11	1.09	1.07	1.05
28	1.09	1.08	1.10	1.11	1.08	1.12	1.11	1.11	1.12	1.08	1.06	1.05	1.05	1.04	1.03	1.00
25	1.07	1.07	1.08	1.09	1.07	1.10	1.11	1.12	1.13	1.10	1.08	1.07	1.07	1.06	1.05	1.04
20	1.07	1.10	1.14	1.14	1.12	1.15	1.13	1.14	1.13	1.10	1.07	1.07	1.06	1.04	1.04	1.04
18	1.07	1.09	1.11	1.13	1.11	1.13	1.13	1.10	1.09	1.07	1.01	0.98	0.98	0.95	0.92	0.91
31	1.05	1.07	1.09	1.09	1.06	1.08	1.10	1.10	1.09	1.07	1.05	1.06	1.06	1.04	1.02	1.00
22	1.00	1.03	1.05	1.04	1.00	1.04	1.06	1.05	1.06	1.01	1.01	1.05	1.02	0.98	0.98	0.96
35	0.95	0.98	0.95	0.93	0.92	0.96	0.97	1.06	0.96	1.00	0.90	0.93	0.91	0.90	0.88	0.82
36	0.92	0.92	0.95	0.96	0.94	0.96	0.97	0.98	0.98	0.97	0.94	0.94	0.93	0.93	0.91	0.89
19	0.91	0.88	0.90	0.90	0.88	0.90	0.90	0.91	0.91	0.88	0.83	0.81	0.80	0.78	0.76	0.75
33	0.87	0.87	0.90	0.91	0.91	0.94	0.93	0.94	0.95	0.94	0.91	0.91	0.92	0.92	0.91	0.90
32	0.86	0.90	0.98	1.03	1.01	1.04	1.05	1.03	1.01	1.03	0.99	0.98	1.00	0.99	0.98	0.94
30	0.63	0.66	0.74	0.80	0.78	0.83	0.81	0.80	0.82	0.82	0.90	0.91	0.95	0.96	0.97	0.95

Table 23: Specialization pattern of CEECs to intra-EU15

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
18	1.29	1.31	1.35	1.38	1.39	1.42	1.42	1.42	1.41	1.38	1.32	1.30	1.28	1.26	1.23	1.23
20	1.28	1.30	1.35	1.38	1.39	1.42	1.40	1.42	1.43	1.36	1.32	1.31	1.28	1.25	1.24	1.25
17	1.18	1.19	1.23	1.27	1.27	1.30	1.31	1.33	1.32	1.31	1.28	1.28	1.27	1.26	1.24	1.24
26	1.18	1.18	1.21	1.23	1.23	1.26	1.25	1.27	1.26	1.23	1.19	1.18	1.16	1.13	1.13	1.14
27	1.17	1.13	1.16	1.20	1.18	1.21	1.20	1.20	1.19	1.18	1.13	1.14	1.12	1.10	1.07	1.05
19	1.17	1.13	1.15	1.15	1.14	1.18	1.18	1.19	1.20	1.17	1.12	1.09	1.06	1.04	1.01	1.01
36	1.12	1.12	1.15	1.17	1.17	1.21	1.22	1.27	1.29	1.27	1.24	1.25	1.23	1.22	1.19	1.20
31	1.07	1.10	1.13	1.16	1.16	1.21	1.22	1.25	1.25	1.23	1.21	1.22	1.22	1.20	1.18	1.18
28	1.06	1.06	1.10	1.13	1.12	1.18	1.18	1.20	1.23	1.18	1.17	1.17	1.17	1.17	1.16	1.15
25	1.01	1.01	1.03	1.06	1.07	1.11	1.12	1.15	1.16	1.13	1.11	1.12	1.12	1.11	1.11	1.12
29	1.00	1.00	1.03	1.05	1.05	1.08	1.09	1.13	1.14	1.12	1.11	1.11	1.11	1.11	1.10	1.11
24	0.99	0.98	0.98	0.98	0.96	1.00	0.97	0.97	0.96	0.93	0.91	0.92	0.91	0.93	0.91	0.92
32	0.96	1.00	1.08	1.14	1.12	1.18	1.19	1.22	1.25	1.27	1.23	1.23	1.24	1.26	1.27	1.23
35	0.96	0.99	0.97	1.01	0.99	1.06	1.04	1.13	1.10	1.13	1.06	1.06	1.02	1.02	1.03	0.97
34	0.94	0.96	1.00	1.05	1.05	1.11	1.12	1.14	1.16	1.15	1.14	1.14	1.14	1.14	1.15	1.15
21	0.92	0.93	0.95	0.97	0.97	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
33	0.85	0.85	0.89	0.92	0.93	0.96	0.96	0.98	0.99	0.98	0.95	0.97	0.97	0.98	0.97	0.98
22	0.84	0.87	0.88	0.89	0.88	0.90	0.92	0.92	0.94	0.91	0.91	0.96	0.98	0.96	0.98	0.97
30	0.71	0.74	0.81	0.89	0.87	0.92	0.93	0.93	0.98	0.98	1.05	1.08	1.10	1.14	1.16	1.16

observations is missing at the sectoral level in 1995-2009. For such observations the data is imputed.

When data is missing at the 2-digit level but R&D activity is reported at a higher level of aggregation (1-digit or total manufacturing), we impute sectoral data using the share of the sector in total R&D activity in adjacent years. For three countries R&D activity is never reported for certain industries. This is the case of sectors $17-22$ for Great Britain, sectors $17-19$ for Japan, and $19-22$ for Slovakia. For these countries, sectoral data is imputed using data on the share of the sector in R&D activity in the most similar country in terms of the distribution of R&D activity across sectors. For Great Britain sectoral R&D shares are taken from France. For the Slovak

Table 24: Specialization pattern of EU15 to non-EU emerging economies

ID	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
34	1.41	1.41	1.36	1.33	1.30	1.32	1.29	1.28	1.28	1.30	1.27	1.26	1.25	1.23	1.19	1.18
21	1.20	1.25	1.27	1.22	1.18	1.18	1.17	1.15	1.14	1.15	1.11	1.13	1.11	1.10	1.07	1.05
22	1.19	1.18	1.19	1.18	1.15	1.15	1.15	1.15	1.13	1.11	1.11	1.10	1.04	1.03	1.00	0.99
29	1.15	1.14	1.13	1.12	1.08	1.06	1.04	1.02	1.00	0.98	0.98	0.97	0.97	0.97	0.95	0.92
24	1.12	1.15	1.14	1.12	1.11	1.11	1.09	1.09	1.07	1.07	1.07	1.06	1.06	1.05	1.07	1.03
26	1.11	1.10	1.09	1.05	1.01	0.98	0.99	0.97	0.96	0.95	0.93	0.92	0.90	0.91	0.91	0.88
25	1.06	1.06	1.05	1.02	1.00	0.99	0.99	0.98	0.97	0.98	0.97	0.96	0.96	0.95	0.95	0.93
28	1.02	1.02	1.00	0.98	0.96	0.95	0.94	0.93	0.91	0.91	0.90	0.90	0.90	0.89	0.89	0.87
33	1.02	1.02	1.01	0.99	0.98	0.98	0.97	0.95	0.95	0.96	0.95	0.94	0.95	0.94	0.94	0.92
27	1.00	1.03	1.03	0.99	1.01	0.97	0.98	0.97	0.94	0.93	0.93	0.92	0.91	0.93	0.96	0.94
35	0.99	0.98	0.97	0.92	0.93	0.91	0.93	0.94	0.87	0.89	0.85	0.88	0.89	0.88	0.86	0.85
31	0.98	0.98	0.96	0.94	0.91	0.90	0.91	0.88	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.85
17	0.94	0.94	0.92	0.91	0.89	0.89	0.88	0.86	0.86	0.86	0.84	0.83	0.83	0.81	0.79	0.77
30	0.90	0.89	0.90	0.90	0.89	0.90	0.88	0.86	0.84	0.84	0.85	0.85	0.87	0.84	0.83	0.81
32	0.89	0.90	0.91	0.90	0.90	0.88	0.88	0.85	0.81	0.81	0.81	0.79	0.80	0.79	0.77	0.76
20	0.83	0.84	0.84	0.83	0.81	0.81	0.81	0.80	0.80	0.80	0.81	0.82	0.82	0.83	0.84	0.83
18	0.83	0.83	0.82	0.82	0.80	0.79	0.79	0.78	0.77	0.78	0.76	0.76	0.76	0.76	0.75	0.74
36	0.82	0.83	0.82	0.82	0.80	0.79	0.79	0.77	0.76	0.76	0.76	0.76	0.76	0.77	0.76	0.74
19	0.78	0.78	0.79	0.78	0.77	0.77	0.77	0.76	0.76	0.76	0.74	0.75	0.75	0.76	0.75	0.74

Republic (resp. Japan), sectoral shares are taken from the Czech Republic (resp. South Korea).

When R&D data is missing at higher aggregation levels but reported at the sectoral level in the adjacent or the next-to-adjacent year, the missing observation is replaced with the observation from the closest available year. The procedure first fills in lacking observations with data available in $t-1$, then $t-2$. For observations which are still missing after this step, the data is imputed with available information in $t+1$, then $t+2$. This imputation procedure is based on the assumption of persistence in sectoral employment and R&D expenditure in the short run.

Alternatively, we could have interpolated missing observations by pooling the available data and running a fixed effects regression with year, country, and industry fixed effects. We choose not to impute missing observations in this way because this method puts relatively more structure on the common components of the underlying R&D process across countries and sectors.

For China, the data is taken from the Yearbook Database. For 2001-2009, the full set of information is reported for nominal R&D expenditure and full time equivalent of R&D personnel. The series corresponds to R&D activity of large and medium-sized enterprises, and includes all such enterprises whether of mixed, government, or private ownership. The series is reported in China Statistical Yearbooks on Science and Technology in 2002-2007 for 2001-2006, and in the chapter 'Education, Science, and Technology' of annual China Statistical Yearbooks in 2007-2011 for 2006-2010.

For 1995-2000, we use the data on total R&D activity in manufacturing reported in China Statistical Yearbooks on Science and Technology in 1996-2001. We compute average sectoral shares in total R&D activity in

2001-2003, and we distribute the reported totals using this weighting system. We cross-check the quality of this imputation by comparing obtained data with information on R&D activity in high-tech industries in 1995-2000 which is reported at the sectoral level in China Statistics Yearbooks on High Technology Industry (2002, 2003, 2007).

D.2 Results of two stage estimation

In this part of the appendix we report results of the two stage estimation procedure for the benchmark model in which labor and inputs are assumed to be combined in the same way across sectors. We then report first stage results of the model estimated in the core of the paper in which the share of inputs is allowed to be sector-specific.

D.2.1 Benchmark specification: no sector specifics in the production function

Table 25: First stage: Measured TFP ($\zeta^k = \zeta$)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.009***	0.008***		
	(0.003)	(0.003)		
K-stocks	0.233***	0.234***		
	(0.005)	(0.005)		
R&D: expenditure			0.090***	0.092***
			(0.002)	(0.002)
Workforce	-0.196***	-0.196***	-0.089***	-0.093***
	(0.004)	(0.004)	(0.005)	(0.004)
Cons	-1.601***	-1.751***	0.710***	0.665***
	(0.084)	(0.084)	(0.050)	(0.048)
Obs	4196	4196	4833	4833
R^2	0.384	0.398	0.167	0.177
Shea's $pcorr^2$	0.153	0.156	0.0980	0.0974
Angrist-Pischke F-stat	188.1	192.8	181.5	182.3
Angrist-Pischke χ^2	377.7	387.1	182.1	183.0

Depvar is log of measured TFP: $\overline{z}^k_{i,t}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

As a benchmark, we estimate the model assuming that labor and inputs are combined in the same way by all sectors: $\zeta^k = \zeta$. In this case the proximity mechanism plays no role in the pattern of intersectoral specialization because there is no sectoral variation in the input component of production costs. We allow for a sector-specific component of factor costs because

Table 26: First stage: Hourly wages ($\zeta^k = \zeta$)

	(I)	**(II)**	**(III)**	**(IV)**
R&D: personnel	0.083***	0.083***		
	(0.006)	(0.006)		
K-stocks	0.798***	0.807***		
	(0.009)	(0.009)		
R&D: expenditure			0.362***	0.374***
			(0.006)	(0.006)
Workforce	-0.937***	-0.949***	-0.567***	-0.589***
	(0.008)	(0.007)	(0.012)	(0.012)
Cons	-5.412***	-5.424***	1.846***	2.219***
	(0.143)	(0.145)	(0.114)	(0.109)
Obs	4196	4196	4833	4833
R^2	0.732	0.748	0.439	0.463
Shea's $pcorr^2$	0.290	0.292	0.249	0.247
Angrist-Pischke F-stat	1106	1204	552.1	573.0
Angrist-Pischke χ^2	2220	2417	554.0	575.0

Depvar is log of measured hourly wage: $\overline{\nu}^k_{i,t}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

the skill composition of the workforce remains sector-specific. Consequently, the pattern of revealed comparative advantage is determined by technology stocks and labor endowments. In the first stage, measured TFP and hourly wages are instrumented on a common set of instruments. Instruments include year fixed effects because they are included in the second stage of the estimation.

First stage results for the benchmark case are reported in tab.25 and tab.26, together with statistics on instruments' performance. Results are reported for the four specifications used throughout the paper which differ by the set of instruments. In columns (I) and (II), R&D personnel and real capital stocks are used as proxies of sectoral technology stocks, and the estimation window is 1995-2007. In columns (III) and (IV), deflated R&D expenditure is used as a proxy of technology stocks by considering that real expenditure on R&D captures variation in production costs due to technology improvement. The estimation window is 1995-2009. Specifications differ in the measure of sectoral workforce: it is efficiency-adjusted in (I) and (III), and not efficiency-adjusted in (II) and (IV).

Across specifications, Angrist-Pischke χ^2 attests that the null of underidentification of each of the endogenous regressors is rejected.[69] Shea's partial

[69] AP χ^2 is constructed by partialling out the linear projection of the other endogenous regressor. AP F-stat attests that endogenous regressors are not weakly identified.

R^2 shows that the loss of precision in the second stage linked to using instrumented endogenous regressors is relatively minor.[70]

Tab.27 reports results of the second stage. The heterogeneity parameter θ is precisely estimated at 4.4;4.5. It is little sensitive to the set of instruments used. The coefficient on hourly wages is also precisely estimated at -1.3;-1.4. The range of point estimates corresponds to a factor share of inputs in gross output of $.68 - .71$. The corresponding value of ζ for the EU-15 computed with data on gross output and expenditure on inputs in WIOD increases from .67 in 1995 to .71 in 2009, with a mean of .68 for the whole period.

Table 27: Second stage: Estimated parameters ($\zeta^k = \zeta$)

	(I)	**(II)**	**(III)**	**(IV)**
TFP	4.455***	4.386***	4.512***	4.442***
	(0.364)	(0.363)	(0.412)	(0.416)
$WAGE$	-1.337***	-1.281***	-1.431***	-1.362***
	(0.088)	(0.086)	(0.089)	(0.087)
Obs	4196	4196	4833	4833
Hansen J	2.118	1.578		
Hansen J p-val	0.146	0.209		
Kleibergen-Paap rk LM	385.2	386.5	302.1	298.4
Cragg Donald Wald F	240.1	246.7	221.8	221.4

2-step GMM estimation. Depvar is estimated exporter-sector dummy: $\widehat{fe}^k_{i,t}$.
Regressors are logs of instrumented TFP and wages. Wages are efficiency adjusted in (II) and (IV).
(I)-(IV) differ in the set of instruments. Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV).
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1
Year fixed effects included in each specification.

Reported values of Kleibergen-Paap rk LM and Cragg Donald Wald F statistics attest that instruments pass respectively the underidentification and weak identification tests across specifications.[71] As the equation is overidentified in the first two specifications, we report the result of the test of overidentifying restrictions (Hansen J statistic). The joint null that instruments are uncorrelated with the error term and correctly excluded from the estimation is not rejected at conventional significance levels.

[70] Loss in precision is approx. the reciprocal of the partial correlation: $1/pcorr$.

[71] The underidentification test rejects the null that the matrix of reduced form coefficients is not full rank. The value of the statistic in the weak identification test corresponds to a case in which instrumental variables estimator is not source of bias due to weak instruments.

D.2.2 First stage results with sector specific production functions

Here we report first stage results of the model estimated in the core of the paper in which the proximity mechanism is active because we allow the factor share of inputs to be sector-specific.

Table 28: First stage : Measured TFP (sector specific slopes)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.059***	0.057***		
	(0.004)	(0.004)		
K-stocks	0.235***	0.235***		
	(0.006)	(0.006)		
R&D: expenditure			0.167***	0.167***
			(0.004)	(0.004)
Workforce	-0.232***	-0.229***	-0.154***	-0.154***
	(0.004)	(0.004)	(0.005)	(0.005)
Workforce 19	-0.008***	-0.008***	-0.014***	-0.014***
Workforce 20	-0.000	-0.001	0.007***	0.007***
Workforce $21-22$	0.003*	0.002*	0.012***	0.013***
Workforce 24	-0.039***	-0.041***	-0.043***	-0.046***
Workforce 25	-0.012***	-0.013***	-0.017***	-0.018***
Workforce 26	-0.008***	-0.008***	-0.003*	-0.003*
Workforce $27-28$	-0.014***	-0.015***	-0.011***	-0.012***
Workforce 29	-0.005***	-0.005***	-0.018***	-0.020***
Workforce $30-33$	-0.016***	-0.017***	-0.031***	-0.034***
Workforce $34-35$	-0.035***	-0.037***	-0.044***	-0.047***
Workforce $36-37$	0.009***	0.009***	0.001	0.001
Obs	4196	4196	4833	4833
R^2	0.522	0.536	0.367	0.380
Shea's $pcorr^2$	0.175	0.193	0.147	0.162
Angrist-Pischke F-stat	223.1	245.3	275.2	301.4
Angrist-Pischke χ^2	449.2	493.8	276.8	303.2

Depvar is log of measured TFP: $\overline{z}^k_{i,t}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Coef. Workforce corresponds to elasticity for sector $17-18$.
For every other sector: elasticity given by sum of coef. Workforce and coef. of sector.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

To estimate this model we need instrumented sectoral hourly wages and instrumented TFP. Consequently, in the first stage we run 13 regressions in which measured TFP and hourly sectoral wages are regressed on a common set of instruments which include R&D personnel and real capital stocks in (I) and (II) (deflated R&D expenditure in (III) and (IV)) together with the workforce of each of the 12 sectors. In (I) and (III) sectoral workforce is efficiency-adjusted. In (II) and (IV) we use raw data on hourly wages and number of persons engaged in the sector.

Tab.28 reports first stage results for measured TFP in (I)-(IV). Tab.29

reports first stage results for hourly wages in the benchmark sector (textiles).[72] Across specifications the instruments pass underidentification and weak identification tests for each of the endogenous regressors.

Table 29: First stage : Hourly wages (sector specific slopes)

	(I)	(II)	(III)	(IV)
R&D: personnel	0.131***	0.132***		
	(0.009)	(0.009)		
K-stocks	0.829***	0.836***		
	(0.010)	(0.010)		
R&D: expenditure			0.513***	0.525***
			(0.009)	(0.009)
Workforce	-1.002***	-1.010***	-0.705***	-0.723***
	(0.008)	(0.008)	(0.012)	(0.012)
Workforce 19	-0.015***	-0.016***	-0.031***	-0.034***
Workforce 20	-0.002	-0.002	0.028***	0.030***
Workforce $21-22$	-0.000	-0.001	0.038***	0.041***
Workforce 24	-0.059***	-0.064***	-0.070***	-0.077***
Workforce 25	-0.024***	-0.026***	-0.037***	-0.041***
Workforce 26	-0.038***	-0.041***	-0.015***	-0.017***
Workforce $27-28$	-0.012***	-0.013***	0.004	0.004
Workforce 29	-0.003	-0.003	-0.049***	-0.054***
Workforce $30-33$	-0.017***	-0.019***	-0.071***	-0.077***
Workforce $34-35$	-0.034***	-0.036***	-0.062***	-0.068***
Workforce $36-37$	0.011***	0.012***	-0.012***	-0.013***
Obs	4196	4196	4833	4833
R^2	0.770	0.784	0.544	0.567
Shea's $pcorr^2$	0.334	0.379	0.311	0.341
Angrist-Pischke F-stat	394.7	644.2	270.2	394.9
Angrist-Pischke χ^2	794.5	1297	271.8	397.2

Depvar is log of measured hourly wage: $\overline{\nu}^k_{i,t}$. Regressors are logs of corresponding variables.
(I)-(IV) differ in the set of regressors. Workforce is efficiency adjusted in (II) and (IV).
Years: 1995-2007 for (I)-(II); 1995-2009 for (III)-(IV). Year fixed effects included in each specification.
Coef. Workforce corresponds to elasticity for sector $17-18$.
For every other sector: elasticity given by sum of coef. Workforce and coef. of sector.
Estimates robust to an arbitrary form of heteroskedasticity.*** p<0.01, ** p<0.05, * p<0.1

E Addenda on the proximity characteristic

E.1 Proximity endowment

Here we show that it is sufficient to assume away bilateral trade frictions to get the result that the measure of proximity to suppliers is not contingent on the distribution of market shares. We refer to the indicator of proximity to suppliers as the proximity endowment of the supplier.

[72]Results for the other 11 sectors are qualitatively similar (not reported).

Assume that bilateral sectoral trade costs τ_{ij}^k are well approximated by a destination-sector specific cost $\tau_j^{M,k}$ and an exporter-specific cost $\tau_i^{E,k}$.

$$\tau_{ij}^k = \tau_j^{M,k} * \tau_i^{E,k}$$

Trade costs on the export side measure the restrictiveness of trade policy which the exporter faces in supplying goods to world markets. Destination-specific trade costs correspond to a synthetic measure of trade frictions linked to the structure of the trade network. We think of this destination-specific mark-up incurred in sourcing inputs as a synthetic measure of location centrality.

Recall that the sectoral price index (see eqn.2) can be written:

$$P_j^k = \kappa \left[\Phi_j^k\right]^{-1/\theta}$$

where $\kappa = \left[\Gamma(\frac{\theta+1-\sigma}{\theta})\right]^{1/(1-\sigma)}$, and the sectoral price distribution parameter, given the assumption on trade costs, is given by:

$$\Phi_j^k = \sum_{i'=1}^{I} \left[\omega_{i'}^k \tau_j^{M,k} \tau_{i'}^{E,k} / z_{i'}^k\right]^{-\theta}$$

Define $\widetilde{\Phi}^k$ the realized least cost distribution of varieties in sector k common across countries:

$$\widetilde{\Phi}^k = \sum_{i'=1}^{I} \left[\omega_{i'}^k \tau_{i'}^{E,k} / z_{i'}^k\right]^{-\theta}$$

The country-specific distribution of least cost varieties can be written as a rescaled world distribution of least cost varieties:

$$\Phi_j^k = \left[\tau_j^{M,k}\right]^{-\theta} \widetilde{\Phi}^k$$

The sectoral price index has three components:

$$P_j^k = \kappa \tau_j^{M,k} \left[\widetilde{\Phi}^k\right]^{-1/\theta} \tag{23}$$

The only country-specific component of the sectoral price index is the destination-specific trade cost $\tau_j^{M,k}$ which is an indicator of the ease with which country j gets access to the world distribution of least cost varieties in sector k.

The overall price index is a Cobb-Douglas aggregator of sectoral price indices. Plugging (23) in (3), the price index can be written as a product of a country-specific index of trade frictions and of two components common to all countries: the product of sectoral price distribution parameters weighted by the share of each sector in total expenditure and the constant κ.

$$P_j = \underbrace{\prod_{s=1}^{S}\left[\tau_j^{M,s}\right]^{\gamma^s}}_{SPECIFIC}\underbrace{\prod_{s=1}^{s}\left[\widetilde{\Phi}^s\right]^{-\gamma^s/\theta}\kappa}_{COMMON} \tag{24}$$

Recall that for $i = j$ the price index in j gives the cost of the input bundle in i. In other words the composite index of sectoral trade frictions $\prod_{s=1}^{S}\left[\tau_j^{M,s}\right]^{\gamma^s}$ for $i = j$ captures how difficult it is for exporter i to get access to the best world technology in sourcing inputs. Switching sides consider production costs in country i. For $i = j$, the reciprocal of the composite index of trade frictions in j corresponds to the proximity endowment of exporter i:

$$PROX_i^M \quad = \quad \left\{\prod_{s=1}^{S}\left[\tau_j^{M,s}\right]^{\gamma^s}\right\}^{-1} \tag{25}$$

Plugging (24) in the expression of the exporter-sector specific production cost ω_i^k, we get:

$$\omega_i^k \quad = \quad \underbrace{\epsilon^k\kappa^{\zeta^k}\left\{\Pi_{s=1}^{S}\left[\widetilde{\Phi}^s\right]^{-\gamma^s/\theta}\right\}^{\zeta^k}}_{COMMON}\underbrace{\left\{\Pi_{s=1}^{S}\left[\tau_j^{M,s}\right]^{\gamma^s}\right\}^{\zeta^k}\left\{\nu_i\right\}^{1-\zeta^k}}_{SPECIFIC} \tag{26}$$

The measure of countries' proximity endowment is the correct way to measure proximity to world technology in a world in which we can abstract from the actual distribution of best practice across countries in the world. To see why this is the case, go back to the sectoral market share equation which according to the model is the probability that a given source is least cost in a given sector. Our modelling of trade costs entails that exporter-specific sectoral market shares are invariant across markets. Destination-specific price distribution parameters are $\Phi_j^k = \left[\tau_j^M\right]^{-\theta}\widetilde{\Phi}^k$, and the sectoral market share equation simplifies to an expression independent of j:

$$\pi_{ij}^k = \frac{X_{ij}^k}{X_j^k} \quad = \quad \frac{\left[\omega_i^k\tau_i^{E,k}/z_i^k\right]^{-\theta}}{\widetilde{\Phi}^k}$$

For this to hold, the distribution of best practice must be common across destinations in every sector: $\pi_{ij}^k = \pi_{ij'}^k = \pi_i^k$, $\forall k$. This would be the case if there were no pair-specific trade frictions. Such a world shares with the frictionless world the feature that any supplier revealed least cost in a variety in some market would be revealed least cost in that variety in all markets.

By definition the indicator of proximity endowment is a measure of country centrality in terms of its proximity to both its clients and its suppliers. Consequently, the identification of the role played by the input cost channel in determining the pattern of intersectoral specialization crucially relies on the interaction term between the indicator of proximity endowment and the measure of sectoral input intensity. Indeed, the interaction term is specific to a backward linkage while the indicator of relative proximity is not.

To see this, notice that the indicator of relative proximity to clients corresponds to an exporter-specific component of trade frictions common across sectors which measures trade restrictiveness the country faces in getting its products to world markets. Consequently, in relative terms, proximity to clients has a proportional effect on relative market share across the full range of sectors, and no incidence on the pattern of intersectoral specialization.

Undeniably, proximity to suppliers matters across the full range of sectors as well because higher proximity lowers production costs. However, it also matters at the intersectoral level because the ability to source inputs at lower cost matters relatively more in sectors which use inputs more intensively. Consequently, a positive and significant coefficient on the interaction effect between sectoral input intensity and the relative proximity characteristic uniquely identifies the role of a supply-side linkage in shaping the pattern of intersectoral specialization even though proximity endowment itself captures both proximity to clients and proximity to suppliers.

E.2 Persistence of time-varying proximity

In fig.13-16 we document the persistence of the proximity characteristic. We plot the evolution of the microfounded measure of distance to suppliers, $PROX^{-1}$, which can be interpreted as a synthetic indicator of (inverse) centrality measured in kilometers. Inverse proximity is plotted in 1995-2009 for a subsample of countries, including the least and the most distant countries in each subgroup.

E.3 Domestic and Foreign Supply Based Proximity

The lack of output data impedes computing the proximity characteristic which we use in the core of the paper for a wider set of countries. We refer to this indicator as domestic supply based proximity (DSBP) because it weighs

Figure 13: Inverse proximity: subset of EU15

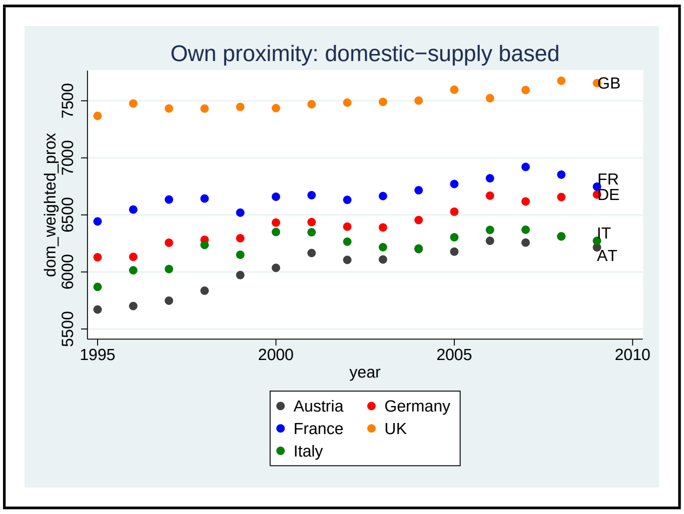
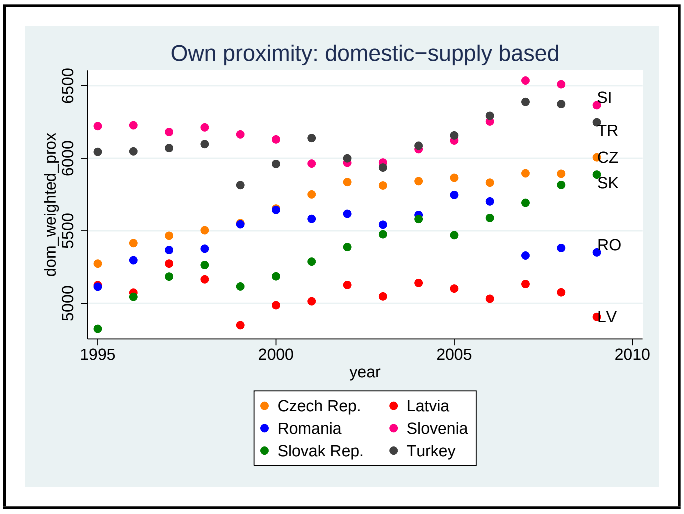


Figure 14: Inverse proximity: subset of CEECs



the components of the distance vector based on the distribution of sectoral market shares including expenditure on domestically produced varieties.

However, the availability of data on bilateral trade allows computing a proximity indicator restricted to the distribution of market shares across foreign suppliers (Foreign Supply Based Proximity) for a wider set of countries. If FSBP approximates DSBP sufficiently well, it would allow keeping more

Figure 15: Inverse proximity: subset of non-European developed

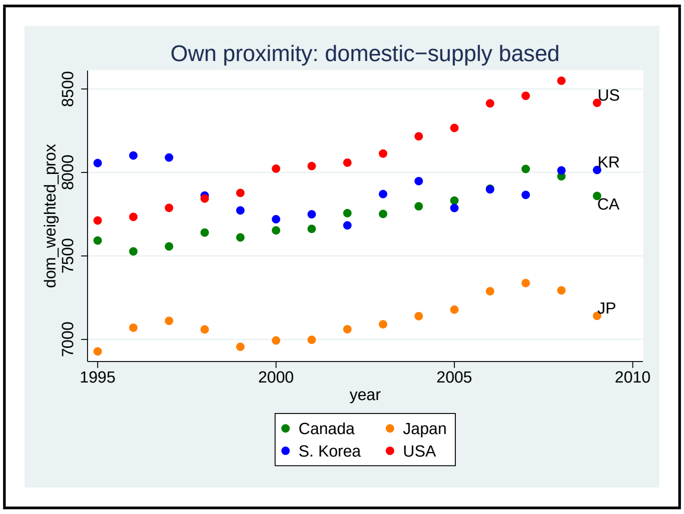
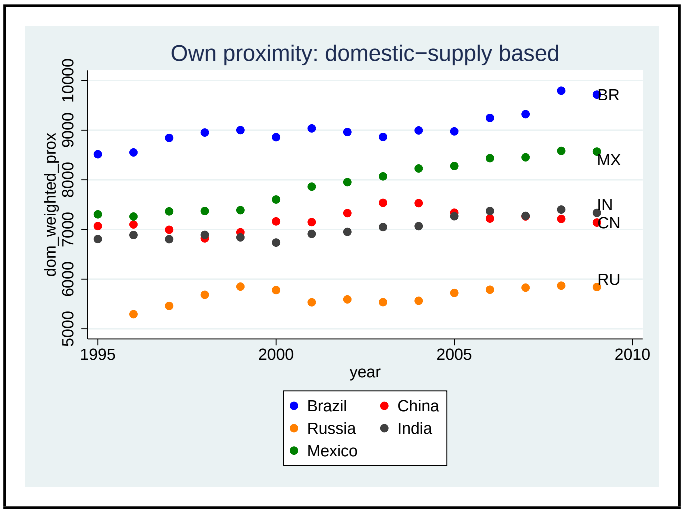


Figure 16: Inverse proximity: subset of non-European emerging



countries in the sample since FSBP can be constructed for all countries except Taiwan.

We compute this indicator for our sample, and find that FSBP picks up most of DSBP variance: 87% in 1995-2009 for the common set of countries. Fig.17-18 illustrate that the relationship becomes tighter in the recent period. Fig.19 shows the persistence of the FSBP indicator for a subset of countries.

Figure 17: DSBP and FSBP in 1996

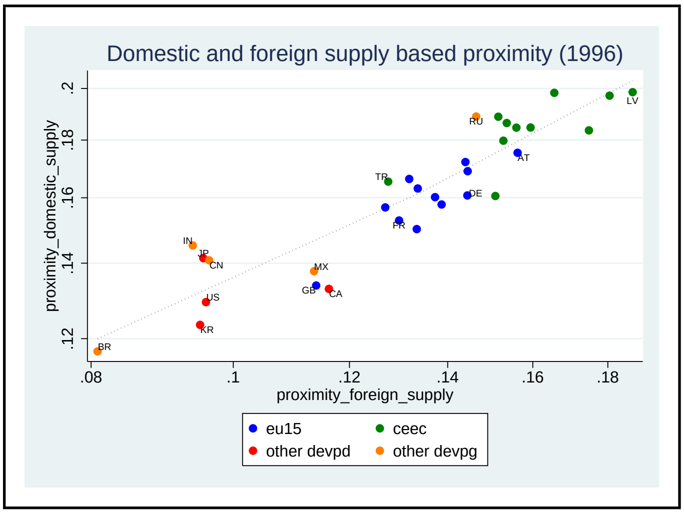
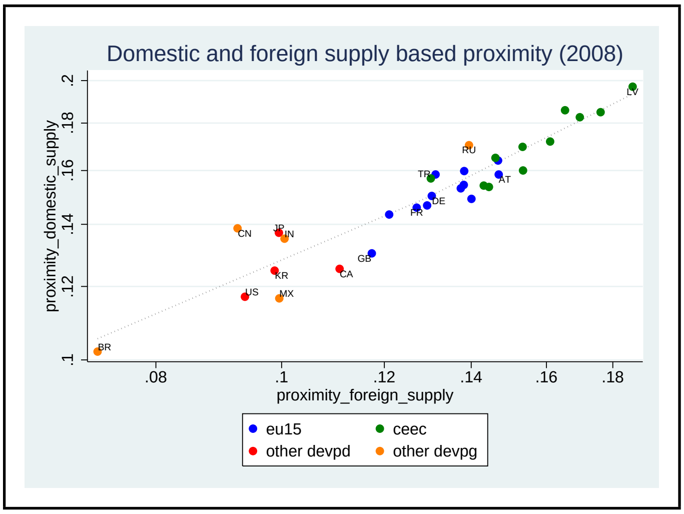
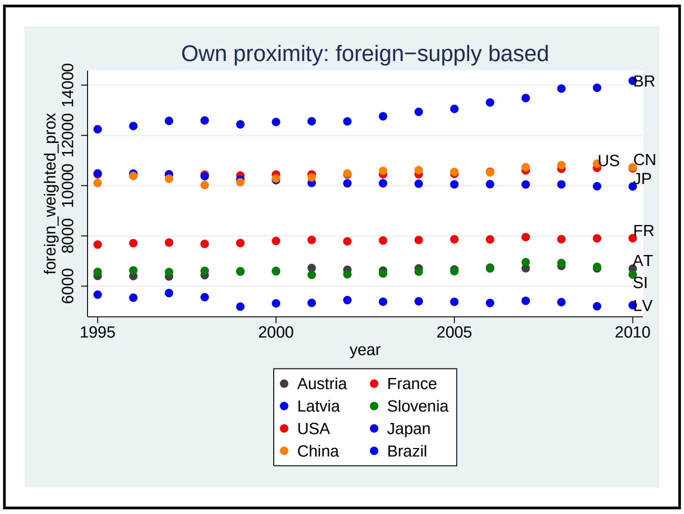


Figure 18: DSBP and FSBP in 2008



Two factors explain differences in centrality measures obtained with these indicators: country size and differences in domestic market openness to foreign supply.

Figure 19: FSBP (subset of countries)

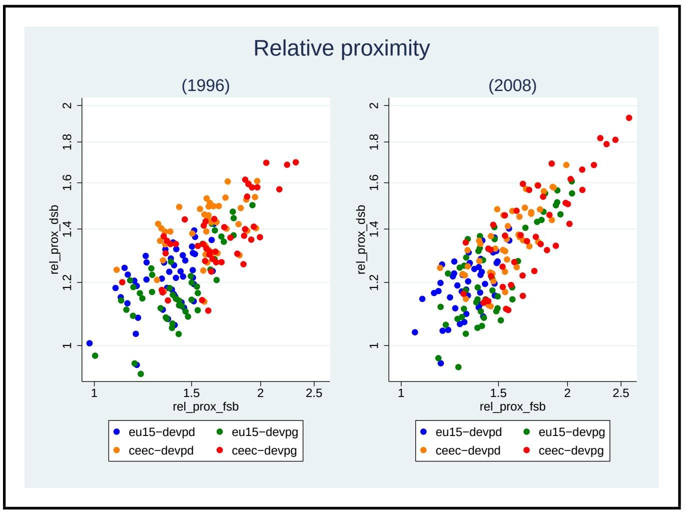


In the first case, the DSBP indicator is unbiased while the FSBP indicator underestimates proximity for relatively big countries because by construction this indicator ignores differences in domestic market size. Indeed, if trade restrictiveness is similar across destinations, then a relatively big country in terms of economic size sources a lot from itself because it is effectively least cost for a substantial chunk of varieties on the domestic market, while the smaller country sources most varieties from abroad. If internal distances are lower than external distances, then the FSBP mismeasures proximity for relatively big countries by putting excessive weight on bigger elements of the distance vector while being accurate for relatively small economies.

In the second case, both indicators are biased by a scalar which corresponds to trade restrictiveness of the domestic market τ_j. In this case, the FSBP indicator is an unbiased indicator of relative proximity for countries with a similar level of trade restrictiveness. However, the bias in the DSBP indicator is more severe because it also suffers from measurement error in the weighting function. Indeed, if the mark-up charged on varieties imported from abroad is relatively high, the share of varieties sourced domestically will also be relatively high, but not because the country is cost efficient in terms of fundamental productivity. Ignoring differences in trade restrictiveness leads to overestimation of domestic market share and underestimation of foreign market shares in the DSBP, and this mismeasurement of the weighting function is more severe for relatively closed markets. If internal distances are lower than external distances, the DSBP overestimates the true underlying proximity for relatively closed markets while it is relatively accurate for markets with a low overall level of trade restrictiveness (the next subsection makes this argument formally).

In our data, fig.18 illustrates that the DSBP predicted by the FSBP would be lower for several non-European economies such as Russia, China, and India while it would be higher for countries such as the USA and Canada. To explain this discrepancy, the argument of FSBP bias linked to differences in country size is not helpful because all of these countries are relatively big. Rather, this finding indicates that differences in domestic market openness to foreign supply may explain the discrepancy between DSBP and FSBP, i.e. that observed DSBP overestimates proximity for Russia, China, and India, and underestimates proximity for USA and Canada because Russia, China, and India are relatively less open to foreign supply than the USA and Canada. This line of argument is corroborated by the fact that the relationship between DSBP and FSBP becomes tighter in the recent period as illustrated in fig.20 which plots relative proximity by partner type measured with both indicators in 1996 and 2008. This evolution is consistent with the idea that initially relatively closed emerging economies are gradually opening up their domestic markets to foreign supply.

Figure 20: Relative proximity: DSBP and FSBP



We conclude that in our data mismeasurement in the weighting function due to not correcting for destination-specific trade restrictiveness may be a bigger source of bias in the DSBP than is mismeasurement due to omitting the share of varieties sourced from the domestic market in the FSBP.

Measurement error provides an additional motivation for instrumenting the DSBP indicator. A feasible alternative would consist in working with the FSBP indicator.

E.4 Proximity and trade restrictiveness of the destination

In this subsection we discuss the potential bias of domestic supply based proximity indicators constructed in the core of the paper. This bias is linked to the assumption that the trade policy component of bilateral sectoral trade costs $\tau_{ij,t}^{k}$ is fully accounted for by an exporter-specific cost component $\tau_{i,t}^{E,k}$ common across destination markets. Indeed, if destinations differ in the degree of trade restrictiveness towards foreign supply, these trade impediments will not be picked up by the exporter-specific trade cost component, and the trade cost function will be misspecified.

Waugh (2010) presents empirical evidence that the assumption of an exporter-specific trade cost component matches the data better than the alternative of trade restrictiveness common across suppliers. Nonetheless, the incidence of a destination-specific trade policy restrictiveness index $\tau_{j,t}$ needs to be considered in this paper because it would enter directly in the calculation of the proximity characteristic. Indeed, the microfounded measure of proximity is subject to measurement error in cross-section and overtime if there are substantial differences in trade restrictiveness across destinations in terms of domestic market openness to foreign supply. Ignoring this trade policy component induces a bias on relative proximity measured in cross-section. Furthermore, it may lead to misinterpreting the evolution of proximity.

As differences in trade restrictiveness across destinations are verified in the data (see below), it is necessary to control for the impact of this trade cost component on the measure of proximity to suppliers. We define the destination-specific overall trade policy restrictiveness index (OTRI) $\tau_{j,t} = 1 + t_{j,t}$ as in Kee et al. (2009) where $t_{j,t}$ is the uniform ad valorem equivalent tariff which if applied to all manufacturing imports of destination j would leave its aggregate manufacturing imports unchanged.[73]

Waugh (2010) discusses the impossibility of separately identifying exporter- and destination-specific components of trade costs using the structure of the EK model. Similarly, it is not possible to separately identify the contribution of exporter-specific trade restrictiveness (MA-OTRI) and destination-specific trade restrictiveness (OTRI) in determining bilateral trade flows using the theoretically grounded methodology developed by Kee et al. (2009). How-

[73] See Anderson and Neary (1996) for the theory behind welfare-based TRI.

ever, we do not actually compute exporter-specific trade cost components in the paper. Rather, we specify under which conditions it is possible to conduct the estimation while leaving this component of trade costs in the residual. Consequently, there is no conceptual conflict in correcting the DSBP indicator constructed in the core of the paper by considering a destination-specific trade cost component in the derivation of the proximity characteristic instead of an exporter-specific trade cost common across destinations.

First, we show that the weighting function must be adjusted to take into account the index of trade restrictiveness. Omitting time subscripts to simplify notation, sectoral expenditure is defined in the core of the paper: $X_j^k = PROD_j^k - EXP_j^k + IMP_j^k$. But the imports data is not corrected for the tariff charged on imports by the destination. Thus, the true underlying sectoral expenditure is: $\overline{X}_j^k = PROD_j^k - EXP_j^k + \tau_j IMP_j^k$. Domestic absorption of domestic varieties is still defined by: $X_{jj}^k = PROD_j^k - EXP_j^k$, but the true underlying value of bilateral imports is: $\overline{X}_{ij}^k = \tau_j X_{ij}^k$.

The true underlying domestic market share is: $\overline{\pi}_{jj}^k = X_{jj}^k / \overline{X}_j^k$ which is strictly smaller than domestic market share π_{jj}^k used in the core of the paper if sectoral imports and the uniform tariff t_j are positive: $\pi_{jj}^k = X_{jj}^k / X_j^k$, and $\overline{\pi}_{jj}^k < \pi_{jj}^k$ whenever $\overline{X}_j^k > X_j^k$. The gap with the true underlying domestic market share $\overline{\pi}_{jj}^k$ is increasing in the degree of trade restrictiveness in the destination.

Similarly, the true underlying market shares for foreign partners are higher than those used in the weighting function in the core of the paper if $X_{jj}^k > 0$ and $t_j > 0$ since $\overline{\pi}_{ij}^k = \overline{X}_{ij}^k / \overline{X}_j^k = \tau_j X_{ij}^k / \overline{X}_j^k$ and $\overline{\pi}_{ij}^k > \pi_{ij}^k$ whenever $\overline{X}_j^k < \tau_j X_j^k$.

Since $\overline{X}_j^k = X_j^k + t_j IMP_j^k$, it is obvious that the lower the degree of overall trade restrictiveness, and the closer are uncorrected market shares to the true underlying weights which should be used in the computation of proximity. We conclude that under the assumption that the internal component of the distance vector is smaller than all external components, the higher the trade restrictiveness in the destination, and the stronger is the bias in the uncorrected proximity indicator linked to giving excessive weight to the internal component and reduced weight to external components of the distance vector.

Second, we show that in constructing the proximity characteristic, the overall indicator must be rescaled by the trade restrictiveness index and domestic market share must be rescaled by the reciprocal of the trade restrictiveness index. Again, the higher the trade restrictiveness in the destination, and the stronger is the bias due to giving excessive weight to the internal component of the distance vector. Furthermore, relative proximity must be rescaled by relative overall trade restrictiveness.

Going back to the sectoral market share equation which states the probability that country i is the least cost producer for country j across the spectrum of varieties in sector k:

$$\overline{\pi}_{ij}^{k} = \frac{\left[\omega_i^k \tau_{ij}\tau_j / z_i^k\right]^{-\theta}}{\Phi_j^k}$$

We bring trade cost components to the left hand side of the equation and sum across all suppliers to market j in sector k including domestic consumption of domestic varieties with $\tau_j = 1$ only for domestically sourced varieties:

$$\sum_{i \neq j} \overline{\pi}_{ij}^{k} \left(\tau_{ij}\tau_j\right)^{\theta} + \overline{\pi}_{jj}^{k}\tau_{jj}^{\theta} = \frac{\sum_{n=1}^{N}\left[\omega_n^k / z_n^k\right]^{-\theta}}{\Phi_j^k}$$

Define $\widetilde{\Phi}^k = \sum_{n=1}^{N}\left[\omega_n^k / z_n^k\right]^{-\theta}$. The sectoral price distribution parameter summarizes the price distribution of best practice in the world within sector k.

Rewrite domestic expenditure: $\overline{\pi}_{jj}^{k}\tau_{jj}^{\theta} = \left\{\overline{\pi}_{jj}^{k}/\tau_j^{\theta}\right\}\left(\tau_{jj}\tau_j\right)^{\theta}$ to factor out OTRI in the sum on the LHS. We now write Φ_j^k using $\widetilde{\Phi}^k$:

$$\Phi_j^k = \widetilde{\Phi}^k \tau_j^{-\theta}\left\{\sum_{i \neq j}\tau_{ij}^{\theta}\overline{\pi}_{ij}^{k} + \tau_{jj}^{\theta}\overline{\pi}_{jj}^{k}/\tau_j^{\theta}\right\}^{-1} \tag{27}$$

As in the core of the paper, we write the sectoral price index using (27):

$$P_j^k = \kappa\left[\widetilde{\Phi}^k\right]^{-1/\theta}\tau_j\left\{\sum_{i \neq j}\tau_{ij}^{\theta}\overline{\pi}_{ij}^{k} + \tau_{jj}^{\theta}\overline{\pi}_{jj}^{k}/\tau_j^{\theta}\right\}^{1/\theta} \tag{28}$$

Using (28), the cost of the input bundle in country j is:

$$P_j = \kappa\left\{\prod_{k=1}^{K}\left[\widetilde{\Phi}^k\right]^{-\gamma^k/\theta}\right\}\tau_j\left\{\prod_{k=1}^{K}\left[\sum_{i \neq j}\tau_{ij}^{\theta}\overline{\pi}_{ij}^{k} + \tau_{jj}^{\theta}\overline{\pi}_{jj}^{k}/\tau_j^{\theta}\right]^{\gamma^k/\theta}\right\} \tag{29}$$

The corrected microfounded proximity characteristic of the exporter in a world with destination-specific trade cost components is given by:

$$\left[\widetilde{PROX}_{i,t}^{M}\right]^{-1} = \tau_i\prod_{s=1}^{S}\left\{\sum_{n \neq i}\tau_{ni}^{\theta}\overline{\pi}_{ni}^{k} + \tau_{ii}^{\theta}\overline{\pi}_{ii}^{k}/\tau_i^{\theta}\right\}^{\gamma^s/\theta} \tag{30}$$

Relative proximity indicators constructed in the core of the paper must be corrected by the scalar $\tau_i/\tau_{i'}$. Furthermore, domestic market share must be rescaled by $\tau_i^{-\theta}$ in computation of sectoral proximity. Finally, all bilateral trade shares must be adjusted to take into account trade restrictiveness of the destination.[74]

Table 30 shows that differences in trade restrictiveness across destinations are non negligible. For example, the uniform AVE tariff is three times higher for China relatively to EU countries, and almost seven times higher for Brazil.

Table 30: OTRI (t_i) in 2008 (KNO)

	OTRI	**OTRI-AGRI**	**OTRI-MANUF**	**ratio-all**	**ratio-agri**	**ratio-manuf**
EU27	**0.05**	0.37	0.03	**1.0**	1.0	1.0
Canada	**0.05**	0.19	0.04	**1.0**	0.5	1.2
USA	**0.06**	0.18	0.05	**1.1**	0.5	1.5
Switzerland	**0.05**	0.54	0.01	**1.0**	1.5	0.4
Turkey	**0.06**	0.24	0.05	**1.3**	0.7	1.9
Norway	**0.08**	0.60	0.01	**1.7**	1.6	0.3
Japan	**0.09**	0.50	0.04	**1.7**	1.4	1.3
China	**0.09**	0.15	0.09	**1.9**	0.4	3.0
Singapore	**0.13**	0.51	0.12	**2.6**	1.4	4.0
India	**0.15**	0.48	0.14	**3.0**	1.3	4.8
Russia	**0.16**	0.25	0.15	**3.2**	0.7	5.0
Mexico	**0.17**	0.29	0.15	**3.3**	0.8	5.2
Brazil	**0.20**	0.22	0.20	**4.0**	0.6	6.8
Croatia	**0.01**	0.09	0.01	**0.3**	0.2	0.2

Source: Data provided by Kee, Nicita, and Olarreaga as an .xls file.

Unfortunately, we do not have time-series data on OTRI in manufacturing. Data on average tariffs in the MacMAp database can be used to get additional data points.[75] Tab. 31 presents average tariffs in manufacturing in 2001 for countries of our sample in their trade with developed and developing countries. Even though this measure is not directly comparable to OTRI, the variability of average tariffs indicates that differences in trade restrictiveness across destinations are likely to be a permanent source of measurement error in domestic supply based proximity.

We use OTRI in manufacturing computed by Kee et al. (2009) in 2008 (see col.3 in tab.30) to construct proximity indicators which take into account differences in trade restrictiveness across destinations. Fig.21 plots OTRI-corrected DSBP and FSBP in 2008 in the RHS panel and the indicators used

[74] It is straightforward to show that to correct for trade restrictiveness gaps in FSBP, it suffices to rescale by $\tau_i/\tau_{i'}$. Indeed, market shares across foreign suppliers are invariant to a destination-specific mark-up on imports.

[75] See Bouet et al. (2008).

Table 31: Average tariffs in manufacturing (MacMAp database, 2001)

	Devpd t_i	**Devpg t_i**	**Ratio Devpd t_i/t_{eu15}**	**Ratio Devpg t_i/t_{eu15}**
EU15	**0.022**	**0.008**	**1.00**	**1.00**
Japan	0.006	0.003	0.29	0.38
Canada	0.020	0.006	0.89	0.77
USA	0.016	0.007	0.70	0.88
Switzerland	0.012	0.014	0.53	1.67
Korea	0.055	0.047	2.44	5.61
Estonia	0.000	0.000	0.00	0.00
Latvia	0.006	0.006	0.27	0.73
Lithuania	0.002	0.009	0.10	1.07
Turkey	0.020	0.034	0.88	4.03
Croatia	0.024	0.024	1.07	2.86
Poland	0.036	0.069	1.61	8.25
Hungary	0.040	0.042	1.77	4.99
Slovakia	0.043	0.026	1.92	3.06
Czech Republic	0.043	0.026	1.93	3.06
Bulgaria	0.047	0.045	2.09	5.41
Romania	0.076	0.085	3.36	10.14
Indonesia	0.055	0.045	2.44	5.42
Mexico	0.073	0.128	3.27	15.28
Russian Federation	0.100	0.082	4.45	9.75
Taiwan	0.102	0.060	4.55	7.19
Thailand	0.110	0.085	4.88	10.19
Malaysia	0.122	0.095	5.45	11.37
Brazil	0.126	0.084	5.61	10.01
China	0.137	0.099	6.08	11.77
India	0.311	0.292	13.84	34.84

Source: MacMAp database (2001). Data provided by Kee, Nicita, and Olarreaga as an .xls file.

in the core of the paper on the LHS. The fit is slightly improved by correcting for differences in trade restrictiveness. On the RHS, FSBP explains 94% of total variance in DSBP, while it explains a slightly lower share (89%) on the LHS. This finding corroborates the intuition that measurement error in the DSBP linked to overestimating centrality for relatively closed markets partly explains the discrepancy between domestic and foreign supply based proximity indicators.

Further, the correlation of the DSBP measure with the indicator of proximity endowment is enhanced once it is corrected for differences in domestic market openness. Proximity endowment is a strong predictor of domestic supply based proximity whether or not DSBP is OTRI-corrected (see fig.22). As the proximity endowment is an instrument which captures the component of proximity independent of a specific distribution of market shares, it is likely to reduce measurement error in the DSBP indicator linked to not controlling for differences in overall trade restrictiveness of the destination.

Figure 21: OTRI-correction: DSBP and FSBP

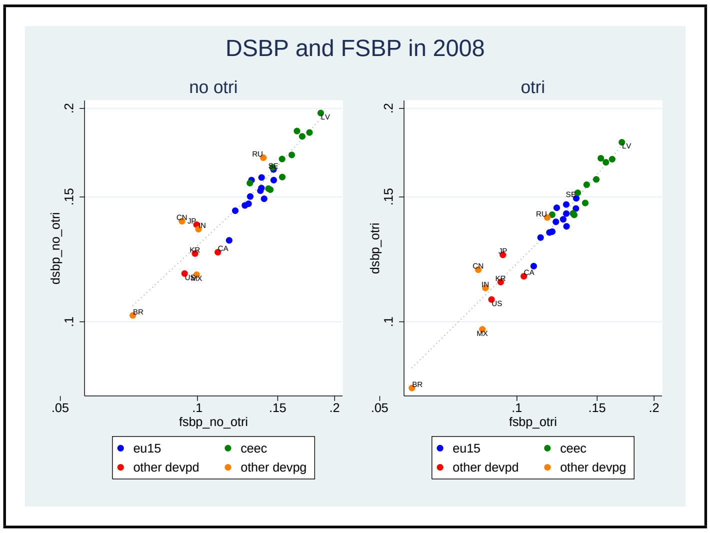


Figure 22: OTRI-correction: DSBP and ENDOWMENT