Department of Economics- FEA/USP

R&D intensity and Vertical Differentiation

RICARDO D. BRITO, EDUARDO CORREIA DE SOUZA AND RODRIGO MOITA

WORKING PAPER SERIES Nº 2021-16

Department of Economics, FEA-USP
Working Paper Nº 2021-16

R&D intensity and Vertical Differentiation

Ricardo D. Brito (ricardobrito@usp.br)
Eduardo Correia de Souza (eduardocs@insper.edu.br)
Rodrigo Moita (rodmoita@usp.br)

Abstract:

Product market characteristics vary considerably across R&D intensity-based "technology levels" of the OECD-STI taxonomy, as well as across categories of the Rauch (1999) classification. Both higher technology and more differentiated products display lower price elasticity of demand and longer quality ladders. However, variety proliferation decreases with the technology level and increases with the Rauch category, while price dispersion increases with technology but not with the Rauch. Additionally, Rauch categories do not differ in factor intensities, while higher tech industries are more capital intensive than lower tech ones. From this evidence, we conclude that R&D intensity is an appropriate measure of vertical differentiation, while the Rauch classification mainly captures horizontal differentiation.

Keywords: R&D intensity; Product differentiation; Product market characteristics.

JEL Codes: L10; F14

R&D intensity and Vertical Differentiation [ℵ]

Ricardo D. Brito, Eduardo Correia de Souza* and Rodrigo Moita

FEA-USP, Insper Institute and FEA-USP, São Paulo, Brazil

Abstract

Product market characteristics vary considerably across R&D intensity-based "technology levels" of the OECD-STI taxonomy, as well as across categories of the Rauch (1999) classification. Both higher technology and more differentiated products display lower price elasticity of demand and longer quality ladders. However, variety proliferation decreases with the technology level and increases with the Rauch category, while price dispersion increases with technology but not with the Rauch. Additionally, Rauch categories do not differ in factor intensities, while higher tech industries are more capital intensive than lower tech ones. From this evidence, we conclude that R&D intensity is an appropriate measure of vertical differentiation, while the Rauch classification mainly captures horizontal differentiation.

Keywords*:* R&D intensity; product differentiation; product market characteristics

JEL classification*:* L10 ; F14

March 2021

ℵ We appreciate the helpful comments and suggestions of Bernardo Blum, Emanuel Ornelas and Luciana Ferreira, and seminar participants at Insper and USP. Ítalo Franca provided excellent research assistance. All errors are our own. Ricardo Brito additionally thanks to the financial support of CNPq grant n. 310360/2016-1.

* Corresponding author. E-mail address: eduardocs@insper.edu.br. Postal address: Rua Quatá, 300, São Paulo, SP 04546-042, Brazil.

1. Introduction

The Rauch (1999) classification is widely used to empirically capture between-industry heterogeneity, being treated as a measure of both horizontal and vertical differentiation. However, another classification system that captures vertical differentiation only would be desirable. Sutton (1998) argues for the use of research and development intensity as a proxy for the scope of vertical differentiation. This paper investigates the appropriateness of R&D intensity as a measure of vertical differentiation.

For our investigation, we examine how product market characteristics, such as the price elasticity of demand, quality ladders[1], price dispersion, variety proliferation and factor intensities, vary along the Rauch classification and along the OECD-STI taxonomy. Elaborated by the OECD Directorate for Science, Technology and Industry, this taxonomy divides industries into "technology levels" based on the cross-industry distribution of R&D intensity.

Using Robert Feenstra's data, we estimate the US import demand for a comprehensive set of manufactured products from different countries and industries. We find that both high-tech and Rauch differentiated products display longer quality ladders and lower price elasticities than their low-tech and Rauch homogeneous counterparts. At the same time, other characteristics separate the two classifications. While the number of varieties per product increases along the Rauch classification, it decreases along OECD-STI technology levels. Price dispersion increases with the technology level, but not significantly so with the Rauch classification. While factor intensities are practically

[1] Based on Khandelwal (2010), a "quality ladder" is a measure of within-industry quality dispersion. "Quality" is here understood as a demand shifter: a certain product (exported by a certain country or supplied by a certain firm) is identified as higher quality when, conditional on price, it has a higher market share.

identical across Rauch categories, high-tech products are more physical and human capital intensive than low-tech ones.

These findings corroborate Kugler and Verhoogen's (2012) conjecture that while R&D intensity captures vertical differentiation, the Rauch classification mainly captures horizontal differentiation. The stylized facts concerning the OECD-STI taxonomy also throw light on what R&D investments are predominantly directed at. In terms of Sutton's (1998) framework, high-tech products conform much more to the view that R&D improves quality in industries with limited variety (i.e., the "escalation mechanism") than to the view that R&D proliferates variety in industries with little quality differentiation (as in Sutton's "flowmeters" example).

In the remainder of the paper, section 2 first provides microfoundations to the product market characteristics we want to measure, and then links patterns of these characteristics to horizontal and vertical differentiation. Section 3 presents the data. Section 4 presents cross-industry averages of the characteristics according to the OECD-STI technology levels or Rauch categories. In section 5, we double sort products by technology level and by Rauch category to show that the OECD-STI taxonomy is relevant on its own in systematizing characteristics. In section 6, we consider the alternative, more disaggregated R&D intensity measures that could be used instead of the OECD-STI taxonomy. Section 7 concludes.

2. Setup

To provide a unified structure to the product market characteristics we use in this paper, we start from a utility maximization problem involving products' prices, varieties, and qualities. Consider thus a Dixit-Stiglitz utility function like Hummels and Klenow's (2005). In an international trade context, as is the case for our data, the expression below describes the representative consumer's preferences over products (and varieties of the same product) supplied by different exporting countries:

$$U = \left[\sum_{j=1}^{J}\sum_{i=1}^{I} Q_{ij} \cdot N_{ij} \cdot \left(x_{ij}^{\frac{\sigma-1}{\sigma}}\right)\right]^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where N_{ij} is the number of symmetric varieties of product i exported by country j ; Q_{ij} is the quality of country j's varieties of product i ; and x_{ij} is the quantity of each variety of product i imported from country j. The $\sigma > 1$ is the single elasticity of substitution between different products and between different varieties within each product.

The representative consumer maximizes equation (1) subject to:

$$\sum_{j=1}^{J}\sum_{i=1}^{I} N_{ij} \cdot P_{ij} \cdot x_{ij} \leq Y , \tag{2}$$

where P_{ij} denotes the price exporter country j charges for product i varieties, and Y denotes the representative consumer's (or importing country's) income.

Suppose now that the same product i is exported by countries j and k . Standard maximization of equation (1) subject to equation (2) yields:

$$\frac{x_{ij}}{x_{ik}} = \left(\frac{P_{ij}}{P_{ik}}\right)^{-\sigma} \cdot \left(\frac{Q_{ij}}{Q_{ik}}\right)^{\sigma} \quad \forall \ i,j,k \ ; \tag{3}$$

and given that the observed total quantity of i exported by j is actually $X_{ij} \equiv x_{ij} \cdot N_{ij}$, equation (3) becomes:

$$\frac{X_{ij}}{X_{ik}} = \frac{N_{ij}}{N_{ik}} \cdot \left(\frac{P_{ij}}{P_{ik}}\right)^{-\sigma} \cdot \left(\frac{Q_{ij}}{Q_{ik}}\right)^{\sigma} \quad \forall \ i,j,k \,. \tag{4}$$

Like Hummels and Klenow (2005), assuming a reference country k where $P_k = X_k = N_k = Q_k = 1$ and taking logs of equation (4), we arrive at the fundamental relation between the dependent variable "relative exported quantity" and the explanatory variables "relative variety", "relative price" and "relative quality":

$$lnX_{ij} = lnN_{ij} - \sigma \cdot lnP_{ij} + \sigma \cdot lnQ_{ij} \quad \forall \, i,j,k \,. \tag{5}$$

Equation (5) motivates the relation between demanded quantities and exports' characteristics. Exporters can attract demand by supplying more variety at lower prices and better quality.

Although quantities, prices and numbers of varieties are directly observable, quality is not. Thus, following Khandelwal (2010), in the empirical implementation we estimate imports demand functions as:

$$lnX_{ijt} = \beta_0 + \sigma_N \cdot lnN_{ijt} + \sigma_P \cdot lnP_{ijt} + \beta_t + \beta_{ij} + \varepsilon_{ijt} \quad \forall \ i,j,t \,; \tag{6}$$

where X_{ijt} is exporting country j's quantity of product i in U.S. imports at time t ; P_{ijt} is country j's unit price for product i that includes the FOB price plus transportation costs and tariffs; and N_{ijt} is the number of 10-digit HS (Harmonized System) varieties that country j exports within product i . The β_{ij} is a product-country fixed-effect; β_t is a common time-effect; and ε_{ijt} is the error term.

Equation (6) is estimated separately for each industry. An "industry" is a pooling of products in which each product i belongs in a single industry. In our sample, the product disaggregation level corresponds to the 5-digit SITC (Standard International Trade Classification), and the industry level corresponds to the 4-digit SITC.

Next, by imposing the restriction implied by equation (5) that $\sigma_N = 1$ in equation (6), we measure the quality for each 5-digit product-country-year triple (i, j, t) as the sum of the product-country fixed-effect β_{ij}, the common time-effect β_t, and the error term ε_{ijt}, similar to Khandelwal's (2010):[2]

$$-\sigma_P \cdot lnQ_{ijt} = \beta_t + \beta_{ij} + \varepsilon_{ijt} \quad \forall\, i, j, t\,. \tag{7}$$

The logic behind this estimated quality is that, controlling for price in a given industry, countries with larger quantities must be exporting better quality products. This estimate is also controlled for observed variety: a country may lose quantity in the industry not because of a decline in quality, but because it is exporting fewer varieties than before.

The quality obtained in equation (7) allows intra-industry comparisons among different countries that export to the US.[3] From equation (7)'s quality estimates for each product-country, we calculate yearly *quality ladders* for each 4-digit industry through interquartile ranges:

[2] Khandelwal (2010) assumes Berry's (1994) nested logit system and infers quality directly from the sum $(\beta_t + \beta_{ij} + \varepsilon_{ijt})$. Provided we assume equation (5), we need to divide it by σ_P.

[3] Implicit in the estimation of (7) is the reasonable assumption that domestic (i.e., US) production affects equally all imported varieties. Unfortunately, we cannot directly measure the quality of US domestic varieties because the NBER-CES database does not contain information on prices and the number of varieties of US domestic production at the high level of disaggregation used here.

$$quality\ ladder_{I4t} = \underset{ij \in I4}{\text{3rd. quartile}}\{lnQ_{ijt}\} - \underset{ij \in I4}{\text{1st. quartile}}\{lnQ_{ijt}\}\ , \tag{8}$$

where $I4$ is the 4-digit industry set of 5-digit products i exported to the US by countries j, and t denotes the respective year. The ladder is a measure of quality dispersion inside the 4-digit industry.

Analogous to equation (8), we define the price dispersion as the industry's "price ladder":

$$price\ dipersion_{I4t} = \underset{ij \in I4}{\text{3rd. quartile}}\{lnP_{ijt}\} - \underset{ij \in I4}{\text{1st. quartile}}\{lnP_{ijt}\}\ . \tag{9}$$

Considering price dispersion as a separate market characteristic is justified when we notice that equations (6) and (7) make our quality estimate orthogonal to price. Therefore, we do not have reasons to expect a high correlation between quality ladder and price dispersion.[4]

Finally, we define "variety proliferation" or simply the "number of varieties" for each 5-digit product i. Let $I5_{ij}$ be the set of HS varieties of product i exported by country j, which implies that $\#(I5_{ij}) = N_{ij}$. Then,

$$N_i = \#\left(\bigcup_j I5_{ij}\right) \tag{10}$$

[4] In our sample, the cross-industry correlation between price dispersion and the quality ladder is only 0.18 .

, where " # " and " ∪ " denote the cardinality and union operators respectively. That is, the number of varieties of product i is the number of *distinct* HS varieties when we consider all exporting countries. Notice that in principle N_i is different from (and smaller than) $\sum_j N_{ij}$ since one same HS variety may be exported by two or more countries.

Having defined the product market characteristics, we proceed to consider how patterns of characteristics correspond to theoretical models of vertical and horizontal differentiation. This is shown in Table 1 that we constructed drawing on Sutton (1998) and on Beath and Katsoulacos (1991).

Table 1: Patterns of Characteristics and Models of Product Differentiation

horizontal / vertical	low	high
high	• quality ladders = long • price elasticity = in general, low (but there may be exceptions) • number of varieties = moderate • price dispersion = big	• quality ladders = long • price elasticity = low • number of varieties = large • price dispersion = big
low	• quality ladders = short • price elasticity = high • number of varieties = small • price dispersion = small	• quality ladders = short • price elasticity = low • number of varieties = large • price dispersion = small, but depends on costs

In Table 1, the bottom left case corresponds to product markets where both vertical and horizontal differentiation are low. This case is closer to homogeneous goods markets.

Keeping vertical differentiation constant, an increase in horizontal differentiation will typically bring about an increase in variety proliferation and a reduction in the price elasticity of demand.[5] The classic example is when horizontal differentiation is high and

[5] Through equations (1) to (5) of this section, we see that the price elasticity is structurally related to the elasticity of substitution between different products or varieties of the same product. In Schmalensee (1978) and Broda and Weinstein (2006), a low elasticity of substitution indicates horizontal differentiation.

vertical differentiation is low (bottom right case in Table 1), which corresponds to Schmalensee's (1978) ready-to-eat cereal and to Sutton's (1998) flowmeters industries. According to Sutton, in such industries all R&D efforts are targeted at variety proliferation and not at improving quality.

Keeping horizontal differentiation constant, the characteristic that better captures an increase in vertical differentiation is the lengthening of quality ladders. The classic example is when vertical differentiation is high and horizontal differentiation is low (top left case in Table 1), which corresponds to Sutton's (1998) "escalation mechanism". According to Sutton, in industries marked by escalation all R&D efforts are targeted at improving quality that leads to concentration.

The number of varieties and price elasticity are less clear indicators of vertical differentiation. Although some variety proliferation may be needed to accommodate quality differences, both Sutton (1998) and Beath and Katsoulakos (1991) argue that "finiteness", i.e. the existence of an upper limit to the equilibrium number of varieties, is a characteristic of vertically differentiated markets.[6] This is why we mark the number of varieties as "moderate" in Table 1's top left case.

As for the price elasticities in vertically differentiated markets, as compared to non-differentiated markets, conclusions are not straightforward. As pointed out by Coibion et al (2007), what we observe in a market are equilibrium elasticities. These elasticities increase with consumers' price sensitivity but also with the size of the market segment, since a large segment attracts more firms, and more firms indicate closer substitutes. And they decrease with the scope for horizontal differentiation and with entry

[6] Lahmandi-Ayed (2007) challenges this view. According to him, finiteness may also arise in horizontally differentiated markets – as long as consumers agree on the ranking of products when these are sold at marginal costs.

sunk costs, as both things mean that less firms are producing in that region of the product space.

In spite of Table 1's top left case ruling out (by assumption) horizontal differentiation, there remain two forces leading to low price elasticities. First, in a vertically differentiated market, we expect consumers who buy high quality varieties to be less price sensitive. Second, high quality is in general the result of R&D investments that represent higher entry sunk costs. This is why we mark the price elasticity as "low" in Table 1's top left case. Notwithstanding, in view of the points raised by Coibion et al (2007), we admit that there may be exceptions.

Finally, price dispersion is still used in the literature as a proxy for vertical differentiation (see Fernandes and Paunov 2013 and Bastos and Silva 2010). This is done on the assumption of a positive correlation between price and quality, whereas our quality estimate à la Khandelwal (2010) is by construction orthogonal to price. For this reason, we regard price dispersion as an alternative measure to the quality ladder and make the entries in Table 1 coincide.

In the results section (section 4 below), we analyze how product market characteristics vary along each industry classification: the OECD-STI "technology levels" and the Rauch (1999). Observing how variations in characteristics fit in Table 1, we concord the two classifications with models of product differentiation. Moving up within a classification (from low-tech to high-tech products, or from Rauch homogeneous to differentiated products) will move from the "baseline" bottom left case in Table 1 to either vertical or horizontal differentiation, or both.

3. Data

In this paper, we analyze a panel of over 1,500 SITC (Standard International Trade Classification) and almost 10,000 HS (Harmonized System) products imported by the United States from 159 countries. Our primary source is Robert Feenstra's US import database at UC Davis Center for International Data (http://cid.econ.ucdavis.edu/usix.html). This database has the values, quantities, transportation costs and tariffs for each 5-digit SITC and 10-digit HS product of each exporting country. [7]

The period analyzed is from 2002-2006 that reflects the CID-econ's availability but also has the advantage of comprising years of relative normalcy in trade flows – after the 2001 recession in the US and before the international financial crisis of 2008. We consider only manufactured goods (SITC codes 5 to 8). Throughout, we use the 10-digit HS disaggregation level solely to provide a count of the number of "varieties" behind each 5-digit SITC "product", while we call the SITC 4-digit level the "industry". [8] Unless otherwise stated, all values are in 2006 constant dollars deflated by the US CPI.

Worldwide, policymakers are familiar with classification systems of industries based on R&D intensity (e.g., R&D expenditure/value added). Here we use the OECD Directorate for Science, Technology and Industry (OECD-STI) taxonomy in Galindo-

---

[7] This database is described in detail in Feenstra et al. (2002).

[8] So that, if country j exports four HS goods and country l exports two HS goods all classified under the 5-digit SITC product $ZZZZZ$ tag, then we will say j exports twice as many $ZZZZZ$ varieties as l. If only countries j and l produce and export $ZZZZZ$, being that j exports the varieties $ZZZZZ1$, $ZZZZZ2$, $ZZZZZ3$, $ZZZZZ4$, and l exports $ZZZZZ4$ and $ZZZZZ5$, then we will say that there exist five $ZZZZZ$ varieties. This way of interpreting a "variety" is consonant with Hummels and Klenow's (2005), but is not unique. For example, Broda and Weinstein (2006) define a "variety" as a (disaggregated product, exporting country) pair.

Rueda and Verger's (2016) revision[9] that we couple with Hatzichronoglou's (1997) high technology product classification.

The OECD-STI taxonomy sorts ISIC (International Standard Industrial Classification) industries in four R&D intensity (or "technology") levels: high, medium-high, medium and medium-low. The "low" level is for services only. The Hatzichronoglou's (1997) detailing allows a closer correspondence with international trade classifications, in particular with the SITC Revision 3 that is precisely the product level classification in our US import database. [10]

The SITC 4-digit industries are also coded according to the Rauch classification that is available at http://econweb.ucsd.edu/~jrauch/rauch_classification.html. As is well known, the empirical criterion behind this classification is merely whether products possess "quoted" (reference) prices or not. In particular, we use the so-called Rauch "conservative" classification, but our results are similar when using the Rauch "liberal" classification. [11] Because there were only 17 Rauch "organized" industries out of 502 4-digit industries in our final sample, we group Rauch "Organized" and "Reference priced" categories into a "Homogeneous" category, as opposed to the "Differentiated" one. [12]

---

[9] Because the most recent year available in the US import database is 2006, we cannot estimate product market characteristics for a period around 2011, which is the base year for Galindo-Rueda and Verger's revision. With our sample period being 2002-2006, we have alternatively tried the older revision in http://www.oecd.org/sti/ind/48350231.pdf, based on the OECD Science, Technology and Industry Scoreboard 2003. The results (available on request) are practically identical to the ones presented in the following sections.

[10] Besides the Hatzichronoglou's (1997) classification, we use the Affendy's et al. (2010) ISIC-SITC concordance table to assign an OECD-STI technology level to each SITC 5-digit product and 4-digit "industry". In a few cases, we had to resort to our judgment based on descriptive contents.

[11] The two classifications do not differ much. Out of the 502 4-digit industries in our final sample, 374 are "differentiated" by Rauch "conservative", and 364 by Rauch "liberal".

[12] That is precisely what Rauch (1999) does.

To illustrate, Table 2 lists the top five industries (in terms of imported values) in each OECD-STI–Rauch category.

Table 2 - Largest Industries in Each Tech-Rauch Level (STIC Rev.3, 4-digit)

SITC	*Medium-low - Homogeneous*	SITC	*Medium-low - Differentiated*
6413	Other printing and writing paper,coated	8514	Other footwear with uppers of leather or composition leather
6343	Plywood of wood sheets	8426	Women's and girls' trousers,breeches and shorts,woven fabrics
6421	Cartons,boxes,cases n.e.s.of paper	8414	Men's and boys' trousers, breeches and shorts,woven fabrics
6417	Other paper and paperboard,coated,covered,printed	8211	Seats and parts thereof (excl. medical etc.seats)
6516	Other synthetic filament yarn (excl.sewing thread)	8453	Jerseys,pullovers,waistcoats,knitted
SITC	***Medium - Homogeneous***	**SITC**	***Medium - Differentiated***
6672	Diamonds (excl.sorted industrial diamonds)	6251	Tyres,new,for motor cars
6841	Aluminium and aluminium alloys,unworked	6252	Tyres,new,for busses and lorries
6842	Aluminium and aluminium alloys,worked	6726	Semi-finished products of iron or non-alloy steel cont. less than 0.25%
6821	Copper and copper alloys,unworked	6794	Other tubes and hollow profiles,of iron or steel
5822	Other plates,sheets,etc.of plastics,non-cellular,not combined with other materials	6791	Seamless tubes,of iron or steel
SITC	***Medium-high - Homogeneous***	**SITC**	***Medium-high - Differentiated***
5157	Other heterocyclic compounds;nucleic acids	7812	Motor vehicles for transport of persons
5111	Acyclic hydrocarbons	7611	Colour tv receivers (incl.monitors)
5156	Lactams;heterocyclic compounds with oxygen heteroatoms	7132	Internal combustion piston engines for cars and tractors
5623	Mineral or chemical fertilizers,potassic	7843	Other parts and accessories of motor vehicles
5922	Albuminoidal substances,modified starches and glues	7232	Excavators and shovel-loaders
SITC	***High - Homogeneous***	**SITC**	***High - Differentiated***
5251	Radio-active chemical elements and isotopes	5429	Other medicaments
5743	Polycarbonates,alkyd resins and other polyesters	7522	Digital adp machines
5311	Synthetic org.colouring matter	7643	Transmission apparatus for radio-telephony,-telegraphy, -broadcasting or
5411	Provitamins and vitamins,unmixed	7523	Digital processing units
5259	Other isotopes;compounds of rare-earth metals	7638	Sound and video recording apparatus,etc.

Notes: In each Tech-Rauch group, industries are ordered by value. Medium-low, Medium, Medium-high and High are OECD tech classification levels. Homogeneous and Differentiated are Rauch categories.

We start with a panel of 655 4-digit industries that appear for at least two years in the US import database. Several problems lead to data exclusion: in some 4-digit industries the 5-digit products have their quantities reported in different, incompatible measurement units; while the Rauch Classification is originally designed for SITC Revision 2, the industries in our data are coded by SITC Revision 3. We also drop industries for which we find insignificant price elasticities when estimating imports demand functions.

As usual, we trim the sample across a few dimensions: first, we exclude observations for which the relative FOB price plus transportation costs and tariffs (the relative unit value, or RUV, defined in Appendix A) is smaller than 0.1 or greater than 10, as in Johnson (2012). Also, we exclude observations with a positive reported quantity, but the value is equal to zero, or the positive value has zero quantity.

Our final sample is an unbalanced panel of country-product-year observations in 502 4-digit industries.[13] This represents a trade volume of 2.685 trillion dollars that is almost 91% of the total imports in the original sample. In Appendix A, we detail the construction of variables: exported quantity, price, number of varieties, transportation costs and tariff.

We also calculate revealed factor intensity (RFI) indices for the SITC 5-digit products and 4-digit industries. This is accomplished by applying the Cadot's et al. (2010) RFI methodology on UN-Comtrade (http://comtrade.un.org/data/) world trade flows and on physical and human capital endowments.[14] By the RFI index, a product's factor

[13] In our webpage (https://www.insper.edu.br/docentes/eduardo-correia-souza/), under this paper's title, we provide the list of 4-digit industries in our sample, classified by technology level and by Rauch category, and with the number of SITC 5-digit products, the number of exporting countries and the total value of U.S. imports (2002-2006) by industry.

[14] Factor endowments data are also from Cadot et al. (2010). Physical capital is in US dollars per worker, and human capital is in average years of schooling for the working age population. Because some SITC

intensity is the weighted average of its exporting countries' endowments. However, the weights reflect the importance of the product under consideration in each country's exports, not the importance of each country in that product's world market (see Appendix A for details).

4. Product market characteristics across each taxonomy

Among our characteristics, while the numbers of varieties and price dispersions are observable, we have to estimate the demand-price elasticities (σ_P in equation 6) and quality ladders. For each SITC 4-digit industry panel, we first estimate equation (6) by Instrumental Variables (IV) with transportation costs and real exchange rates as instruments to account for price endogeneity;[15] and also by Ordinary Least Squares (OLS) for reference in Table B.1 of Appendix B.

We examine how product market characteristics vary across the OECD-STI and the Rauch classifications for the 502 4-digit industries with significant IV demand-price elasticities estimates. Table 3 presents descriptive statistics for all the characteristics we use in our analysis: price elasticity, price dispersion, quality ladder, number of varieties, number of exporting countries per 5-digit product (proxying for competition, i.e., a product market with more exporters is more competitive) and products' factor intensities ("supply-side" variables).

---

products appear in the US import database but not in Comtrade's, we were able to estimate factor intensities for 411 4-digit industries only.

[15] The real exchange rates relative to the US are from http://bruegel.org/2012/03/real-effective-exchange-rates-for-178-countries-a-new-database/.

Table 3 - Descriptive Statistics

Variable	*Mean*	*Median*	*1st quartile*	*3rd quartile*
IV Price Coeff.	-1.46	-1.40	-1.75	-1.16
Quality Ladder	2.68	2.41	1.92	3.08
Price dispersion	1.74	1.64	1.34	2.00
Number of HS Varieties per product	6.59	2	3.5	7
Number of exporting countries per product	60.03	55	39	73
Median Physical K Intensity [a]	116,873	120,323	85,673	147,628
Median Human K Intensity [a]	8.63	8.87	8.04	9.46

Notes : 502 industries with significant (at 95% confidence) IV price coefficient (out of 552 described in Table B.1) are included in this table. The variables' definitions are in Appendix A. The "mean", "median, "1st quartile" and "3rd quartile" are from the cross-section distribution of all 4-digit industries in the sample. *t* statistics are in parentheses, computed from standard errors clustered by exporting country. a. COMTRADE provides Physical and Human capital intensity for 411 out of these 502 industries.

For each OECD-STI technology level and Rauch category, Table 4 gives the cross-industry average of the price elasticities estimated in the IV Price Coeff. regressions. A "Mean" is the cross-section average of the 4-digit industries' price slopes within the respective group: the OECD-STI Medium-low, Medium, Medium-high or High tech, and Rauch Homogeneous or Differentiated. The standard error of an average price slope is the cross-section standard error divided by the square root of the number of industries, if the industries' slopes are independent.

We also report the *t*-statistics from the tests for equality between the averages of two groups' price slopes, that is, the difference in the averages divided by the standard error of this difference. Assuming independence between the groups, the standard error of this difference is the square root of the sum of variances of the average slopes of the groups being compared. [16]

[16] In Table B.2 of Appendix B, we reestimate equation (6) in a panel that pools all industries belonging to a certain technology level or Rauch category together, instead of first estimating (6) for each industry, and then taking cross-industry averages. Because the different price-elasticities are estimated jointly in this panel, the correlations across industries are accounted for, providing a sense of their importance. The implied *t*-statistics of equal price-coefficients in Table B.2.2 (i.e., the square root of the *F*-test statistics of

Table 4 - Means Across 4-digit Industries of Market Characteristics - OECD-STI & Rauch

Characteristics:	*Grouping by:* Technology levels: ***Medium-low***	***Medium***	***Medium-high***	***High***	Rauch categories: ***Homogeneous***	***Differentitated***
	4.1. Means of the Market Characteristics and Tests					
IV demand price coefficient	-1.63	-1.55	-1.30	-1.37	-1.69	-1.38
	(0.03)	(0.06)	(0.03)	(0.06)	(0.05)	(0.02)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		-1.16	-3.51	0.97		-5.91
Quality ladder	2.17	2.51	2.96	3.31	2.25	2.83
	(0.06)	(0.11)	(0.10)	(0.21)	(0.08)	(0.07)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		-2.78	-3.12	-1.51		-5.41
Price dispersion	1.56	1.35	1.94	2.05	1.62	1.78
	(0.03)	(0.06)	(0.05)	(0.08)	(0.07)	(0.03)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		2.98	-7.11	-1.1		-2.13
Median no. of HS Varieties per 5-digit product	9.88	4.08	5.18	5.57	4.02	7.47
	(1.04)	(0.54)	(0.56)	(0.73)	(0.37)	(0.56)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		4.94	-1.4	-0.42		-5.14
No. of exporting countries per 5-digit product	42.94	25.45	21.78	24.88	20.97	32.62
	(2.18)	(1.61)	(0.80)	(1.97)	(1.08)	(1.18)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		6.45	2.04	-1.46		-7.28
Median physical capital intensity (in dollars per worker)	86,502	109,144	138,676	139,474	116,561	116,989
	(3,303)	(5,339)	(2,809)	(3,918)	(4,087)	(2,536)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		-3.61	-4.9	-0.17		-0.09
Median human capital intensity (in years of schooling)	7.86	8.74	9.07	9.19	8.75	8.59
	(0.11)	(0.12)	(0.07)	(0.10)	(0.11)	(0.07)
	H_0:	*M.Low = Med.*	*Med. = M.High*	*M.High = High*		*Homog. = Diffe.*
		-5.43	-2.4	-1.01		1.33
	4.2. Number of observations and values					
No. observations	87,812	20,735	47,218	19,976	36,818	138,923
Total value	607,338	271,440	1,269,504	537,510	377,401	2,308,320
No.4-dig.industries	163	78	192	69	128	374
No.5-dig.products	511	220	606	234	484	1,087
No. countries	159	148	157	155	150	159

Notes: 502 industries with significant (at 95% confidence) IV price coefficient (out of 552 described in Table B.1) are included in this table. Equation (6) in the text is estimated for each 4-digit industry. The variables' definitions are in Appendix A. Means of Medium-low (M.Low), Medium (Med.), Medium-high (M.High), High, Homogeneous (Homog) and Differentiated (Diffe.) are calculated from the cross-section distribution of the 4-digit industries in each group. Standard-errors of "Mean" are in parentheses. The *t*-statistics of "Mean" uses the standard error of mean; i.e., the cross-section standard error divided by the square root of the number of 4-digit industries included in the mean. "H_0: XX=YY" rows report *t-statistics of equal means between XX and YY*. Assuming independence between groups, the standard error of the difference in means is the square root of the sum of the variances of the average slopes of the compared groups.

equal coefficients) are higher than the respective ones in Table 7, indicating the conservativeness of our inferences.

In Table 4, we see that price elasticities are significantly different between Medium and Medium-high (*t*-statistics of -3.51), and between Homogeneous and Differentiated (*t*-statistics of -5.91). With industries classified by the OECD-STI taxonomy, the "Top-tech group", which adds Medium-high- and High-tech industries, is on average less price sensitive than the "Bottom-tech group", which adds the Medium-low- and Medium-tech.[17] Similarly, with industries organized by the Rauch taxonomy, the Differentiated group is on average less price sensitive. Recalling Table 1, these patterns indicate that the substitutability among products is stronger in low-tech and homogeneous goods markets, weakening as we go up in both classifications.

Table 4 also depicts the average quality ladders across industries. We find that the quality ladders increase in both the technology levels and in the Rauch categories, reaching their longest for the high tech industries.[18] Recalling Table 1, this indicates that vertical differentiation increases along both classification schemes.

Additionally, Table 4 presents the average price dispersion across industries for each technology level and Rauch category. We find that price dispersion is significantly bigger for Top-tech industries than for Bottom-tech ones but not convincingly different between the Rauch categories.[19] According to Table 1, this suggests that vertical differentiation is stronger across technology levels than across Rauch categories.

The next characteristic we consider is variety proliferation that is defined in equation (10). In table 4, we take the cross-product median number of varieties within

---

[17] Not shown in Table 4, we note that the *t*-statistics of the difference in means between -1.61 (Bottom-tech mean) and -1.32 (Top-tech mean) is -6.70 .

[18] Khandelwal (2010) also finds that quality ladders are increasing with the industry R&D intensity.

[19] Not shown in Table 4, we note that the *t*-statistics of the difference in means between 1.56 (Medium-low mean) and 1.94 (Medium-high mean) is -6.01 . And the *t*-statistics of the difference in means between 1.49 (Bottom-tech mean) and 1.97 (Top-tech mean) is -8.68 .

each industry, and then compute the cross-industry average of medians for each tech level and Rauch category. The number of varieties per product is increasing in the Rauch categories. Recalling Table 1 above, this pattern of variety proliferation suggests that the Rauch classification captures increasing horizontal differentiation along its scheme. The same cannot be said about the OECD-STI taxonomy, because the number of varieties is biggest for the Medium-low technology level.[20]

One might conjecture that the number of HS varieties behind each SITC 5-digit product simply reflects the way the HS and the SITC classifications where built, with some products naturally disaggregating into a greater number of varieties. Put another way, can we rely on a number of HS varieties comparison between two different SITC products?[21]

To give our variety results additional support we offer an alternative measure, which consists of regarding each pair (product, exporting country) as a distinct variety, like in Broda and Weinstein (2006). Table 4 shows that the number of exporting countries per product increases with the Rauch category. This is compatible with what we have seen for the number of varieties and horizontal differentiation in a broad sense. The opposite happens along technology levels: the Medium-low-tech displays the largest number of exporters per product.[22]

The last characteristics shown in Table 4 are factor intensities. They are potentially relevant to identify vertically differentiated industries because exporting countries'

[20] Notice that our results concerning variety proliferation corroborate Kugler and Verhoogen's (2012) choice of the Rauch classification as a proxy to horizontal differentiation, and of R&D and advertising intensity as a proxy to vertical differentiation. Drawing on Shaked and Sutton (1987), Beath and Katsoulacos (1991) further associate finiteness (a limited number of varieties) and vertical differentiation to R&D intensive goods.

[21] We thank the editor for this conjecture.

[22] Not shown, the *t*-statistics of the difference in between the Medium-low- and High-tech groups is 4.94 .

physical and human capital endowments are associated with higher quality products (see Khandelwal 2010 and Schott 2004). That is, besides quality improving R&D investments, physical and human capital can be at the origin of vertical differentiation.

As mentioned, we use Cadot's et al. (2010) RFI index to construct measures of physical and human capital intensities for the 5-digit products and 4-digit industries. There is a clear difference between the taxonomies. While both physical and human capital intensities are almost identical across Rauch categories, they are monotonously increasing in the technology levels.[23]

These results for factor intensities should not be surprising. Based on the trading criterion of whether products have prices quoted or not, the Rauch classification is not directly related to supply-side conditions and production technologies. On the other hand, the OECD-STI classification is based on R&D intensity, and R&D is often regarded as complementary to physical and human capital (see Howitt and Aghion 1998, and Frantzen 2000, among others).

Table 5 summarizes this section's results, showing how product market characteristics behave across each classification.

[23] Using the NBER-CES database (https://www.nber.org/research/data/nber-ces-manufacturing-industry-database) to calculate capital-labor ratios for the US industrial sectors in the same period as our sample's, the picture we get is: the Medium Low-tech level displays a cross-NAICS sectors average capital intensity of US$ 153,940 per worker (with US$ 11,578 standard deviation), the Medium-tech level of US$ 317,162 per worker (with US$ 20,469 std. dev.), the Medium High-tech of US$ 363,213 per worker (with US$ 36,091 std. dev.) and the High-tech of US$ 322,559 per worker (with US$ 23,924 std. dev.). The Rauch Homogeneous category displays an average capital intensity of US$ 510,100 per worker (with US$ 29,445 std. dev.) and the Differentiated category of US$ 191,141 per worker (with US$ 8,091 std. dev.). We thank Professor Wayne B. Gray for helping us with the NBER-CES database.

Table 5 - Behavior Along the Classification Sc

Product/Market characteristic	*OECD-STI*	*Rauch*
Price elasticity	↓	↓
Quality ladder	↑	↑
Price dispersion	↑	—
Number of varieties	↓	↑
Number of exporting countries	↓	↑
Physical capital intensity	↑	—
Human capital intensity	↑	—

Notes: This table summarizes what happens to product/market characteristics as one moves: (i) from low-tech to high-tech products; or (ii) from Rauch homogeneous to Rauch differentiated products.

Leaving the comparison with the Rauch classification aside, the results for the OECD-STI taxonomy allow us to decide whether real-world R&D efforts are directed mainly at quality improvement, variety proliferation, or cost reduction. Cost reduction (or cost competition) is important in markets with high price elasticity of demand, which we have seen is not a characteristic of high-tech, R&D intensive products. Also, these products display a small number of varieties.

When we put together the long quality ladders, a big price dispersion and a small number of exporting countries that characterize high-tech products, what stands out is Sutton's (1998) "escalation" view of R&D as a strategic investment in product quality that is aimed at escaping competition and that leads to industry concentration.

5. The OECD-STI *X* Rauch Matrix

So far, we have studied how market characteristics behave, considering each classification separately. In this section, we double sort products by OECD-STI technology level and Rauch category. By doing so, do we find empirical counterparts of the theoretical cases depicted in Table 1? For example, given our earlier conclusions that the OECD-STI captures vertical differentiation, and the Rauch classification captures horizontal differentiation, do high-tech homogeneous products fit in the pure vertical differentiation top left case of Table 1?

Before proceeding, we briefly describe an aggregation we do in order to simplify the analysis. As we can see in Table 4, with respect to many characteristics Medium Low-tech industries are not statistically significantly different from Medium-tech industries, just as Medium High-tech industries are not significantly different from High-techs. Therefore, in this section we group the first two technology levels into a "Bottom" technology level, and the next two into a "Top" technology level. While this grouping does not change our qualitative results, it makes a joint OECD-STI X Rauch analysis more tractable by reducing the 4 X 2 sorting to a 2 X 2 dimension.

Consider thus Table 6, that brings the joint OECD-STI–Rauch frequency distribution of SITC 4-digit industries. Homogeneous industries represent approximately 25% of the total, and Differentiated industries represent approximately 75% at both the Top and Bottom technology levels. Pearson's Chi-Squared test does not allow us to reject independence between the Rauch and the OECD-STI classifications at conventional significance levels, which rejects Rauch's (1999, page 27) conjecture that differentiated goods are also more technologically sophisticated.

Table 6 - Contingency Table (OECD-STI x Rauch Independence), Industry-level

	Homogeneous	*Differentiated*	Total	
Bottom-tech	28% (68)	72% (173)	100% (241)	
Top-tech	23% (60)	77% (201)	100% (261)	
Total	25% (128)	75% (374)	100% (502)	
Peason's Chi-Squared statistics				1.89
p-value				0.17

Notes: 502 industries with significant (at 95% confidence) IV price coefficient (out of 552 industries described in Table B.1) are included in this table. Total Value is in constant 2006 million US$. Pearson's Chi-Squared test of independence between Rauch and OECD-STI technology categories.

The goal of this section is to understand the pattern of change in product market characteristics when we move across the joint distribution of the two classifications. Besides within-group average characteristics, Table 7 also presents *t*-statistics of the tests for equality between the averages of two groups' characteristics when moving across OECD-STI technology levels holding the Rauch category constant, or as we move across Rauch categories holding the OECD-STI tech level constant.

Table 7 - OECD-STI x Rauch Grid

	Grouping by:			
Characteristics:	***Bottom-homogeneous***	***Bottom-differentiated***	***Top-homogeneous***	***Top-differentiated***
	7.1. Means of the Market Characteristics and Tests			
IV Price Coeff.	-1.74	-1.55	-1.64	-1.22
	(0.08)	(0.03)	(0.06)	(0.03)
	H_0: B-h = B-d -2.29		H_0: T-h = T-d -6.34	
	H_0: B-h = T-h -1.02		H_0: B-d = T-d -7.57	
Quality Ladder	2.17	2.32	2.33	3.27
	(0.11)	(0.06)	(0.12)	(0.11)
	H_0: B-h = B-d -1.14		H_0: T-h = T-d -5.77	
	H_0: B-h = T-h -0.96		H_0: B-d = T-d -7.72	
Price dispersion	1.34	1.55	1.94	1.98
	(0.06)	(0.03)	(0.11)	(0.05)
	H_0: B-h = B-d -2.98		H_0: T-h = T-d -0.29	
	H_0: B-h = T-h -4.77		H_0: B-d = T-d -7.1	
Median Number of HS Varieties per 5-digit product	4.07	9.55	3.97	5.67
	(0.51)	(1.00)	(0.53)	(0.57)
	H_0: B-h = B-d -4.9		H_0: T-h = T-d -2.2	
	H_0: B-h = T-h -0.15		H_0: B-d = T-d -3.38	
Number of exporting countries per 5-digit product	47.97	76.40	43.17	55.06
	(2.36)	(2.42)	(1.69)	(1.48)
	H_0: B-h = B-d -8.41		H_0: T-h = T-d -5.29	
	H_0: B-h = T-h -1.65		H_0: B-d = T-d -7.52	
Median Physical K Intensity (in dollars per worker)	102,265	90,109	131,917	141,301
	(5,855)	(3,295)	(4,937)	(2,559)
	H_0: B-h = B-d -1.81		H_0: T-h = T-d -1.69	
	H_0: B-h = T-h -3.87		H_0: B-d = T-d -12.27	
Median Human K Intensity (in years of schooling)	8.52	7.98	9.00	9.14
	(0.16)	(0.10)	(0.12)	(0.06)
	H_0: B-h = B-d -2.83		H_0: T-h = T-d -0.99	
	H_0: B-h = T-h -2.34		H_0: B-d = T-d -9.81	
	7.2. Number of observations and values			
Number of Observations	22,016	86,531	14,802	52,392
Total Value	240,087	638,604	137,314	1,669,717
Number of 4-digit industri	68	173	60	201
Number of 5-digit product	233	498	251	589
Number of Countries	144	159	132	157

Notes: 502 industries with significant (at 95% confidence) IV price coefficient (out of 552 described in Table B.1) are included in this table. Equation (6) in the text is estimated for each 4-digit industry. The variables' definitions are in Appendix A. Means of "Bottom-homogeneous" (B-h), "Bottom-differentiated" (B-d), "Top-homogeneous" (T-h) and "Top-differentiated" (T-d) are calculated from the cross-section distribution of the 4-digit industries in each tech-level group. Standard-errors of "Mean" are in parentheses. The *t*-statistics of "Mean" uses the standard error of mean; i.e., the cross-section standard error divided by the square root of the number of 4-digit industries included in the mean. "H0: XX=YY" rows report t-statistics of equal means between XX and YY. Assuming independence between groups, the standard error of the difference in means is the square root of the sum of the variances of the average slopes of the compared groups.

In Table 7, as we move from OECD-STI Bottom-tech to Top-tech holding the Rauch category constant at Homogeneous, there is no significant difference between the average price coefficients or quality ladders. Now, as we move from OECD-STI Bottom-tech to Top-tech holding the Rauch category constant at Differentiated, the differences become very salient. We get significantly less negative price elasticities (with a -7.57 *t*-statistics), and significantly longer quality ladders (with a -7.72 *t*-statistics).

What do these findings mean? They show that R&D has distinct effects on markets of homogeneous and differentiated goods. R&D investment in homogenous goods industries neither leads to lower price elasticities nor to longer quality ladders but to more price dispersion (with a -4.77 *t*-statistics). This evidence indicates that R&D in homogeneous industries is used mainly for cost reductions and not for vertical differentiation. Price dispersion arises due to differences in costs among firms and countries.

Different from the homogenous case, R&D investment in differentiated goods markets means lower price elasticities and longer quality ladders. Horizontal distance on the product space is a force that reduces the price elasticity, and the characteristic space is large for these markets. Here, R&D leads to vertical differentiation[24], and the bigger price dispersion when we move from Bottom to Top-techs (with a 7.10 *t*-statistics) reflects quality differences.

As an example, we can see in Table 2 that the Top-homogenous segment consists mostly of chemicals' industries for which the scope for differentiation is narrow. In fact, that is what defines a homogenous good industry. In contrast, electronics such as

[24] Because differentiated industries represent 75% of the industries in our dataset, this type of R&D is the prevalent one, which justifies our conclusions regarding the OECD-STI classification at the end of section 4; namely, that real-world R&D efforts are directed mainly at improving quality.

computers etc., where the scope for differentiation is wide, are mostly on the Top-differentiated group. The data show that R&D has distinct purposes in these markets.

How does Table 7 help to empirically disentangle horizontal and vertical differentiation? It shows that contrary to intuition, the Top-homogeneous industries do not correspond to the pure vertical differentiation, top left case of Table 1. In the data, there is no vertical differentiation without horizontal differentiation.

The Top-differentiated entries in Table 7 show that this group is a mix of horizontal and vertical differentiation that corresponds to the top right case of Table 1. It displays the smallest price-elasticity, but with the longest quality ladder and the greatest price dispersion, and smaller number of varieties than the Bottom-differentiated case.

Bearing in mind that our identification strategy takes the Bottom-homogeneous group as the no-vertical and no-horizontal differentiation bottom left case of Table 1, there remains judging the Bottom-differentiated group. This group is pure horizontal differentiation: short quality ladder and small price dispersion but price elasticity substantially smaller than the Bottom-homogeneous'. This group also displays the largest numbers of HS varieties and exporting countries. It corresponds to the Sutton's (1998) flowmeters, bottom right case of Table 1.

Overall, the above results indicate that the OECD-STI taxonomy shows variations in market characteristics beyond the Rauch classification. This is particularly the case for the too coarse Rauch differentiated category, which responds for 75% of the industries in our sample. There, the variation in product market characteristics as we move from Bottom to Top-techs is clearly indicative of increased vertical differentiation.[25]

---

[25] One might conjecture if the means in Table 7 are not hiding heterogeneity within groups, and if these patterns hold for important sectors and industries as well. In Appendix C, we group the 4-digit SITC industries into 20 ISIC sectors, and report average product market characteristics for them. Broadly, the

6. Alternative R&D intensity measures

The OECD-STI taxonomy is an ordinal classification based on the cross-industry distribution of R&D intensities. Originally, it was 4-partite, but in the last section we have further simplified it to a 2-partite division of Top-techs and Bottom-techs. Is it too coarse? Should we use more disaggregated, quantitative measures of R&D intensity directly?

To answer those questions, we compare the OECD-STI taxonomy with three alternative classifications. The first is the OECD's R&D intensity measure, reported by Galindo-Rueda and Verger (2016), and shown in the last column of Table C of Appendix C. It is defined at the disaggregation level of the ISIC sector, and here we use the median R&D expenditure (across OECD member countries) as a percentage of the value-added. The second is Nunn and Trefler's (2013) much more disaggregated measure of R&D intensity. Based on the Orbis database, it is defined at the HS disaggregation level, which we concord with the 4-digit SITC industries. The third is the Rauch (1999) full classification, sorting products into "Organized", "Reference Priced" and "Differentiated".

---

patterns detected for the four-group grid analysis also hold for the ISIC sectors classified by technology level and Rauch category.

Table 8 : Spearman's Rank Correlations with alternative taxonomies

	OECD-STI 4-levels	Rauch's 3-levels	OECD R&D	Nunn's R&D	2-levels OECD-STI	2-levels Rauch
	8.1. Between taxonomies					
Rauch's 3-levels	0.02 (0.69)					
OECD R&D intensity	0.84 (0.00)	0.27 (0.00)				
Nunn's R&D intensity	0.62 (0.00)	0.11 (0.02)	0.54 (0.00)			
2-levels OECD-STI	0.91 (0.00)	0.08 (0.08)	0.87 (0.00)	0.57 (0.00)		
2-levels Rauch	0.01 (0.84)	0.99 (0.00)	0.26 (0.00)	0.10 (0.04)	0.07 (0.16)	
	8.2. Taxonomies with product market characteristics					
Price elasticity	0.30 (0.00)	0.27 (0.00)	0.25 (0.00)	0.20 (0.00)	0.32 (0.00)	0.25 (0.00)
Quality ladder (interquartile)	0.36 (0.00)	0.24 (0.00)	0.32 (0.00)	0.20 (0.00)	0.32 (0.00)	0.22 (0.00)
Price dispersion (interquartile)	0.35 (0.00)	0.12 (0.01)	0.38 (0.00)	0.24 (0.00)	0.38 (0.00)	0.11 (0.02)
Median no. of HS Varieties per product	-0.19 (0.00)	0.19 (0.00)	-0.03 (0.56)	-0.17 (0.00)	-0.15 (0.00)	0.19 (0.00)
No. of exporting countries per product	-0.36 (0.00)	0.26 (0.00)	-0.16 (0.00)	-0.22 (0.00)	-0.33 (0.00)	0.27 (0.00)

Notes : The number of observations for all correlations is 461, corresponding to the SITC 4-digit industries which are classified in the four taxonomies. p-values are in parentheses. The 4-digit estimates of the demand price-elasticities and quality ladders in Panel 2 come from the regressions in Table B.1.

Table 8 shows the Spearman correlations between the classifications in Panel1, and between a given classification and product market characteristics in Panel 2.[26] To use

[26] The Spearman's correlation ranks each 4-digit SITC industry by the product market characteristic in question and by the R&D intensity. Given that the OECD-STI taxonomy is an ordinal variable, this is the natural correlation to use for a comparison with alternative R&D intensity measures.

a common sample, the correlations in Table 8 are based on the 461 4-digit SITC industries included in the four classifications.

In Panel 1, we confirm that the three R&D intensity-based classifications are highly correlated: that is, the 4-levels OECD-STI, the OECD R&D intensity measure, and Nunn and Trefler's. Most importantly, the correlations between the OECD-STI and the Rauch classifications are very low – an insignificant 0.07 between the respective two levels used in this paper.

Panel 2 shows that the OECD-STI 2-level taxonomy is more correlated with price elasticity, the quality ladder, and price dispersion than the Rauch's, or than the more disaggregated Nunn and Trefler's and OECD R&D intensities measures. The number of varieties is more strongly correlated with the Rauch classification that again indicates that it captures horizontal differentiation. The negative correlations of the OECD-STI with the numbers of varieties and exporting countries are also in accordance with Shaked and Sutton's (1987) narrative of escalation and concentration.

In sum, Table 8 supports that the ordinal OECD-STI 2-level classification is better than, or at least as appropriate as, the alternative R&D intensity measures to systematize cross-industry variation in product market characteristics. And the variation it captures is complementary to the Rauch classification.

7. Conclusions

Using a sample of US imports, we investigate how product market characteristics vary across the R&D intensity-based "technology levels" of the OECD-STI classification. The motivation to test this taxonomy comes from the fact that Sutton (1998) recommends using the R&D intensity as a proxy for the scope of vertical differentiation. However, the

Rauch classification is by far more widely used in the literature as a measure of vertical as well as horizontal differentiation.

We considered six product market characteristics: price elasticity of demand, quality ladders, price dispersion, number of product varieties, number of exporting countries, and factor intensities. While Fernandes and Paunov (2013) and other authors rely on price dispersion alone, we used quality ladders to identify vertical differentiation. It is precisely because the quality-ladders' estimation controls for variety that we can identify vertical differentiation.

When we consider the two classifications separately, the patterns of variation in product market characteristics indicate that the Rauch captures mainly horizontal but also some vertical differentiation. This is as expected because the Rauch classification only separates differentiated goods from commodities according to whether products possess quoted prices or not.

In contrast, the behavior of all market characteristics along the technology levels of the OECD-STI classification is consistent with vertical differentiation. This is suggestive of the "escalation mechanism" envisaged by Sutton (1998), where R&D is a strategic investment in product quality that leads to industry concentration.

However, when we double sort products by Rauch category and technology level, we find that R&D targets quality improvements and vertical differentiation only in the Rauch differentiated industries. Therefore, our recommendation for International Trade and Industrial Organization applications that seek to refine the empirical identification of vertical differentiation to not simply use the OECD-STI instead of the Rauch classification. Rather, R&D intensity-based classifications should be used after filtering out homogeneous goods.

REFERENCES:

Affendy, A., Yee, L., Satoru, M., 2010. Commodity-industry classification proxy: a correspondence table between SITC revision 2 and ISIC revision 3. MPRA Paper No. 27626.

Bastos, P. and Silva, J. (2010). Identifying Vertically Differentiated Products. *Economics* Letters 106, 32-34.

Beath, J., Katsoulacos, Y., 1991. The Economic Theory of Product Differentiation. Cambridge University Press.

Berry, S., 1994. Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics* 25, 242–262.

Broda, C., Weinstein, D., 2006. Globalization and the gains from variety. *Quarterly Journal of Economics* 121, 541-585.

Coibion, O., Einav, L. and Hallak, J.C., 2007. Equilibrium demand elasticities across quality segments. *International Journal of Industrial Organization* 25, 13-30.

Erkel-Rousse, H., Mirza, D., 2002. Import price elasticities: reconsidering the evidence. *Canadian Journal of Economics* 35, 282-306.

Feenstra, R., Romalis, J., Schott, P., 2002. U.S. imports, exports, and tariff data, 1989-2001. NBER Working Papers No 9387.

Fernandes, A. and Paunov, C. (2013), Does trade stimulate product quality upgrading?, *Canadian Journal of Economics* 46, 1232-1264.

Frantzen, D., 2000. R&D, Human Capital and International Technology Spillovers: A Cross-country Analysis. *The Scandinavian Journal of Economics* 102, 57-75.

Galindo-Rueda, F., Verger, F., 2016. OECD Taxonomy of Economic Activities Based on R&D Intensity. *OECD Science, Technology and Industry Working Papers*, 2016/04, OECD Publishing.

Hatzichronoglou, T., 1997. Revision of the high-technology sector and product classification. OECD Science, Technology and Industry Working Papers 02, OECD Publishing.

Howitt, P. and Aghion, P., 1998. Capital Accumulation and Innovation as Complementary Factors in Long-Run Growth. *Journal of Economic Growth* 3, 111-130.

Hummels, D., Klenow, P., 2005. The variety and quality of a nation's exports. *American Economic Review* 95, 704-723.

Johnson, R., 2012. Trade and prices with heterogeneous firms. *Journal of International Economics* 86, 43-56.

Khandelwal, A., 2010. The long and short of quality ladders. *Review of Economic Studies* 77, 1450-76.

Kugler, J., Verhoogen, E., 2012. Prices, plant size, and product quality. *Review of Economic Studies* 79, 307-339.

Lahmandi-Ayed, R., 2007. Finiteness property with vertical and horizontal differentiation: does it really matter? *Economic Theory* 33, 531-548.

Nunn, N., Trefler, D., 2013. Incomplete Contracts and the Boundaries of the Multinational Firm. *Journal of Economic Behavior & Organization* 94, 330-344.

Rauch, J., 1999. Networks versus markets in international trade. *Journal of International Economics* 48, 7-35.

Schmalensee, R., 1978. Entry Deterrence in the Ready-to-Eat Breakfast Cereal Industry. *Bell Journal of Economics* 9, 305-327.

Schott, P. "Across-Product Versus Within-Product Specialization in International Trade, *Quarterly Journal of Economics* 119, 2004, 647-678.

Shaked, A. and Sutton, J., 1987. Product Differentiation and Industrial Structure. *The Journal of Industrial Economics* 36, 131-146.

Sutton, J., 1998. Technology and Market Structure: theory and history. The MIT Press.

APPENDICES

Appendix A: Variables Construction

The variables available at CID-econ US import database are explained in detail in Feenstra et al. (2002). X_{ijt} is the quantity of 5-digit SITC product i exported by country j to the US. in year t; V_{ijt} is the corresponding total value (in current US$) deflated by the U.S. GDP deflator (from http://www.econstats.com/weo/V005.htm), not including transportation costs or tariffs; $charge_{ijt}$ corresponds to transportation costs (in constant US$); and $duty_{ijt}$ is the total value (also in constant US$) levied by importation tariffs' application. With these variables, we construct the unit value variable which we call "price" in the main text:

$$P_{ijt} = \frac{V_{ijt} + charge_{ijt} + duty_{ijt}}{X_{ijt}}. \qquad (A.1)$$

And, dividing by cross-country averages, get the relative price used as a trimming criterion, as explained in section 3:

$$RUV_{ijt} = \frac{P_{ijt}}{{M_{it}}^{-1}\sum_{n=1}^{M_{it}} P_{int}}. \qquad (A.2)$$

where n denotes an exporter country (to the US) and there are M_{it} different exporters of product i in year t.

Using the variable $charge_{ijt}$ from the CID-econ database, we also construct a measure of unit transportation costs, to be used as an instrument in the estimation of equation (6):

$$transp_{ijt} = \frac{charge_{ijt}}{X_{ijt}} \, . \qquad (A.3)$$

We take countries' factor endowments from Cadot et al. (2010) database: $PHYSK_j$ and $HUMK_j$ denote respectively country j's physical capital (in constant dollars) per worker and human capital per worker, this latter being actually the average schooling years for the population aged 15 or more, as in Barro and Lee database.

In order to construct 5-digit products' factor intensities, we use the Cadot's et al. (2010) revealed factor intensity index:

$$PHYSKINT_{it} = \sum_{j=1}^{M_{it}} \omega_{ijt} \cdot PHYSKINT_{jt} \, , \qquad (A.4)$$

where:

$$\omega_{ijt} = \frac{V_{ijt}/V_{jt}}{\sum_{k=1}^{M_{it}} V_{ikt}/V_{kt}} \, . \qquad (A.5)$$

So that 5-digit product i's physical capital intensity is a weighted average of the physical capital endowments of i's exporting countries. The weights ω_{ij} are such that V_{ij}/V_j is the value of i's exports by country j divided by the total value of j's exports (of all products). In order to calculate the weights ω_{ij} we resort to the UN-Comtrade database, so here we are considering countries' exports to the whole world, not only to the United States.[27] Notice also that $\sum_{j=1}^{M_{it}} \omega_{ijt} = 1$.

[27] Here, the number of countries which are product i's exporters, M_{it} , coincides with the number of countries for which we have factor endowments data, from Cadot et al. (2010).

Analogously, for human capital intensities:

$$HUMKINT_{it} = \sum_{j=1}^{M_{it}} \omega_{ijt} \cdot HUMKINT_{jt} \,. \qquad (A.6)$$

Appendix B: Demand Estimates

For each SITC 4-digit industry panel, we first estimate equation (6) by Instrumental Variables (IV) with transportation costs and real exchange rates as instruments to account for price endogeneity; [28] and also by Ordinary Least Squares (OLS) for reference.

Table B.1 presents the estimates under the labels of "OLS" and "Unrestricted IV". It reports averages and medians, across all 4-digit industries, of the estimated slopes (σ_N and σ_P) and some descriptive statistics. We see that exported quantities depend positively on variety, with coefficients close to one, confirming the trade literature assumption of "love of variety". Like in Erkel-Rousse and Mirza (2002), as we move from the OLS to the IV estimates, we get higher price elasticities. Since our price variable is the CIF price plus tariffs, it approximates what consumers pay for; and price elasticities greater than one are indeed expected, provided exporting firms operate on the elastic portion of the demand curve in equilibrium.

By imposing the restriction $\sigma_N = 1$ in (6), labeled simply as "IV Price Coeff." in Table B.1, the IV price coefficients estimates almost do not change, and 91% of the 4-digit industries' price elasticity estimates have absolute *t*-statistics higher than 1.96. The

[28] The real exchange rates relative to the U.S. are from http://bruegel.org/2012/03/real-effective-exchange-rates-for-178-countries-a-new-database/.

rejections of H_0s in under-identification and weak-identification tests confirm that this "restricted" IV model is identified and not weakly identified.

Table B.1 - 4-digit Industries Demands Estimations Summary

	Mean	*Median*	*1st quartile*	*3rd quartile*
OLS Price coefficient	-1.08	-1.09	-1.30	-0.89
	(-6.88)	(-6.23)	(-9.04)	(-4.09)
OLS Variety coefficient	0.92	0.99	0.57	1.31
	(2.90)	(2.46)	(1.37)	(3.94)
Unrestricted IV Price Coeff.	-1.42	-1.38	-1.71	-1.11
	(-7.28)	(-6.85)	(-9.53)	(-4.32)
Unrestricted IV Variety Coeff.	0.85	0.93	0.44	1.24
	(3.32)	(2.96)	(1.61)	(4.62)
IV Price Coeff.	-1.41	-1.38	-1.73	-1.11
	(-7.27)	(-6.86)	(-9.54)	(-4.25)
1st. stage LM-test (underidentificatic	14.17	12.01	7.20	18.38
1st. stage F-stat. (weak indentificatic	37.44	22.34	9.86	43.85
Over identifying restric. p-value	0.46	0.44	0.18	0.71
Observations per estimation	327	244	120	431
Significant IV Price Coeff.	-1.46	-1.40	-1.75	-1.16
t-stat. of Signif. IV Price Coeff.	(-7.88)	(-7.29)	(-9.92)	(-4.97)
	Overall		***Significant IV Price Coeff.***	
Price coeff.s with |*t*-stat.| > 1.96	0.91			
Number of industries		552		502
Total obs. across all industries		180,317		174,248

Notes: Equation (6) in the text is estimated for each 4-digit industry. The variables' definitions are in Appendix A. The "mean", "median, "1st quartile" and "3rd quartile" are from the cross-section distribution of all 4-digit industries in the sample. *t*-statistics are in parentheses, computed from standard errors clustered by exporting country.

Because there are alternative ways to test for heterogeneity across groups in the price elasticities of demand – rather than our choice of estimating elasticities for each 4-digit industry and then testing differences in means across technology or Rauch levels –, Table B.2 tests differences in the elasticities of the different groups in a panel of all observations from 4-digit industries for which the IV price coefficient is significant in Table B.1. Common price coefficients by group are estimated jointly. Because the correlations across industries and between groups are accounted in the panel estimation, hypothesis testing of price coefficients significance and equality are standard *t*-test and *F*-test. To facilitate comparison with the difference in means tests from Table 7, in Panel B.2.2, we compute *t*-statistics in parentheses that are the square root of the *chi2(1)* from the *F*-statistics of equal price elasticity.

Table B.2 - IV Panel Demand Estimates

B.2.1. Demand estimates and t-statistics

	Botom-Homogeneous	***Bottom-Differentiated***	***Top-Homogeneous***	***Top-Differentiated***
Mean IV Price Coeff.	-1.61	-1.49	-1.63	-1.13
	(-40.66)	(-83.77)	(-32.66)	(-63.08)

B.2.2. F-statistics of equal price elasticities and p-values

	chi2(1) =	*p-value =*
H_0: Bot-Hom = Top-Hom	0.23	0.63
	(0.48)	
H_0: Bot-Dif = Top-Dif	250.67	0.00
	(15.83)	
H_0: Bot-Hom = Bot-Dif	9.99	0.00
	(3.16)	
H_0: Top-Hom = Top-Dif	102.76	0.00
	(10.14)	
H_0: Bot-Hom = Top-Dif	150.95	0.00
	(12.29)	
H_0: Bot-Dif = Top-Hom	8.06	0.00
	(2.84)	

Notes: Data from the 502 4-digit industries with significant (at 95% confidence) IV price coefficients are included in this panel. Equation (6) is estimated with different price-elasticity coefficients representing each group ("Bottom-Homogeneous", "Bottom-Differentiated", "Top-Homogeneous" and "Top-Differentiated") in a panel that pools all observations, with 5-digit-product-country and 4-digit-industry-time fixed-effects. Robust standard-errors clustered by country. *t*-statistics are in parentheses. In the second panel, the *t*-statistics in parentheses are the square root of the *chi2(1)* from the *F*-test of equal coefficients.

Appendix C: Going more disaggregated - Product Market Characteristics for ISIC sectors

In Table C, we group 469 of the 502 4-digit SITC industries from the sample into 20 ISIC sectors and report their average product market characteristics, technology level and Rauch category. Given that the OECD-STI taxonomy sorts ISIC industries, each ISIC sector is unambiguously in a tech-level. However, the Rauch classification applies to SITC 4-digit industries, and an ISIC sector can gather both homogeneous and differentiated industries. This is the case of 12 sectors, for which we present market characteristics for both Rauch categories. The numbers in parentheses of column 4 are the proportions of 4-digit industries belonging to each Rauch category within the ISIC sector.

Table C broadly conveys the same "grid patterns" as in Table 7. Within ISIC sectors, as we move from Rauch homogeneous to differentiated, in general, we get lower price elasticities (in 9 out of 12 sectors), longer quality ladders (in 8 sectors), and bigger price dispersions (in 9 sectors). Among the Rauch differentiated industries, we notice that the shortest quality ladder is for "leather and related products" (a Bottom-tech ISIC sector) and the longest is for "machinery and equipment, nec" (a Top-tech ISIC sector). Among the Rauch homogeneous industries, the shortest quality ladder is for "fabricated metal products except weapons and ammunition" (a Bottom-tech ISIC sector) and the longest is for "machinery and equipment, nec" (a Top-tech ISIC sector).

Among Rauch differentiated industries, the highest price elasticity is for a Bottom-tech ISIC sector ("leather and related products"). Among the Rauch homogeneous industries, the highest price elasticities are for "fabricated metal products except weapons and ammunition" and "textiles", both Bottom-tech ISIC sectors.

Table C - Product market characteristics by ISIC sector

ISIC code	description	tech level	Rauch category	Price elast. (abs. value)	Quality ladder (interquartile)	Price dispersion	OECD R&D intensity (%)
26	computer, eletronic and optical products	Top	Dif (97%)	-1.40	2.60	1.17	19.92
			Hom (3%)	-1.25	2.72	1.23	
21	pharmaceuticals	Top	Dif (88%)	-1.90	3.05	2.34	13.57
			Hom (12%)	-1.67	2.26	1.94	
252	weapons and ammunition	Top	Dif (100%)	-1.30	2.70	1.69	11.53
29	motor vehicles, trailers, etc.	Top	Dif (100%)	-1.66	2.68	1.31	6.11
27	electrical equiment	Top	Dif (100%)	-1.25	3.25	6.28	5.45
28	machinery and equipment, nec	Top	Dif (98%)	-0.97	3.95	1.90	4.85
			Hom (2%)	-1.70	2.87	1.84	
325	medical and dental instrumments	Top	Dif (100%)	-2.19	1.67	1.91	4.31
20	chemicals	Top	Dif (30%)	-1.42	2.34	1.45	3.54
			Hom (70%)	-1.55	2.39	0.87	
22	rubber and plastic	Bottom	Dif (45%)	-1.51	2.71	1.90	2.28
			Hom (55%)	-1.91	2.09	0.99	
301	building of ships and boats	Bottom	dif (100%)	-0.95	3.32	4.34	2.22
13	textiles	Bottom	Dif (78%)	-1.85	1.80	1.04	1.78
			Hom (22%)	-2.13	1.48	1.55	
14	wearing apparel	Bottom	Dif (100%)	-1.59	2.41	1.52	1.71
24	basic metals	Bottom	Dif (19%)	-1.31	3.04	1.90	1.64
			Hom (81%)	-1.69	2.40	0.91	
25X	fabricated metal products except weapons and ammunition	Bottom	Dif (95%)	-1.29	2.39	2.19	1.39
			Hom (5%)	-2.21	0.98	1.86	
23	other non-metallic mineral products	Bottom	Dif (77%)	-1.22	2.44	2.82	1.22
			Hom (33%)	-1.54	2.31	1.82	
15	leather and related products	Bottom	Dif (86%)	-2.26	1.38	1.76	1.04
			Hom (14%)	-1.91	1.05	1.22	
31	furniture	Bottom	Dif (100%)	-1.21	2.99	1.51	0.98
17	paper and related products	Bottom	Dif (34%)	-1.39	2.24	1.96	0.95
			Hom (66%)	-1.43	2.40	1.27	
16	wood and cork	Bottom	Dif (50%)	-1.13	2.51	1.26	0.7
			Hom (50%)	-1.15	2.68	1.50	
18	printing and reproduction of	Bottom	Dif (100%)	-1.44	1.85	1.07	0.5

Notes: a characteristic (quality ladder, for example) for a given 2-digit ISIC sector is the average of that characteristic across the 4-digit SITC industries that belong to the ISIC sector. Out of 502 industries with significant demand price elasticities in the sample, 469 4-digit industries are included in this table, as the default "32X - other manufacturing except medical and dental instruments" ISIC sector is left out. In parentheses is the proportion of 4-digit industries classified in each Rauch category within the ISIC sector.