The Evolution of Comparative Advantage: The Impact of the Structure of the Product Space

Ricardo Hausmann
Bailey Klinger

Center for International Development and
Kennedy School of Government
Harvard University

June 2006

Valuable comments have been provided by Philippe Aghion, Laura Alfaro, Albert-Lazlo Barabasi, Olivier Blanchard, Ricardo Caballero, Oded Galor, Elhanan Helpman, Cesar Hidalgo, Asim Khwaja, Robert Lawrence, Daniel Lederman, Lant Pritchett, Roberto Rigobon, Andres Rodriguez-Clare, Dani Rodrik, Charles Sabel, Ernesto Stein, Federico Sturzenegger, & David Weil. All errors & omissions are our own.

Comments: ricardo_hausmann@harvard.edu & bailey_klinger@ksgphd.harvard.edu

1. Introduction

What determines the evolution of a country's comparative advantage across products? Or put another way, what governs the pattern of structural transformation? Does the initial pattern of specialization affect its future evolution? Does specialization in e.g. electric toaster-ovens impact the future evolution of a country differently than if it had started with coffee instead?[1]

The foundational models of trade say no. The Heckscher-Ohlin model suggests that in an open economy, a country's pattern of production will depend on its relative factor endowments. Over time, changes in these underlying endowments –through the accumulation of physical, human and institutional capital – will be reflected in a change of the export mix. Controlling for factor endowments, the initial pattern of specialization has no independent effect on future comparative advantage, as it is just one way of exporting the relative factor abundance[2]. The Ricardian model argues that technological differences across countries determine comparative advantage, and changes in the product mix will therefore depend on the relative evolution of productivity across products. To make such a model say something on our question, it is necessary to complement it with some story of what drives productivity growth. The two dominant approaches – the varieties model (Romer 1988) using Dixit-Stiglitz production functions and the quality ladders model (Aghion-Howitt, 1992, Grossman-Helpman (1991) – assume a degree of homogeneity across products that eliminates the possibility to capture the impact of initial specialization[3]. For example, in a varieties model, initial specialization in coffee or microwave ovens do not affect the expected productivity levels in the production flat-panel TVs. In Aghion-Howitt (1992) quality improvements happen across all products at the same time while in Grossman and Helpman (1991) they happen independently in each product.

In these models, the structure of the product space is of no importance and hence does not create sources of path dependence. In this paper, we seek to examine the product space empirically. Our main finding is that changes over time in the revealed comparative advantage of individual nations are associated with the pattern of relatedness across products. As countries change their export mix there is a strong tendency to move towards related goods rather than to goods that are less related. To establish this we first develop an outcomes-based measure of the relatedness between pairs of products using cross-country export data. We find that the space of relatedness or what we call the product space is highly heterogeneous: there are very dense parts of the product space with highly inter-connected products, and goods that are in very sparse sections of the

[1] Hausmann, Hwang and Rodrik (2006) argue that the pattern of specialization affects future growth. They show that initial specialization in goods exported by countries with higher income are associated with faster growth. In this paper we shall not look at the implications of our findings for growth.

[2] For example, in Leamer (1987) countries find a cone of diversification, i.e. a linear combinations of goods that produced in the right combinations exhaust the resource endowment. As relative endowments change, the country initially changes only the relative proportions of the goods in the cone of diversification. Larger changes involve the abandonment of one or more goods and a change in the cone of diversification.

[3] Models such as Matsuyama (199?) assume exogenously the difference in the productivity patterns across different sectors by assuming that some offer increasing returns while others do not.

products space[4]. Moreover we establish that the pattern of relatedness across products is only weakly determined by broad factor intensities, at least as expressed in the 10-category Leamer (1984) classification. Relatedness is also weakly related to the Lall (2000) levels of technological sophistication[5].

Furthermore, we find highly robust evidence that the evolution of comparative advantage in a country is significantly affected by these patterns of relatedness. Countries develop comparative advantage preferentially in nearby goods. A particular product's proximity to existing areas of comparative advantage is one of the most significant determinants of whether a country will develop an advantage in that product in the future.

These two findings, that the product space is highly heterogeneous and that it regulates the evolution of comparative advantage, implies that a country's current location in the product space significantly affects its opportunities for future productive transformation.

After reviewing the trade literature and showing how it does not account for these observations, in section 2 we propose a model that explains these findings in a very straightforward way. In this model, human capital is highly product-specific, and imperfectly substitutable. The specific human capital required to produce one good is an imperfect substitute for that required to produce another good, with the degree of substitutability determining product relatedness[6]. This causes the process of structural transformation to favor nearby goods in the product space.

In section 3 we develop our empirical measure of distance, and show some of its characteristics. We measure the relatedness between pairs of products based on the probability that countries in the world export both. This is a purely outcomes-based measure of revealed relatedness that is agnostic as to its source. It does not depend on *a priori* identification of the drivers of relatedness, such as factor intensity or technological sophistication, although we show that such broad patterns capture little of the structure of the product space. We show this structure to be highly heterogeneous, with both dense and sparse areas. In section 4 we present highly robust evidence showing that this product-level relatedness governs the process of structural transformation. Section 5 concludes.

This paper attempts to dig deeper into the process of structural transformation and the determinants of the evolution of a country's export mix, which relates to several strands in the literature. As mentioned above, the work on quality ladders or variety models (Grossman and Helpman 1989 & 1991, Aghion and Howitt 1992) assume implicitly a perfectly homogeneous product space in the sense that the cost of developing a new variety is independent of the distance between old and new products[7].

---

[4] This finding contradicts the typical assumption of homogeneity assumed in models that use Dixit-Stiglitz production functions, Romer (1990), Grossman and Helpman (1991) or Aghion and Howitt (1998).

[5] It is also weakly related to interactions between Leamer (1984) and Lall (2000) categories.

[6] We describe the model in terms of human capital specificity. However, the specificity could be in physical capital, infrastructure, regulatory framework or property rights regime.

[7] Segerstrom (1991) introduces heterogeneity in the R&D technology across countries, but not in the pattern of imitation between products.

Some models focus on the determinants of the technology choices of firms which then govern country-level structural transformation. For example, Acemoglu Antras & Helpman (2006) stress the importance of contracting institutions on technology, by assuming that contracting institutions avoid hold-up problems that would prevent investment in specific intermediate goods. Better contracting institutions allow for more specialization, more intermediate inputs and higher productivity. However, contracting institutions are a one-dimensional national characteristic and the production function is assumed to be Dixit-Stiglitz so that the product space is continuous and does not affect outcomes[8].

Our work is closest in spirit to Jovanovic and Nyarko (1996) who model learning by doing and technology upgrading at the individual level. In their model, experience provides agents with information that improves their productivity in the given technology (vertical shift). But gains in this dimension are limited, and agents must also 'jump' to new products (horizontal shift). The degree of the similarity of the new products to the old determines how transferable the accumulated knowledge is, with less similar products having a higher productivity loss. However, they assume that there is always a product at the right distance for the country to jump into. The heterogeneity of the product space suggests that this assumption may be unwarranted and we shall relax it. In this paper we do not focus on improvements within products, but instead concentrate on the varying distances between goods[9].

2. A Model of Structural Transformation and the Product Space

We develop a model of human capital that is product-specific, and the degree of substitutability across products is heterogeneous. The micro-foundations for such a model are in the spirit of Lazear (2003), who models firm-specific human capital as a weighted combination of skills specific to a firm, but which are an imperfect substitute for the combination of skills required in another firm. Generalizing this structure to the product-level by country is consistent with models of national learning-by-doing (e.g. NBER 1999)[10].

Consider a model of overlapping generations. In each period there is a young untrained worker and an old trained worker. In the first period, the young worker does not produce, but is trained by the old worker in the production of the specific product the old worker produces. That is, training is done through learning-by-watching. In the second period,

[8] The contracting environment is a potentially time-varying national characteristic. Our results are robust to the inclusion of country-year dummies implying that we capture something other than the logic implied by Acemoglu et al (2006). One alternative would be to assume that the contracting environment is product specific: property rights on oil and gas resources are very different to those for medicines or downloaded music. One could imagine a model where the pattern of relatedness across products emerged from the pattern of similarity in the requisite contracting environment.

[9] For within product improvements see Schott (2004), Hwang (2006).

[10] Relatively specific human capital is just one way to create product relatedness. Any relatively specific non-tradable asset would have the same effect such as infrastructure, property rights, regulations or any specific public good in general.

the young worker is now an old trained worker, and possesses the specific human capital required to produce the good in which she was trained. She can either produce that same good, or jump to an alternative good for which her product-specific human capital is an imperfect substitute. Given this choice, she will then train the newly born young unskilled worker in the production of that product.

Fixing the output of each skilled worker to 1, we can order goods on a line so that their price increases linearly with distance. Note that there is no reason for the product space to be continuous. The additional revenues earned by moving from the current good *i* to another good *j* are:

$$(1)\ \Delta P_{i,j} = f\delta_{i,j}$$

where $\delta_{i,j}$ is the distance from good *i* to good *j*, equal to 0 if *i=j*, and greater than 0 if *i≠j*.

While the price rises with distance, the substitutability of product-specific human capital decreases with distance, meaning production costs increase. The additional costs from moving from the current good *i* to another good *j* are:

$$(2)\ C(\delta_{ij}) = \frac{c\delta_{ij}^{2}}{2}$$

The trained worker therefore faces the following profit maximization problem:

$$(3)\ \max_{\delta_{i,j}} \Pi = f\delta_{i,j} - \frac{c\delta_{i,j}^{2}}{2}$$

with the optimal distance to jump being simply

$$(4)\ \delta^{*}_{i,j} = f/c$$

The profits from jumping to the new good are

$$(5)\ \Pi_{\delta^{*}_{i,j}} = f^{2}/2c$$

whereas the profits from remaining in good *i* are simply

$$(6)\ \Pi\delta_{i,j=1} = f - c/2$$

which implies that the trained worker will jump δ^{*} as long as $f > c$. More generally, the trained worker will not remain in good *i* as long as there exists a good at distance δ_{min} which satisfies

$$(7)\ \left(f\delta_{\min} - {c\delta_{\min}}^2\!/\!_2 \right) > \left(f - c\!/\!_2 \right)$$

This implies that, given the imperfect substitutability of human capital across products, the pattern of structural transformation is path-dependent. Incremental structural transformation is privately profitable as long as there aren't breaks in the product space, but stagnation will occur if there are breaks in the product space larger than $\delta_{\min}$. Stagnation due to breaks in the product space would represent a coordination failure when jumps that are not privately profitable are nevertheless socially optimal, since future firms benefit from the newly-created specific human capital.

This simple model places a great deal of importance on the distance between products. Our simple model is of distance along one dimension (where $\delta_{i,j} + \delta_{j,k} = \delta_{i,k}$), yet in a world of *n* products, the product space is only fully represented by an *n* x *n* matrix of the pairwise distances:

$$(8)\ \Delta = \begin{bmatrix} 0 & \delta_{1,2} & \delta_{1,3} & \cdots & \delta_{1,n} \\ & \ddots & \delta_{2,3} & \ddots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \delta_{n-1,n} \\ & & & & 0 \end{bmatrix}$$

This space is *n*-dimensional.

The foundational models of trade and growth suggest alternative forms of this matrix. For example, a smooth quality-ladder model (eg. Grossman & Helpman 1989) implies the following form:

$$(9)\ \Delta = \begin{bmatrix} 0 & c & \infty & \cdots & \infty \\ & \ddots & c & \ddots & \vdots \\ & & \ddots & \ddots & \infty \\ & & & \ddots & c \\ & & & & 0 \end{bmatrix}$$

where each product one rung up the ladder is slightly more complex and requires some adaptation or R&D, and leapfrogging isn't possible due to huge distances. Or consider the Hecksher-Ohlin model, where productive opportunities are determined by relative factor endowments, broadly defined. This would be represented as a product space with groupings determined by factor intensities. Firms are limited to products intensive in those factors in which the country has relative factor abundance.

BAILEY: I AM NOT SURE I UNDERSTAND THE HECKSHER-OHLIN MATRIX. THIS IS A MATRIX OF PAIRWISE DISTANCES. HOW DO YOU INTERPRET THE BLOCKS?

$$(10)\ \Delta = \begin{bmatrix} \begin{bmatrix} 0 & l \\ l & 0 \end{bmatrix} & & \infty \\ & \begin{bmatrix} 0 & k \\ k & 0 \end{bmatrix} & \\ \infty & & \begin{bmatrix} 0 & h \\ h & 0 \end{bmatrix} \end{bmatrix}$$

The self-discovery model of Hausmann and Rodrik (2003) would be represented by a substitution matrix with each element off the diagonal as a random variable.

We depart from these assumptions about the product space and their implications for the matrix of pairwise distances, and instead seek to measure it empirically. We then can test whether these distances regulate the process of structural transformation.

3. Data & Methodology

To make progress we need an empirical measure of the pairwise distance between products at a highly disaggregated level that allows us to map the product space. There are many potential measures developed in the literature. For example, product relatedness may be affected by vertical input-output relationships (e.g. Ditezenbacher & Lahr 2001) or by similarity in patent citations (Jaffe 1986, Caballero and Jaffee 1993). Yet these are measures of particular dimensions of similarity between products, which may be dominated by other dimensions. For example, it is not clear that being composed of similar inputs is more important than being sold to the same market, or quoting similar patents is more important than requiring the same types of infrastructure or institutions.

We seek a measure of the distance between products that avoids any priors as to the relevant dimension of similarity, and instead is based on outcomes. Our main idea is that the similarity of requisite specific assets is revealed by the likelihood that countries have revealed comparative advantage in both goods. To develop this measure we use product-level data of exports. This is not only for data availability reasons: exports represent products in which a country has a comparative advantage and must pass a rather strict market test compared to production for the domestic market. For a country to have revealed comparative advantage in an export good it must have the right endowments and capabilities to produce that good and export it successfully. If two goods require the same productive factors, this should show up in a higher probability of a country having comparative advantage in both. We calculate this probability across a large sample of countries.

We must decide which measure of probability to use. Calculating the joint probability that the two goods are exported (i.e $P(A \cap B)$) may appear to be an option, but this measure combines the similarity between two products with the products' overall presence in global trade. That is, if every single country that exports ostrich eggs also exports ostrich meat, these two goods seem extremely similar to one another. Yet if only three countries in the world export these two goods, then the joint probability for any single country exporting the two would be small, instead of large. We therefore need a measure of the distance that isolates the degree of similarity between the two goods from their overall prevalence in world trade.

The conditional probability *P(A|B)* has this characteristic. However, the conditional probability is not a symmetric measure: *P(A|B)* is not equal to *P(B|A)*. Yet standard notions of distance between two goods are symmetric. More importantly, as the number of exporters of any good A falls, the conditional probability of exporting another good given you export A becomes a dummy variable, equal to 1 for every other good exported by that particular country, and 0 otherwise, thus reflecting the peculiarity of the country and not the similarity of the goods. Suppose Australia is the only country in the world that exports ostrich meet. Then all other goods exported by Australia, such as minerals or wine, would appear to be very close to ostrich meat, when in fact they may be quite different.

Hence, for these two reasons we focus on the minimum of the pairs of conditional probabilities going in both directions as an inverse measure of distance: min{*P(A|B)*, *P(B|A)*}. This formulation would imply that the probability of exporting metal ores given that you export ostrich meat is large, but the probability that you export ostrich meat given that you export metal ores is very low, since Chile, Peru and Zambia do not export ostrich meat but do export metals. If the products were really close together in terms of requisite factors, then all countries exporting metal ores would also export ostrich meat, but this is not the case, and our measure captures it. In the robustness section of the Appendix we relax this assumption and allow for asymmetric distance by using the directional conditional probabilities. All our results continue to hold.

We also want a measure that is strict in terms of capturing true similarities and not just marginal exports. In order to impose this strictness on our data we require not only that a country export any amount, but that its exports of the good are substantial. One way to impose this restriction is to require that the country have revealed comparative advantage (RCA) in that good. This means that the share of the country's exports in that product is greater than the country's share of exports in all products[11]. Since every country tends to have a specialized basket of exports, this measure captures all its significant exports but leaves aside the noise[12]. In short, our measure of revealed distance has no priors as to its cause, and goods will only be measured as highly proximate if they indeed strongly tend to be exported together, for whatever reason.

---

[11] We use the Balassa (1965) definition:

$$RCA_{c,i,t} = \frac{xval_{c,i,t} \Big/ \sum_i xval_{c,i,t}}{\sum_c xval_{c,i,t} \Big/ \sum_i \sum_c xval_{c,i,t}}$$

[12] We also repeated our tests using a definition of 'exported' as exports of more than 0.6% and 0.06% of the country's total export basket and also by putting a fixed minimum dollar value of exports. All results continued to hold.

Formally, the inverse measure of distance between goods *i* and *j* in year *t*, which we will call proximity, equals

$$(11)\ \varphi_{i,j,t} = \min\{P(x_{i,t} \mid x_{j,t}), P(x_{j,t} \mid x_{i,t})\}$$

where for any country *c*

$$(12)\ x_{i,c,t} = \begin{cases} 1 & if \quad RCA_{i,c,t} > 1 \\ 0 & otherwise \end{cases}$$

and where the conditional probability is calculated using all countries in year *t*.

Our primary source of export data is the World Trade Flows data from Feenstra et. al. (2005). These data are drawn from the United Nations Commodity Trade Statistics, and available from 1962-2000 at the SITC 4-digit level of desegregation (1006 products). While export data at a higher level of disaggregation can be obtained from the UN COMTRADE database, the advantage of these data is that they are significantly cleaner than the raw data and exist for a longer time period[13].

Exploring Proximity

To get a sense of the data, **w**e can list for each good what other products are close and which tend to be farther away. For example, let us consider the distance of cotton undergarments and CPUs to other products. This is illustrated in Table 1.

Table 1. Illustrating the Product Sapce: Proximity to cotton undergarments and CPUs

Proximity of Cotton Undergarments to:	
Synthetic undergarments	0.78
Overcoats	0.51
Woven fabrics	0.12
Centrifuges	0.02
Proximity of CPUs to:	
Digital central storage units	0.56
Epoxide resins	0.50
Optical glass	0.32
Unmilled rye	0.00

Source: Author's Calculations

We can also see what goods are in a dense part of the product space and which are on the periphery by simply adding the row for that product in the matrix of proximities, and

[13] See Feenstra et. al. 2005 for documentation.

dividing by the maximum possible number of distance-weighted products (J=1006). We define the distance-weighted number of paths leading to product i at time t.

$$(13)\quad paths_{i,t} = \frac{\sum_j \varphi_{i,j,t}}{J}$$

With this definition Table 2 looks at the goods that are in the densest (2.A) and the sparsest (2.B) part of the product space, based on the average of 1998-2000 export data.

BAILEY: COULD YOU REDO THESE TABLES ELIMINATING GOODS WITH TOTAL EXPORTS BELOW A CERTAIN CUTOFF?

Table 2.A The Ten Goods in the Densest Part of the Forest

Code	Product Name	Paths
6996	MISCELLANEOUS ARTICLES OF BASE METAL	0.207
6785	TUBE & PIPE FITTINGS(JOINTS,ELBOWS)OF IRON/STEEL	0.207
6921	RESERVOIRS,TANKS,VATS AND SIMILAR CONTAINERS	0.203
7449	PARTS OF THE MACHINERY OF 744.2-	0.199
6210	MATERIALS OF RUBBER(E.G.,PASTES.PLATES,SHEETS,ETC)	0.199
8935	ART.OF ELECTRIC LIGHTING OF MATERIALS OF DIV.58	0.198
8939	MISCELLANEOUS ART.OF MATERIALS OF DIV.58	0.197
5335	COLOUR.PREPTNS OF A KIND USED IN CERAMIC,ENAMELLI.	0.196
8932	SANITARY OR TOILET ART.OF MATERIALS OF DIV.58	0.195
6632	NATURAL OR ARTIFICIAL ABRASIVE POWDER OR GRAIN	0.194

Table 2.B The Ten Goods in the Sparsest Part of the Forest

Code	Product Name	Paths
0019	LIVE ANIMALS OF A KIND MAINLY USED FOR HUMAN FOOD	0.00
9110	POSTAL PACKAGES NOT CLASSIFIED ACCORDING TO KIND	0.01
6553	KNITTED/CROCHETED FABRICS ELASTIC OR RUBBERIZED	0.01
2655	MANILA HEMP,RAW OR PROCESSED,NOT SPUN;TOW & WASTE	0.01
4245	CASTOR OIL	0.03
2640	JUTE & OTHER TEXTILE BAST FIBRES,NES,RAW/PROCESSED	0.03
2231	COPRA	0.03
6344	WOOD-BASED PANELS,N.E.S.	0.03
2235	CASTOR OIL SEEDS	0.03
6545	FABRICS,WOVEN,OF JUTE OR OF OTHER TEXTILE BAST FIB	0.03

Source: Author's Calculations

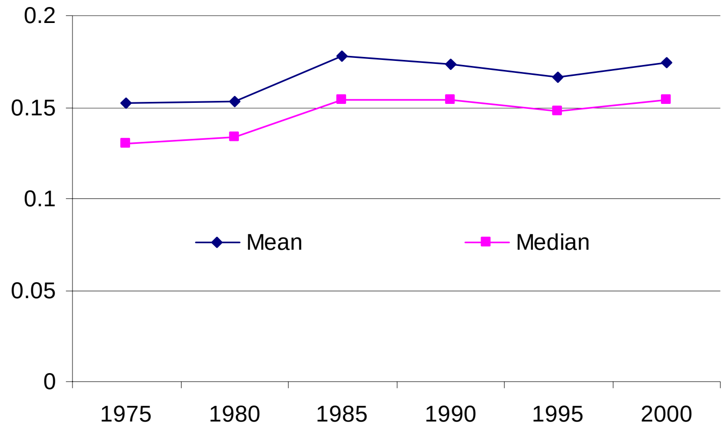
Notice that the densest part of the forest tends to be dominated by manufactured products while the sparsest goods tend to be un-processed agricultural goods such as live animals, castor oil, jute, sisal, cork and mate. Yet broad SITC categories are not a determinant of location in the product space, as some members of almost every SITC category can be found both in the 50 most and 50 least connected products. Additional descriptive statistics for φ, our measure of proximity, can be found in the Appendix.

How dense is the product space around the areas where different countries have specialized? We can look into this question by calculating the country-level average *paths*. Figure % graphs this variable against GDP per capita and shows that in general, rich (poor) countries tend to be specialized in dense (sparse) parts of the product space.

However, there is significant variation in this relationship. In particular, fast growing countries such as China, India and Indonesia seem to be in a denser part of the forest than would be predicted by their income levels. This is reminiscent of the finding in Hausmann, Hwang and Rodrik (2006) where they calculate the level of income associated with a country's exports by looking at the GDP per capita of countries with comparative advantage in the same export mix. We look at the correlation between our variable and theirs and find that ……. BAILEY: PLEASE INCLUDE HERE THE GRAPHS PATHS VS. GDP PER CAPITA AND PATHS VS. EXPY. COULD YOU RUN A REGRESSION BETWEEN TH TWO? I WOULD HOPE THAT EXPY KILLS GDPPC. THIS WOULD SUGGEST THAT PATHS IS WHAT DRIVES THE RESULTS IN HHR.

How stable are these distances over time? Or put another way, has the product space become more connected or disconnected over time? The change in the product space over time is an interesting subject in its own right that we defer to subsequent work. However, to show that the first-order question is of how countries move in the space rather than how the space moves around countries, we present the mean and median distance measures every five years from 1975-2000. Except for a structural break between 1980 and 1985 (corresponding to a methodological change in the world trade flows data) BAILEY: THIS IS WORRISOME. IF THERE IS A METHODOLOGICAL BREAK, DOES THIS BREAK AFFECT OUR RESULTS? HOW DO YOU KNOW IT WAS THE BREAK AND NOT CHANGES IN THE WORLD THAT EXPLAIN THE JUMP BEFORE 1985? MOREOVER, A CORRELATION OF 0.6 APPEARS QUITE LOW AND DOES NOT ALLOW YOU TO SAY THAT THEREFORE CHANGES ARE DOMINATED BY MOVEMENTS OF COUNTRIES AND NOT OF THE PRODUCT SPACE. CAN YOU SHOW INSTEAD THE SAME STUFF USING ANNUAL DATA FOR 1985-2000? CAN YOU LOOK NOT ONLY AT THE MEAN BVUT AT THE MEDIAN? there is no clear trend. The correlation coefficient for the proximity of product pairs declines over time, but is always above .6 within the two periods with the same coding methodologies. Conversely, we will see in the final section that country position in the product space changes significantly over time. Furthermore, as a robustness test we repeat all of our estimations using only those distances calculated in 1975, and all results continue to hold with the same statistical and economic significance.

Figure X: How is the Product Space Changing Over Time?
Mean and Median Proximity 1975-2000



Restricted to pairs of goods exported in all periods by at least one country. Source: Author's Calculations

The Sources of Proximity

Our measure of proximity is outcomes-based, and therefore agnostic as to the sources of product similarity. We can take a first cut at analyzing the sources of similarity by considering how closely our measure conforms to notions of similarity from the literature discussed in Section 2. Or put another way, to what degree the product space matrix can be simplified according to the assumptions of most models of trade.

First, we take our proximity matrix, re-order the products into blocks based on their broad factor intensities by Leamer's classification system (1984), and examine the average proximity within versus between categories. The results are shown in Table 3. A factor proportions model of the world would suggest a high proximity within groups, and a low proximity (high distance) between groups. While it is the case that the highest proximity measures occur within product categories, the degree of relatedness is similar to that observed with other categories, indicating that only a fraction of the pattern of relatedness is accounted for by broad factor intensities as captured by Leamer (1984).

Table 3: **Average φ Within and Between Leamer Commodity Clusters, 1998-2000**

	Petroleum	Raw Materials	Forest Products	Tropical Agriculture	Animal Products	Cereals, etc.	Labor Intensive	Capital Intensive	Machinery	Chemical
Petroleum	**0.21**	0.10	0.11	0.11	0.09	0.09	0.09	0.11	0.07	0.10
Raw Materials		**0.10**	0.09	0.08	0.08	0.07	0.07	0.09	0.07	0.09
Forest Products			**0.17**	0.10	0.10	0.09	0.11	0.13	0.10	0.11
Tropical Agriculture				**0.14**	0.10	0.09	0.10	0.11	0.07	0.09
Animal Products					**0.11**	0.09	0.09	0.09	0.07	0.09
Cereals, etc.						**0.09**	0.08	0.09	0.07	0.09
Labor Intensive							**0.13**	0.13	0.10	0.10
Capital Intensive								**0.16**	0.11	0.12
Machinery									**0.14**	0.12
Chemical										**0.15**

We can look instead at degrees of technological sophistication as a determinant of product relatedness. We repeat the above exercise using Lall's technological classification of exports (2000). The results are shown in Table 4. If this was the dominant story we would expect a very high proximity within blocks, and a very low average proximity between.

Table 4: **Average φ Within and Between Lall Technology Categories, 1998-2000**

	PP	RB1	RB2	LT1	LT2	MT1	MT2	MT3	HT1	HT2
PP	**0.09**	0.09	0.09	0.09	0.09	0.07	0.08	0.06	0.06	0.07
RB1		**0.12**	0.11	0.10	0.12	0.11	0.11	0.10	0.08	0.09
RB2			**0.12**	0.09	0.12	0.11	0.12	0.11	0.09	0.11
LT1				**0.16**	0.13	0.09	0.10	0.09	0.08	0.07
LT2					**0.17**	0.14	0.14	0.14	0.12	0.11
MT1						**0.17**	0.13	0.15	0.11	0.11
MT2							**0.14**	0.13	0.11	0.12
MT3								**0.16**	0.12	0.14
HT1									**0.17**	0.13
HT2										**0.15**

PP	Primary Products
RB1	Resource-Based Products (agriculture)
RB2	Resource-Based Products (other)
LT1	Low-Technology (textile, garment, footwear)
LT2	Low-Technology (other)
MT1	Medium-Technology (automotive products)
MT2	Medium-Technology (chemicals and basic metals)
MT3	Medium-Technology (engineering products)
HT1	High-Technology (electronics)
HT2	High-Technology (other)

Note: All pairs of the same product were first dropped so that block size would not affect the average proximity. Source: Author's Calculations

As in Table 3, Table 4 has slightly higher coefficients along the diagonal, indicating that there is something to the classification, but the values in many of the off-diagonal cells are not that different. This suggests that the specificity that underlies the product space is more complex than that captured with broad classifications. This suggests that our measure of proximity incorporates the structural relationships identified by both Leamer and Lall, but combines these elements of similarity with other relevant dimensions and captures more of the heterogeneity present in the product space.

In the following section, we will use these proximity measures to develop a measure of a country's location in the product space and show it has significant consequences for the pattern of structural transformation.

4. Density and the Speed of Structural Transformation

Our outcomes-based measure of distance between products is a structural relationship common to all countries. We now need to develop a measure of how close is a country to each of the products it currently does not export with comparative advantage. We will then test whether this measure, which we call *density*, is a significant and robust determinant of future changes in comparative advantage.

If our measure of proximity is indeed capturing the degree of factor substitutability between products, then the probability of exporting a particular good in the future depends on its proximity to the current export basket. To test this, we need to combine pairwise proximity measures with each country's pattern of production in a single country-product measure. *Density* is the sum of all paths leading to the product in which the country is present, scaled by the total number of paths leading to the product[14]. That is, it is the percentage of distance-weighted paths leading to a good that are currently occupied by the country. Formally,

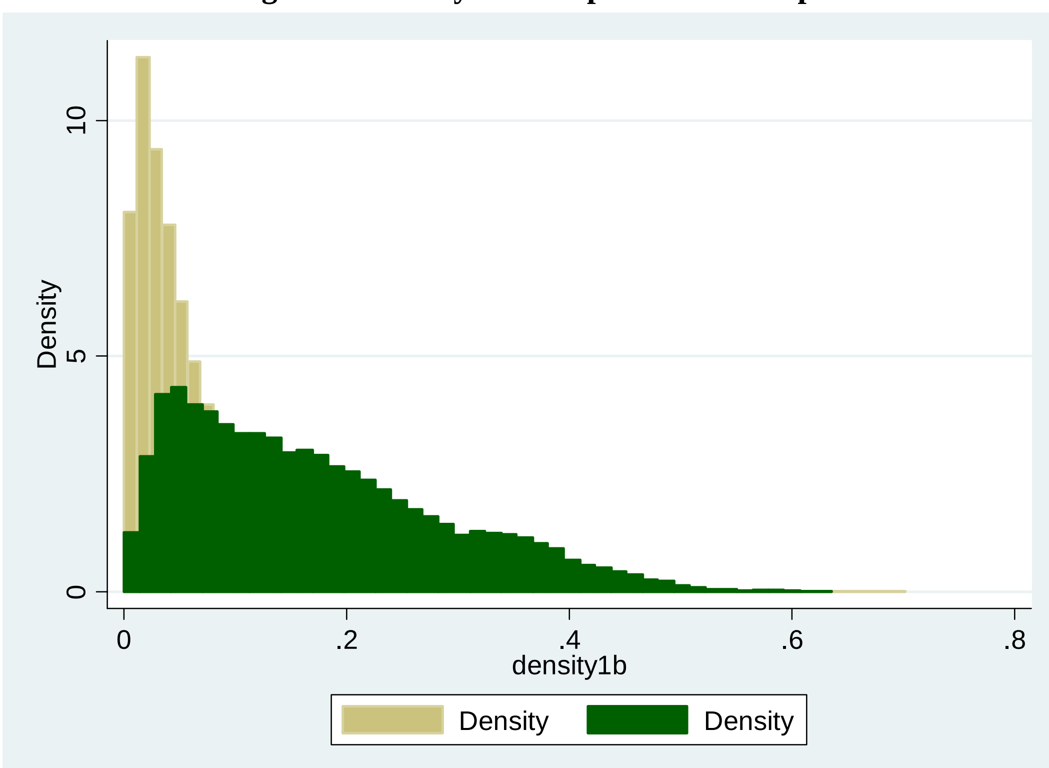
[14] Note that in the robustness section we test unscaled density (i.e. without the denominator). All results continue to hold. In any case, the total number of paths leading to a good is a time-varying product specific variable and would be captured by our product-year dummies.

$$\text{(14)}\ density_{i,c,t} = \left(\frac{\sum_k \varphi_{i,k,t} x_{c,k,t}}{\sum_k \varphi_{i,k,t}} \right)$$

INSERT MORE DESCRIPTIVE STATS ON DENSITY HERE
-densities across countries and products, density over time, etc.

According to our model, the cost of producing a new product rises with distance, and therefore firms will only find it profitable to move to nearby goods, meaning there should be a positive relationship between density and the probability of exporting a new good. We test this proposition by plotting a histogram of density for all products not produced in period t, splitting the sample according to whether the product was also not produced in the next period (brown) versus cases where the country successfully exported the new product with comparative advantage (green). Figure 1 clearly shows that density is higher for products that were subsequently 'discovered', suggesting that structural transformation does indeed depend on distance as we have measured it. IT WOULD BE INTERESTING TO REPORT THE T-TEST OF THE SIMILARITY BETWEEN THE TWO DISTRIBUTIONS

Figure 1: Density for Jumps vs. Non-Jumps



Using all goods without comparative advantage in period t, the density around goods also without comparative advantage in t+1 is shown in brown, and those with comparative advantage in t+1 in green. Source: Author's Calculations

We test the importance of proximity more formally by estimating the following equation on five-year panels from 1975-2000 in the Feenstra export data:

$$(15)\ x_{i,c,t+1} = \alpha + \gamma x_{i,c,t} + \beta density_{i,c,t} + \pi X + \varepsilon$$

where X is a vector of country+year and time+year dummy variables, which control for any time-varying country or product characteristic. Density is normalized by subtracting the mean and dividing by the standard deviation to make the estimated coefficient in units of standard deviations. The results, estimated using OLS[15] with standard errors clustered at the country level, are shown below in Table 4.

Table 4

	(1)	(2)
	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.657	0.655
	(66.27)**	(67.44)**
$density_{i,c,t}$	0.062	0.056
	(7.03)**	(6.36)**
$RCA\_lall_{la,c,t}$		0.004
		(7.46)**
$RCA\_leamer_{le,c,t}$		0.008
		(6.19)**
Observations	398362	389092
R-squared	0.56	0.56

t statistics in parentheses
* significant at 5%; ** significant at 1%

After controlling for *all* time-varying country and product characteristics, we see that density is a highly significant determinant of structural transformation. A 1-standard deviation increase in density leads to a 6 percentage point increase in the probability of moving to the new product. Given that the unconditional probability of exporting a good that was not exported last period is only 1.27%, this is an almost five-fold increase in the probability of moving to a new product.

As suggested by our model, these results show that structural transformation favors nearby products, as we have measured them. We can also determine how much our outcomes-based measure of proximity adds to the broad classifications based on technology and factor intensities. That is, we can test the importance of relaxing the assumption that the matrix of pairwise proximities is in block form by factor intensity or technological sophistication. We do this by controlling for each country's revealed comparative advantage in the respective Leamer (1984) commodity cluster associated with that good, and in the Lall (2000) technology class associated with that good. If our

[15] As illustrated by Greene (2004), the maximum likelihood estimator with fixed effects sizes suffers from an incidental parameters problem which biases the results when groups are small, as is our case, so we can not use a probit estimation. Furthermore, non-linear estimations with over 300,000 observations and over 3000 control variables are too computationally-intensive.

measure of proximity is only capturing patterns of specialization in products based on broad factor endowments or technologies, these controls should swamp our results.

Column 2 shows that this is not the case. While it is true that a country's revealed comparative advantage in broadly-defined commodity clusters and technology classes does affects structural transformation in a statistically significant way, the remaining variation in density continues to be highly significant in determining future structural transformation. Moreover, a 1-standard deviation increase in a country's RCA in the Leamer commodity group or Lall technology class leads to an increase in the probability of gaining RCA in that good of 0.9 and 1.1 percentage points, respectively, while the corresponding figure for the remaining variation captured by our density variable is 5.6 percentage points.

The previous equation does not distinguish between the factors determining the probability of moving into a new export good from the determinants of abandoning goods that are currently being exported. To make this distinction, we estimate the following equation:

$$x_{i,c,t+1} = \alpha + \gamma x_{i,c,t} + \beta_1 \left(x_{i,c,t}\right) density_{i,c,t} + \beta_2 \left(1 - x_{i,c,t}\right) density_{i,c,t} + \pi X + \varepsilon$$

Here, β_1 represents the impact of density in preventing abandonment while β_2 is the coefficient relating density to the probability of moving to a new product. We see that although the estimated coefficient of β_2 is slightly smaller, it remains highly significant, both statistically and economically.

Table 5

	$x_{i,c,t+1}$
$x_{i,c,t}$	0.641
	(51.34)**
(1-$x_{i,c,t}$)*density$_{i,c,t}$	0.046
	(6.75)**
($x_{i,c,t}$)*density$_{i,c,t}$	0.068
	(6.47)**
Observations	398362
R-squared	0.56

t statistics in parentheses
* significant at 5%; ** significant at 1%

A question may arise as to whether our results are dependent on our discontinuous treatment of the left-hand side variable, which is based on a dummy variable that reflects whether the country has comparative advantage in the good or not. As a robustness test, we use the value of the RCA index directly[16], and estimate the following equation:

[16] Because RCA is a ratio, at the highly disaggregated product level there can be extreme outliers in cases where a country with an extremely small share of world exports is the only exporter of a particular good. In the Feenstra dataset, the 99th percentile of the RCA index at the product level is at 24, whereas the largest four values are all from 7000 to 21000. Therefore, we drop the largest 1% of observations to ensure that a few outliers do not affect the results.

$$(17)\ RCA_{i,c,t+1} = \alpha + \gamma RCA_{i,c,t} + \beta density_{i,c,t} + \pi X + \varepsilon$$

The results are provided below in Table 6, and show that our results are not dependent on our definition of x. A 1-standard deviation increase in density is associated with an increase in the RCA index for a product of 0.668. BAILEY: YOU HAD WRITTEN 1 INSTEAD OF 0.668. WAS THAT A MISTAKE?

We relax the assumption of a linear relationship between density and the RCA index by incorporating a quadratic term in density. The results quite robustly show a positive but concave relationship between density and revealed comparative advantage in a product. This concave relationship remains positive even three standard deviations above the mean value of density.

Table 6

	$RCA_{i,c,t+1}$	$RCA_{i,c,t+1}$
$RCA_{i,c,t}$	0.668	0.667
	(46.88)**	(46.30)**
$density_{i,c,t}$	0.366	0.691
	(5.52)**	(6.61)**
$density_{i,c,t}^2$		-0.105
		(4.29)**
Observations	308076	308076
R-squared	0.35	0.35

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

The concave impact of density on structural transformation can be further explored by estimating the following equation, a piece-wise regression:

$$(17)\ \begin{aligned} x_{i,c,t+1} &= \alpha + \gamma x_{i,c,t} + \beta_{10}Q_1 + \beta_{11}Q_1 density_{i,c,t} + \beta_{20}Q_2 + \beta_{21}Q_2 density_{i,c,t} + \\ &\beta_{30}Q_3 + \beta_{31}Q_3 density_{i,c,t} + \beta_{40}Q_1 + \beta_{41}Q_4 density_{i,c,t} + \pi X + \varepsilon \end{aligned}$$

Where

$Q_1 = 1$ if $density_{i,c,t} < 0.05$, 0 otherwise

$Q_2 = 1$ if $0.05 \leq density_{i,c,t} < 0.1$, 0 otherwise

$Q_3 = 1$ if $0.1 \leq density_{i,c,t} < 0.2$, 0 otherwise

$Q_4 = 1$ if $0.2 \leq density_{i,c,t}$, 0 otherwise

Note that here density is not standardized: it is the original measure from 0 to 1. The results are shown in Table 7.

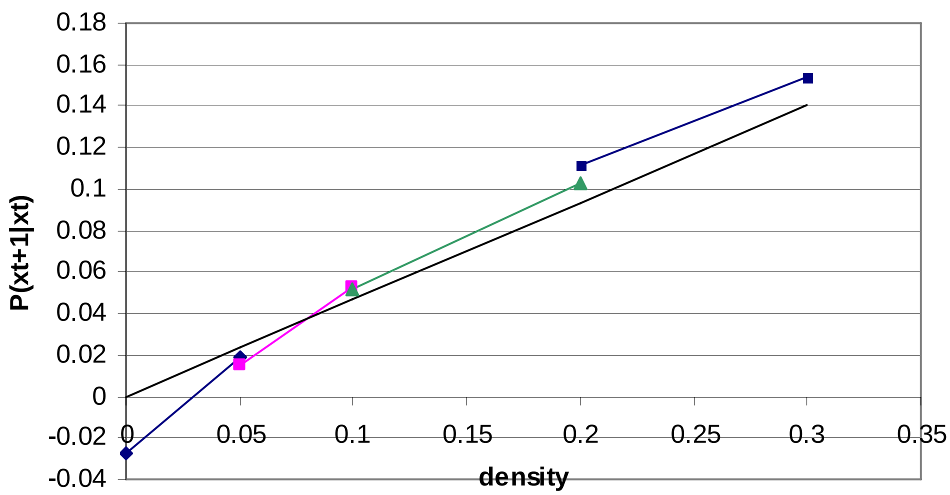
Table 7

	$x_{i,c,t+1}$
$x_{i,c,t}$	0.656
	(66.82)**
Q1	-0.027
	(2.49)*
Q2	-0.023
	(2.02)*
Q3	0.000
	(.)
Q4	0.027
	(1.18)
Q1* $density_{i,c,t}$	0.910
	(6.24)**
Q2* $density_{i,c,t}$	0.758
	(8.22)**
Q3* $density_{i,c,t}$	0.513
	(6.30)**
Q4* $density_{i,c,t}$	0.422
	(4.84)**
Observations	398362
R-squared	0.56

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

Again, these results show a robust concave impact of density, in this case on the probability of achieving RCA in a product next period. However, the degree of concavity is minor. We can plot this estimated relationship and compare it to the linear estimate, shown below in Figure 2. Note that at higher values of density, the estimated probability of the concave curve is above the linear relationship because of rising values of the estimated intercept.

Figure 2



Source: Author's calculations.

Questions may arise as to the impact of reclassifications on our results. If a product gets reclassified into several items, this may cause an apparent move to new products that is only caused by this reclassification. In addition, changes in the distance matrix could potentially be correlated with some other changes, causing some spurious relationship between density and changes in exports. To check for these potential problems, we recalculate density for all years using only the 1975 matrix of pairwise proximities. This implies dropping all product classifications created post-1975. Moreover, it eliminates any potential interactions between changes in density over time caused by changes in the comparative advantage of countries. The results are unchanged.

In addition, we repeat the above estimations using unscaled density (that is, not dividing by the total number of paths leading to a product (see equation 14). The results are presented in the appendix, which show that all of our findings are maintained with equivalent levels of significance. We also relax the requirement that distance be symmetric and redo our results using asymmetric distance (i.e. taking the directional conditional probability rather than the minimum of the pair). We also check whether changing the criteria for the dummy variable x from RCA>1 to a proportion of exports or a threshold dollar value of exports. These results are shown in the appendix. In all cases, the findings were unaffected.

In sum, our measure of distance is a significant determinant of structural transformation: when countries change their pattern of specialization they move preferentially towards nearby products. The pattern of change shows much more structure than that captured by broad measures of factor intensities or levels of technological sophistication. It also shows more structure than what would be expected from any national time-varying factor such as the general contractual environment, rule of law or changes in general human capital. The results are consistent with an interpretation based on the idea that factors of production are very numerous and highly-specific, with varying degrees of substitutability between products.

5. Concluding remarks

Much of the recent theory of trade and growth assumes a homogeneous and continuous product space. This implies that it is always possible to find products through which to move up the ladder of comparative advantage. This paper argues that this assumption is inconsistent with the facts. Depending on where a country has developed its comparative advantage, its opportunities for structural transformation will be affected by the structure of the product space in its neighborhood.

In some sense, this allows us to reinterpret the intuitions of the fathers of development economics. Their belief that industrialization created externalities that if harnessed could lead to accelerated growth, can be interpreted not as being related to forward and backward linkages (Hirschman, 1957) or complementarities in investment requiring a ‘big push’ (Rosenstein-Rodan, 1943) but in terms of the flexibility with which the accumulated capabilities could be redeployed from product to product. The speed of structural transformation is higher in countries which happen to start in products that are in a denser part of the product space.

Our measure of proximity is purely descriptive and outcomes-based. We have shown that it captures the broad relationships of both factor endowments and technological sophistication, yet these dimensions of product similarity are only a part of the picture. The remaining heterogeneity in our measure of distance remains highly significant in predicting patterns of structural transformation, even after controlling for time-varying country and product characteristics.

The work in this paper can be extended in several dimensions. First, it is important to study whether the heterogeneity in the product space matters for aggregate growth. Do countries that are specialized in a denser part of the product space exhibit faster long-run growth? Second, it should be possible to integrate the analysis of transitions across products and quality improvements within products, measured by the changes in the export unit values at the product level. What is the impact of distance to the frontier in a given product to the probability of jumping to a new product? What is the impact of distance on the initial level of quality achievable in a new product?[17]

Third, it would be interesting to study the evolution of the proximity matrix over time in greater detail. How has globalization affected the degree of factor substitutability? Fourth, it should also be of interest to analyze the determinants of structural transformation. What factors are associated with the ability of countries to move to more distant products? Controlling for distance, what factors are associated with more frequent jumps? Fifth, it should be possible to enquire about the potential role for economic policy. The density of products around current areas of comparative advantage represents an externality that captures potential inter-industry spillovers. Have the jumps to new distant products been followed by jumps to nearby products? Is there a case for policies that could move a country from a sparse part of the product space to a denser part and then leave subsequent progress to happen naturally towards nearby products? What factors explain improbable transitions? Were the transitions observed in East Asia the consequence of their position in the product space or were they related to a strategic move towards a denser part of the product space? Are improbable transitions more likely when foreign direct investment is involved? Does the presence of large local conglomerates, as in Korea, help internalize some of the externalities highlighted by this paper? Finally, our study of the proximity matrix could be enhanced by using the tools of network analysis[18].

---

[17] Hwang (2006) finds empirically that when a country adopts a new product it does so at lower qualities (measured as the difference in unit export prices relative to the OECD). It would be interesting to explore how this is affected by the quality reached in the neighboring goods and their distance to the new good.

[18] This work is currently underway with Albert-Lazlo Barabasi and Cesar Hidalgo

References

Aghion, P. & P. Howitt. 1992. “A model of growth through creative destruction” *Econometrica* 60(2): 323-351.

Arrow, K. 1962. "The economic implications of learning by doing" *Review of Economic Studies* 29(3): 155 - 173.

Balassa, B. 1986 “Comparative advantage in manufactured goods: a reappraisal.” *The Review of Economics and Statistics* 68(2): 315-19.

Bardhan, P. 1970. Economic growth, development, and foreign trade. Wiley-Interscience, New York.

Caballero R. & A. Jaffe. 1993. “How high are the giant’s shoulders: an empirical assessment of knowledge spillovers and creative destruction in a model of economic growth” NBER macroeconomics annual 1993, O. Blanchard & S. Fischer (eds.), Cambridge, MA: p. 15-74.

Cabral, L. 2000. “Stretching firm and brand reputation” *RAND journal of economics* 31(4): 658-73.

Diamond, D. 1989. “Reputation acquisition in debt markets” *Journal of Political Economy* 97: 828-62.

Dietzenbacher, E. & M. Lahr. 2001. Input-output analysis: frontiers and extensions. Palgrave, NY.

Feenstra, R. R. Lipsey, H. Deng, A. Ma and H. Mo. 2005. “World Trade Flows: 1962-2000” NBER working paper 11040. National Bureau of Economic Research, Cambridge MA.

Green, W. 2004. “Fixed effects and bias due to the incidental parameters problem in the tobit model” *Econometric Reviews* 23(2): 125-147.

Grossman, G. & E. Helpman. 1989. “Product development and international trade” *The Journal of Political Economy* 97(6): 1261-1283.

Grossman, G. & E. Helpman. 1991. “Quality ladders in the theory of growth” *Review of Economic Studies* 58(1): 43-61.

Hausmann, R. J. Hwang and D. Rodrik. 2005. “What you export matters” NBER Working paper 11905. National Bureau of Economic Research, Cambridge MA.

Hausmann, R. and D. Rodrik. 2003. “Economic development as self-discovery.” *Journal of Development Economics*. 72: 603-633.

Hirschman, A. 1958. The Strategy of Economic Development. New Haven, Conn.: Yale University press.

Jaffe, A. 1986. “Technological opportunity and spillovers of R&D: evidence from firm’s patents, profits, and market value” *American Economic Review* 76(5): 984-1001.

Jaffe, A., M. Trajtenberg & R. Henderson. 1993. “Geographic localization of knowledge spillovers as evidenced by patent citations.” *Quarterly Journal of Economics* 108(3): 577-98.

Jones, R. 1971. “A three-factor model in theory, trade, and history” in Trade, Balance of Payments, and Growth: Papers in International Economics in Honor of Charles P. Kindleberger. J. Bhagwati et al., eds. Amsterdam: North-Holland.

Jovanovic, B. & Y. Nyarko. 1996. “Learning by doing and the choice of technology” *Econometrica* 64(6): 1299-1310.

Lall, S. 2000. “The technological structure and performance of developing country manufactured exports, 1985-1998” Queen Elizabeth House Working Paper #44, University of Oxford.

Lazear, E. 2003. “Firm-specific human capital: a skill-weights approach”. NBER working paper 9679. Cambridge, MA.

Leamer, Edward E. 1984. *Sources of Comparative Advantage: Theory and Evidence*. Cambridge MA: The MIT Press.

Leamer Edward E. 1987 “Paths of Development in the Three Factor, n-Good General Equilibrium Model, *The Journal of Political Economy* 95(5): 961-999.

Matsuyama, K. 1991. “Increasing returns, industrialization, and indeterminacy of equilibrium” *Quarterly Journal of Economics* 106(2): 617-650.

Michaely, Michael. 1984. Trade, Income Levels and Dependence, North-Holland, Amsterdam.

National Bureau of Economic Research (NBER). 1999. “Learning by doing in Markets, Firms, and Countries.” N. Lamoreaux, D. Raff & P. Temin (Eds.). National Bureau of Economic Research Conference Report, University of Chicago Press.

Porter, M. 1998. On Competition. Harvard Business School Press, Cambridge MA.

Scherer, F. 1982. “Inter-industry technology flows and productivity growth.” *The Review of Economics and Statistics* 64(4): 627-634.

Schott, Peter K. (2004) “Across-Product versus Within-Product Specialization in International Trade, *The Quarterly Journal of Economics*, (May), pp. 647-678.

Segerstrom, Peter (1991) “Innovation, Imitation and Economic Growth” *Journal of Political Economy*, Vol. 99, No. 4 (August, 1991), pp. 807-827.

Van Pottelsberghe de la Potterie, B. 1997. “Issues in assessing the effect of interindustry R&D spillovers” *Economic Systems Research* 9(4): 331-356.

Young, Alwyn (1991) “Learning by Doing and the Dynamic Effects of International Trade”, *The Quarterly Journal of Economics*, Vol. 106, No.2 (May 1991), pp. 369-405.

Appendix

Methodological Notes

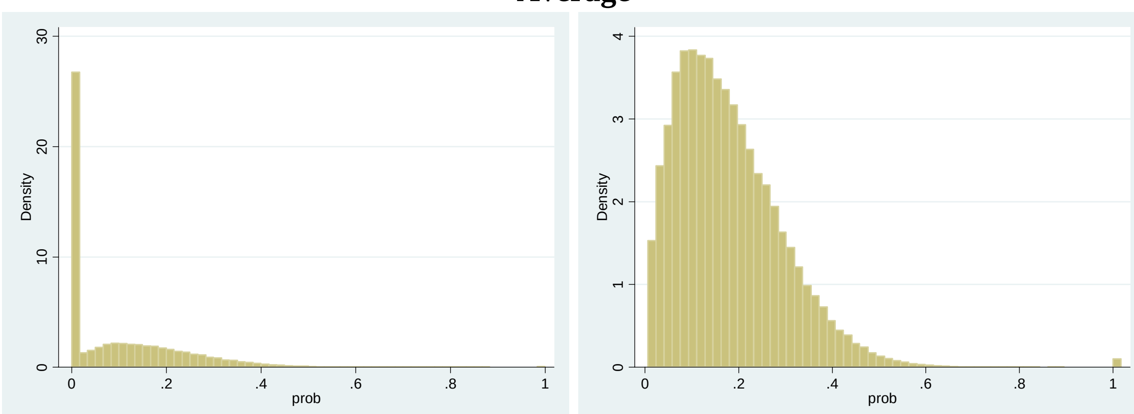
We drop all the artificial 'A' & 'X' product categories from the Feenstra dataset, leaving 1006 products. We drop any countries that reported more than 5% of their total exports in these artificial product categories. We exclude from all regressions countries with a population under 2 million.

Descriptive Statistics for φ
(1998-2000 Average)

Variable	Obs	Mean	Std. Dev.	Min	Max
•	1012036	.1007126	.1240665	0	1

There is a strong mode at 0: most goods are not linked. Excluding the 0s, we see somewhat of a lognormal distribution.

Histogram of proximity (left) and proximity excluding 0 values (right): 1998-2000 Average



Descriptive Statistics for φ
(1985)

Variable	Obs	Mean	Std. Dev.	Min	Max
•	1012036	.129338	.1410314	0	1

Robustness Checks

-X defined by % of exports and by $ value

Results Using φ From 1975

	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$RCA_{i,c,t+1}$	$RCA_{i,c,t+1}$
$x_{i,c,t}$	0.693	0.690		
	(78.15)**	(80.18)**		
$density_{i,c,t}$	0.053	0.046	0.273	0.672
	(6.32)**	(5.63)**	(4.82)**	(5.86)**
$RCA\_leamer_{le,c,t}$		0.003		
		(4.90)**		
$RCA\_lall_{la,c,t}$		0.008		
		(5.89)**		
$RCA_{i,c,t}$			0.698	0.696
			(52.02)**	(51.56)**
$density_{i,c,t}^2$				-0.149
				(5.38)**
Observations	245588	238417	174940	174940
R-squared	0.57	0.57	0.56	0.56

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

Results Using φ From 1975 cont.

	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.686	0.692
	(50.15)**	(77.68)**
$(1-x_{i,c,t})*density_{i,c,t}$	0.049	
	(7.88)**	
$(x_{i,c,t})*density_{i,c,t}$	0.057	
	(4.61)**	
Q1		-0.025
		(2.22)*
Q2		-0.023
		(1.89)
Q3		0.000
		(.)
Q4		0.065
		(2.44)*
$Q1*\ density_{i,c,t}$		0.812
		(5.31)**
$Q2*\ density_{i,c,t}$		0.705
		(6.47)**
$Q3*\ density_{i,c,t}$		0.496
		(6.25)**
$Q4*\ density_{i,c,t}$		0.284
		(3.01)**
Observations	245588	245588
R-squared	0.57	0.57

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

Results with Un-scaled Density

	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$RCA_{i,c,t+1}$	$RCA_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.662	0.661			0.652
	(76.74)**	(77.76)**			(64.28)**
$density_{i,c,t}$	0.045	0.040	0.273	0.489	
	(9.22)**	(8.24)**	(5.57)**	(5.02)**	
$RCA\_leamer_{le,c,t}$		0.004			
		(7.07)**			
$RCA\_lall_{la,c,t}$		0.007			
		(6.10)**			
$RCA_{i,c,t}$			0.670	0.668	
			(46.03)**	(44.82)**	
$density_{i,c,t}^{2}$				-0.050	
				(3.51)**	
$(1-x_{i,c,t})*density_{i,c,t}$					0.037
					(8.75)**
$(x_{i,c,t})*density_{i,c,t}$					0.049
					(7.97)**
Observations	398362	389092	308076	308076	398362
R-squared	0.56	0.56	0.35	0.35	0.56

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

Results with Asymmetric Distance

	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$RCA_{i,c,t+1}$	$RCA_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.663	0.661			0.651
	(66.80)**	(68.18)**			(57.73)**
$density_{i,c,t}$	0.054	0.049	0.343	0.738	
	(5.85)**	(5.37)**	(5.05)**	(7.10)**	
$RCA\_leamer_{le,c,t}$		0.004			
		(7.54)**			
$RCA\_lall_{la,c,t}$		0.008			
		(6.28)**			
$RCA_{i,c,t}$			0.669	0.667	
			(47.30)**	(46.52)**	
$density_{i,c,t}^{2}$				-0.107	
				(5.31)**	
$(1-x_{i,c,t})*density_{i,c,t}$					0.033
					(4.57)**
$(x_{i,c,t})*density_{i,c,t}$					0.056
					(5.54)**
Observations	398362	389092	308076	308076	398362
R-squared	0.55	0.56	0.35	0.35	0.56

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%