The Structure of the Product Space and the Evolution of Comparative Advantage

Ricardo Hausmann and Bailey Klinger

CID Working Paper No. 146
April 2007

Working Papers

Center for International Development at Harvard University

The Structure of the Product Space and the Evolution of Comparative Advantage

Ricardo Hausmann and Bailey Klinger

Abstract

This paper establishes a robust stylized fact: changes in the revealed comparative advantage of nations are governed by the pattern of relatedness of products at the global level. As countries change their export mix, there is a strong tendency to move towards related goods rather than to goods that are farther away. The pattern of relatedness of products is only very partially explained by similarity in broad factor or technological intensities, suggesting that the relevant determinants are much more product-specific. Moreover, the pattern of relatedness of products exhibits very strong heterogeneity: there are parts of this 'product space' that are dense while others are sparse. This implies that countries that are specialized in a dense part of the product space have an easier time at changing their revealed comparative advantage than countries that are specialized in more disconnected products.

Keywords: structural transformation, discovery, technological change

JEL codes: F19, O14, O33, O40

This paper is a shortened and revised version of "Structural Transformation and Patterns of Comparative Advantage in the Product Space," CID Working Paper No. 128, http://www.cid.harvard.edu/cidwp/128.htm.

The Structure of the Product Space and the Evolution of Comparative Advantage

Ricardo Hausmann
Bailey Klinger

Center for International Development and
Kennedy School of Government
Harvard University

February 2007

Ricardo Hausmann
Rubenstein-414
Center for International Development
Kennedy School of Government
Mailbox 34
79 JFK Street
Cambridge, MA 02138-5801
Phone: 617-496-3740
Fax: 617-496-8753
Email: ricardo_hausmann@harvard.edu

Bailey Klinger
Rubenstein-409
Center for International Development
Kennedy School of Government
79 JFK Street
Cambridge, MA 02138
Phone: 617-495-1717
Fax: 617-496-8753
Email: bailey_klinger@ksgphd.harvard.edu

We are especially indebted to Roberto Rigobon and Ernesto Stein for extensive comments on an earlier draft. Valuable comments were also 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, Andres Rodriguez-Clare, Dani Rodrik, Charles Sabel, Federico Sturzenegger, & David Weil. All errors & omissions are our own.

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 frozen vegetables impact the future evolution of a country differently than if it had started with bananas?[1]

This paper establishes a robust stylized fact: changes in the revealed comparative advantage of nations are governed by the pattern of relatedness of products at the global level. Countries tend to move from current products to 'nearby' goods, defined in a way we will make more precise below. This pattern of relatedness of products is only very partially explained by similarity in broad factor intensities or technological sophistication as in Leamer (1984) or Lall (2000), suggesting that the relevant determinants are much more product-specific. Moreover, the pattern of relatedness of products exhibits very strong heterogeneity: there are parts of the product space that are dense while others are sparse. This implies that the structure of this product space governs the evolution of comparative advantage.

How does this finding fit in predominant theory? The foundational models of trade theory suggest that the initial pattern of specialization has little to no effect on its future evolution, as it is merely a reflection of deeper underlying characteristics of the country. 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

[1] Hausmann, Hwang and Rodrik (2005) 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.

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 a means of exporting the relative factor abundance[2]. The Ricardian model argues that technological differences across countries determine comparative advantage, and therefore changes in the product mix will 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 as to 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

[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] Matsuyama (1991) assumes exogenous differences in the productivity patterns across sectors, whereby some offer increasing returns while others do not.

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 products space[4]. Moreover we establish that the pattern of relatedness across products is only partly determined by broad factor intensities, at least as expressed in the 10-category Leamer (1984) classification. Relatedness is also only 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

---

[4] This finding contradicts the typical simplifying 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.

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 some 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, even when controlling for all time-varying country and product characteristics. 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].

---

[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 regimes.
[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. While they may be important, contracting institutions are but one dimension of relatedness, and the production function is assumed to be Dixit-Stiglitz so that the product space is continuous and does not affect outcomes[8].

Porter (1998) highlights the emergence of clusters: critical masses of competitive success in particular fields in a particular geographic location. He states that although these clusters rarely conform to standard industrial classification systems, they naturally emerge due to benefits of co-location such as access to specific institutions, public goods, inputs and information sharing. Here, we are exploring the same type of subtle linkages between products, but we bring greater rigor to the exercise by identifying these linkages directly rather than inferring them from the physical co-location of firms in one area.

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

[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.

(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 that we observe 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]. In addition, we do not explore the determinants of observed proximity, beyond similarities in factor and technological requirements.

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

[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.

old worker produces. That is, training is done through learning-by-watching. In the second period, 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

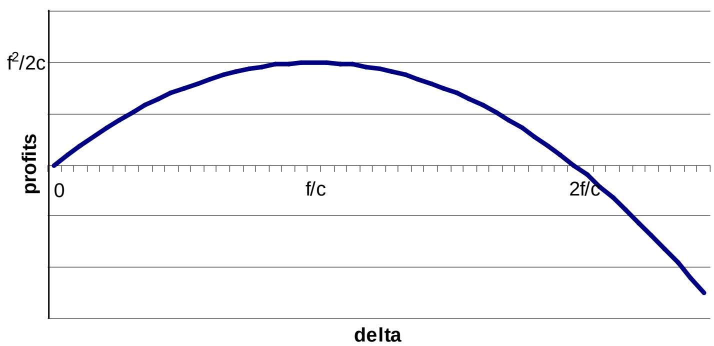
$$(4)\ \delta^{*}_{i,j} = f/c$$

The profits from jumping δ^* are

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

whereas the profits from remaining in good *i* are 0. At distances beyond δ^*, profits begin to decline, and reach zero at $2f/c$. The distance profile of profits are shown in Figure 1.

Figure 1
Distance Profile of Profits



The trained worker will jump δ^* if there is a continuous spectrum of products that could be produced. However, if we move from classical to 'quantum' economics and assume a discontinuous product space, then the trained worker will chose among existing goods that at the distance which maximizes (3). More importantly, if there exists no product nearer than $2f/c$, then the trained worker will remain in the current product, as the higher costs of adapting product-specific human capital outweigh the private benefits.

Yet if the nearest product is at $2f/c + \varepsilon$, this stagnation may be socially sub-optimal. Even if you consider only one subsequent generation and do not allow for any future jumps, it is socially optimal for the trained worker this period to jump as long as $\varepsilon < 2f/c$.

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:

$$(6)\ \Delta = \begin{bmatrix} 0 & \delta_{1,2} & \delta_{1,3} & \cdots & \delta_{1,n} \\ & \ddots & \delta_{2,3} & \ddots & \vdots \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \delta_{n-1,n} \\ & & & & 0 \end{bmatrix}$$

This space is *n*-dimensional. Given the imperfect substitutability of human capital across products, the pattern of structural transformation in this space 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 this space larger than $2f/c$. 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.

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:

$$(7)\ \Delta = \begin{bmatrix} 0 & c & \infty & \cdots & \infty \\ & \ddots & c & \ddots & \vdots \\ & & \ddots & \ddots & \infty \\ & & & \ddots & c \\ & & & & 0 \end{bmatrix}$$

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

We depart from these assumptions about the product space and instead measure the matrix 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

---

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

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

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.

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

$$(8)\ \varphi_{i,j,t} = \min\{P(x_{i,t} \mid x_{j,t}), P(x_{j,t} \mid x_{i,t})\}$$

where for any country *c*

$$(9)\ x_{i,c,t} = \begin{cases} 1 & if \quad RCA_{i,c,t} > 1 \\ 0 & otherwise \end{cases}$$

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.

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

Table 1. Illustrating the Product Space: 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, 1999 proximities

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 dividing by the maximum possible number of distance-weighted products (*J*=1006). This is a measure of product *i*'s centrality in the product space in time *t*. A product that is more central in the product space will be connected to a greater proportion of the 1006 products, and therefore have a higher value for centrality.

$$(10)\ \ centrality_{i,t} = \frac{\sum_j \varphi_{i,j,t}}{J}$$

With this definition, we show the products that are in the densest and the sparsest part of the product space in 1975 (Table 2) and 2000 (Table 3).

Table 2

The Ten Goods in the Densest Part of the Product Space, 1975

Code	Product Name	Paths
6940	NAILS,SCREWS,NUTS,BOLTS ETC.OF IRON.STEEL,COPPER	0.217
6991	LOCKSMITHS WARES,SAFES,STRONG ROOMS OF BASE METAL	0.215
8124	LIGHTING FIXTURES AND FITTINGS AND PARTS	0.212
6794	CASTINGS OR IRON OR STEEL,IN THE ROUGH STATE	0.209
6911	STRUCTURES & PARTS OF STRUC.:IRON/STEEL;PLATES	0.209
5224	METALLIC OXIDES OF ZINC,CHROMIUM,MANGANESE,IRON,	0.208
6573	COATED/IMPREGNATED TEXTILE FABRICS & PRODUCTS NES.	0.208
6282	TRANSMISSION,CONVEYOR/ELEVATOR BELTS OF RUBBER	0.206
6992	CHAIN AND PARTS THEREOF,OF IRON OR STEEL	0.205
7188	ENGINES & MOTORS,N.E.S.SUCH AS WATER TURBINES ETC.	0.204

The Ten Goods in the Sparsest Part of the Product Space, 1975

Code	Product Name	Paths
3330	PETROL.OILS & CRUDE OILS OBT.FROM BITUMIN.MINERALS	0.02
2231	COPRA	0.03
2472	SAWLOGS AND VENEER LOGS,OF NON CONIFEROUS SPECIES	0.04
6330	CORK MANUFACTURES	0.04
721	COCOA BEANS,WHOLE OR BROKEN,RAW OR ROASTED	0.04
2320	NATURAL RUBBER LATEX; NAT.RUBBER & SIM.NAT.GUMS	0.04
4242	PALM OIL	0.04
2654	SISAL & OTHER FIBRES OF AGAVE FAMILY,RAW OR PROCE.	0.04
6871	TIN AND TIN ALLOYS,UNWROUGHT	0.05
6545	FABRICS,WOVEN,OF JUTE OR OF OTHER TEXTILE BAST FIB	0.05

Table 3

The Ten Goods in the Densest Part of the Product Space, 2000

Code	Product Name	Paths
6785	TUBE & PIPE FITTINGS(JOINTS,ELBOWS)OF IRON/STEEL	0.217
6996	MISCELLANEOUS ARTICLES OF BASE METAL	0.209
6921	RESERVOIRS,TANKS,VATS AND SIMILAR CONTAINERS	0.208
6210	MATERIALS OF RUBBER(E.G.,PASTES.PLATES,SHEETS,ETC)	0.207
7849	OTHER PARTS & ACCESSORIES OF MOTOR VEHICLES	0.206
8935	ART.OF ELECTRIC LIGHTING OF MATERIALS OF DIV.58	0.206
8939	MISCELLANEOUS ART.OF MATERIALS OF DIV.58	0.205
7139	PARTS OF INT.COMB.PISTON ENGINES OF 713.2-/713.8-	0.204
7492	TAPS,COCKS,VALVES ETC.FOR PIPES,TANKS,VATS ETC	0.203
5822	AMINOPLASTS	0.202

The Ten Goods in the Sparsest Part of the Product Space, 2000

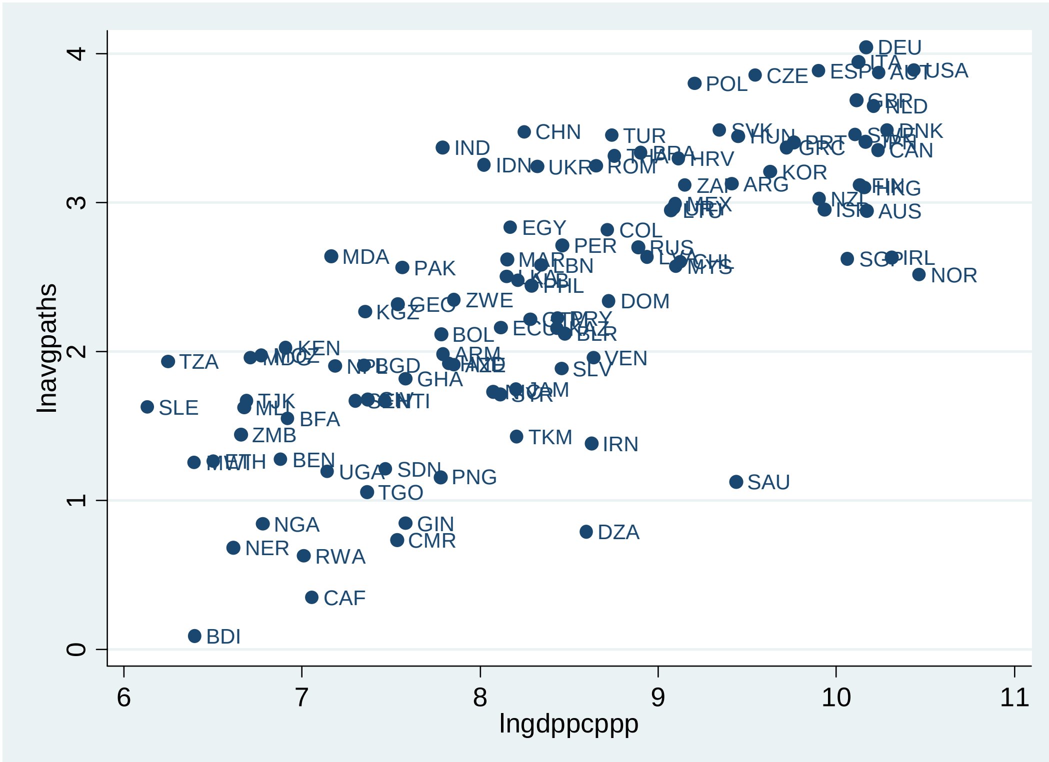
Code	Product Name	Paths
9610	COIN(OTHER THAN GOLD) NOT BEING LEGAL TENDER	0.02
6545	FABRICS,WOVEN,OF JUTE OR OF OTHER TEXTILE BAST FIB	0.02
5723	PYROTECHNIC ARTICLES:(FIREWORK,RAILWAY FOG ETC.)	0.03
4245	CASTOR OIL	0.03
2440	CORK,NATURAL,RAW & WASTE (INCLUD.IN BLOCKS/SHEETS)	0.04
2613	RAW SILK (NOT THROWN)	0.04
0721	COCOA BEANS,WHOLE OR BROKEN,RAW OR ROASTED	0.04
6812	PLATINUM AND OTHER METALS OF THE PLATINUM GROUP	0.04
0573	BANANAS,FRESH OR DRIED	0.04
2876	TIN ORES AND CONCENTRATES	0.05

Source: Author's Calculations. Restricted to goods with at least $100M in world exports for illustrative purposes.

Notice that the densest part of the product space 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 average centrality of all products in which the country has comparative advantage. Figure 2 graphs this variable in 2000 against GDP per capita. This figure 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. Controlling for the level of income, countries such as China, India, Indonesia, Turkey and Poland as specialized in a very dense part of the product space, while countries that specialize in natural resources (particularly oil) have export baskets in disconnected parts of the space.

Figure 2
Average Centrality of Exports vs. GDP per capita (logs), 2000



Source: Author's Calculations (GDP data from World Bank WDI)

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. 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 4, where for each column, the cell is shaded if the average proximity is statistically different at the 1% level from the diagonal. A factor proportions model of the world would suggest a high proximity within groups, and a low proximity (high distance) between groups. So for each column, the diagonal should be the largest number, and statistically significantly different from the off-diagonal entries in the column. We see that this is generally true, although there are some instances where the internal proximity of a Leamer group is smaller than its cross-proximity to another group. This suggests that a broad pattern of relatedness is indeed captured by factor intensities as grouped by Leamer (1984). However, there remains high and variable average proximity between clusters, meaning that a great deal of heterogeneity remains under the surface of a factor proportions view of the product space.

Table 4: **Average φ Within and Between Leamer Commodity Clusters**

	Petroleum	Raw Materials	Forest Products	Tropical Agriculture	Animal Products	Cereals, etc.	Labor Intensive	Capital Intensive	Machinery	Chemical
Petroleum	**0.214**	0.102	0.115	0.118	0.096	0.089	0.095	0.114	0.077	0.109
Raw Materials	0.102	**0.097**	0.095	0.085	0.078	0.075	0.076	0.089	0.071	0.089
Forest Products	0.115	0.095	**0.177**	0.106	0.104	0.087	0.114	0.135	0.105	0.110
Tropical Agriculture	0.118	0.085	0.106	**0.143**	0.109	0.095	0.105	0.111	0.070	0.095
Animal Products	0.096	0.078	0.104	0.109	**0.110**	0.085	0.088	0.095	0.070	0.093
Cereals, etc.	0.089	0.075	0.087	0.095	0.085	**0.085**	0.078	0.090	0.069	0.089
Labor Intensive	0.095	0.076	0.114	0.105	0.088	0.078	**0.135**	0.135	0.105	0.100
Capital Intensive	0.114	0.089	0.135	0.111	0.095	0.090	0.135	**0.158**	0.118	0.121
Machinery	0.077	0.071	0.105	0.070	0.070	0.069	0.105	0.118	**0.143**	0.127
Chemical	0.109	0.089	0.110	0.095	0.093	0.089	0.100	0.121	0.127	**0.152**

Note: All pairs of the same product were first dropped so that block size would not affect the average proximity. Source: Author's Calculations using 2000 distances. Shaded cells indicate that the average proximity in that cell is statistically different from the diagonal in that column, at the 1% level.

We can perform the same analysis considering 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 5.

Table 5: **Average φ Within and Between Lall Technology Categories**

	PP	RB1	RB2	LT1	LT2	MT1	MT2	MT3	HT1	HT2
PP	**0.090**	0.095	0.092	0.087	0.089	0.070	0.083	0.067	0.063	0.073
RB1	0.095	**0.123**	0.110	0.106	0.126	0.115	0.114	0.101	0.086	0.092
RB2	0.092	0.110	**0.123**	0.096	0.125	0.115	0.125	0.113	0.097	0.114
LT1	0.087	0.106	0.096	**0.164**	0.134	0.094	0.100	0.092	0.083	0.073
LT2	0.089	0.126	0.125	0.134	**0.171**	0.151	0.140	0.145	0.119	0.118
MT1	0.070	0.115	0.115	0.094	0.151	**0.178**	0.140	0.158	0.119	0.117
MT2	0.083	0.114	0.125	0.100	0.140	0.140	**0.138**	0.137	0.119	0.130
MT3	0.067	0.101	0.113	0.092	0.145	0.158	0.137	**0.161**	0.130	0.142
HT1	0.063	0.086	0.097	0.083	0.119	0.119	0.119	0.130	**0.173**	0.135
HT2	0.073	0.092	0.114	0.073	0.118	0.117	0.130	0.142	0.135	**0.156**

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 & 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 using 2000 distances. Shaded cells indicate that the average proximity in that cell is statistically different from the diagonal in that column, at the 1% level.

As in Table 4, there are a small number of cases where the cross-group proximities are significantly higher than the within-group average, but for the majority of groupings the within-cluster proximity is highest. This shows that the Lall technological classifications capture an important dimension of relatedness between products. Yet this block form is far from absolute, which suggests that the specificity underlying the product space is more complex than that captured with broad technological classifications. Table 4 and 5 indicate that our measure of proximity incorporates the structural relationships identified by Leamer and Lall, but that it combines these relationships 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 a country exporting a particular good in the future depends on that good's proximity to the current export basket. To test this, we need to combine pairwise proximity measures, which are global characteristics of products, with each country's particular export package. This country-product level measure is '*density*': the density of a country's exports around a particular good. Density is the sum of proximities from good *i* to all products that are currently exported with comparative advantage, divided by the sum of proximities of all products[14]. Put another way, it is the distance-weighted proportion of products connected with good *i* that country *c* exports. If country *c* exports all of the goods connected with product *i*, then density will equal 1. But, if country *c* only has achieved comparative advantage in a small proportion of products which are only weakly connected with good *i*, then density will be much smaller, and will equal 0 if the country exports none of the products connected to good *i*. Formally,

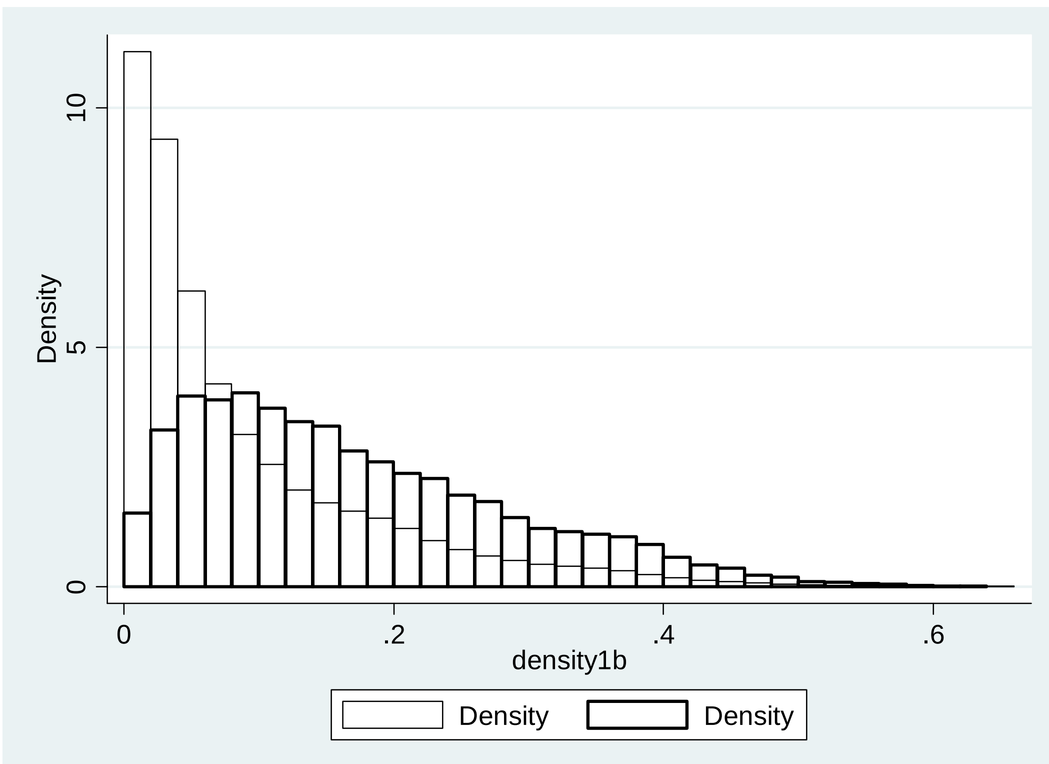
$$(11)\ density_{i,c,t} = \left(\frac{\sum_k \varphi_{i,k,t} x_{c,k,t}}{\sum_k \varphi_{i,k,t}} \right)$$

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

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

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 (thin bars) versus cases where the country successfully exported the new product with comparative advantage (thick bars). Figure 3 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[15].

Figure 3: 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:

$$(12)\ x_{i,c,t+1} = \alpha + \gamma x_{i,c,t} + \beta density_{i,c,t} + \pi X + \varepsilon$$

[15] T t-test for equality of average density for these two distributions is strongly rejected.

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[16] with standard errors clustered at the country level, are shown below in Table 6.

Table 6

	(1)	(2)
	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$[17]	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

[16] 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.

[17] Note that this coefficient is not a measure of persistence because one must consider the country/year and product/year dummies, as well as average density. The unconditional probability of exporting a good with RCA next period given it was exported this period is 75%, while the unconditional probability of exporting a good with RCA next period given it was not exported this period is 1.66%

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 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:

$$(13)\quad x_{i,c,t+1} = \alpha + \gamma x_{i,c,t} + \beta_1 \left(x_{i,c,t}\right) density_{i,c,t} + \beta_2 \left(1 - x_{i,c,t}\right) density_{i,c,t} + \pi X + \varepsilon$$

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 7

	$x_{i,c,t+1}$
$x_{i,c,t}$	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[18], and estimate the following equation:

$$(14)\ RCA_{i,c,t+1} = \alpha + \gamma RCA_{i,c,t} + \beta density_{i,c,t} + \pi X + \varepsilon$$

The results are provided below in Table 8, 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 0.366.

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

[18] 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.

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 8

	$RCA_{i,c,t+1}$	$RCA_{i,c,t+1}$
$RCA_{i,c,t}$	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:

$$(15)\quad \begin{aligned} x_{i,c,t+1} &= \alpha + \gamma x_{i,c,t} + \beta_{10}Q_1 + \beta_{11}Q_1 density_{i,c,t} + \beta_{20}Q_2 + \beta_{21}Q_2 density_{i,c,t} + \\ &\beta_{30}Q_3 + \beta_{31}Q_3 density_{i,c,t} + \beta_{40}Q_1 + \beta_{41}Q_4 density_{i,c,t} + \pi X + \varepsilon \end{aligned}$$

Where

$Q_1 = 1$ if $density_{i,c,t} < 0.05$, 0 otherwise
$Q_2 = 1$ if $0.05 \le density_{i,c,t} < 0.1$, 0 otherwise
$Q_3 = 1$ if $0.1 \le density_{i,c,t} < 0.2$, 0 otherwise
$Q_4 = 1$ if $0.2 \le density_{i,c,t}$, 0 otherwise

Note that here density is not standardized: it is the original measure from 0 to 1. The results are shown in Table 9.

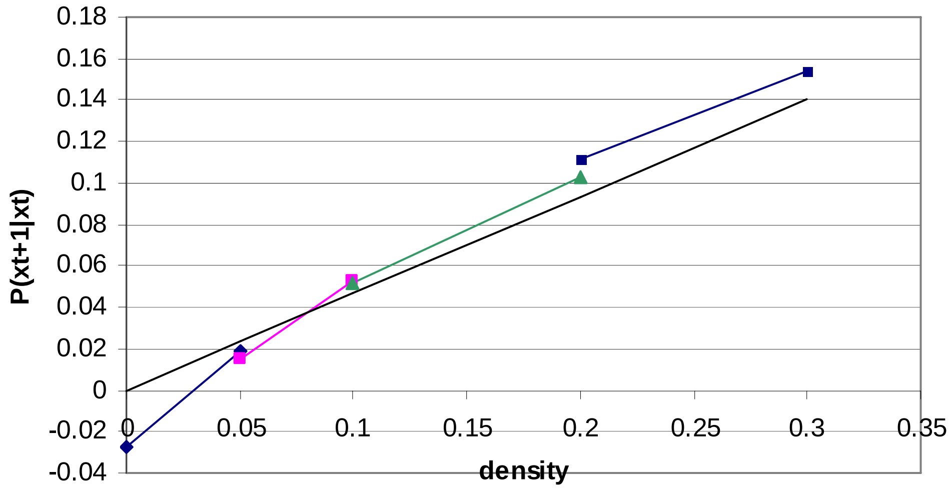
Table 9

	$x_{i,c,t+1}$
$x_{i,c,t}$	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 4. 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 4



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, provided in the Appendix, are unchanged.

In addition, we repeat the above estimations using unscaled density (that is, not dividing by the sum of distance-weighted products (see equation 11). 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 repeat 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 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.

Going back to our opening question, does specialization in frozen vegetables impact the future evolution of a country differently than if it had started with bananas, we now know that the answer is indeed yes. Although both bananas are primary products (Lall) and tropical agriculture (Leamer), specialization in frozen vegetables with a centrality value of .14 leads to far greater opportunities for future structural transformation, as opposed to fresh bananas with a centrality measure of .04. Although our model motivated this impact through specific human capital, any relatively specific non-tradable asset would have the same effect. The infrastructure, property rights, and sanitary regulations specific to frozen vegetables seem to be more easily adapted to other products as compared to fresh bananas, resulting in differing opportunities in the evolution of comparative advantage.

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. Matsuyama's (1991) assertion that certain goods face increasing returns can also be reinterpreted with these empirical results: goods intensive in easily redeployed human capital will offer more self-sustaining growth, and 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?[19]

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

[19] 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.

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 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.

References

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

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

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.

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.

Hwang, J. 2006. "Introduction of New Goods, Convergence and Growth." Department of Economics, Harvard University, mimeo.

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.

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.

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. “Clusters and the new economics of competition.” *Harvard Business Review* 76(6): 77-90.

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.

Appendix

Methodological Notes

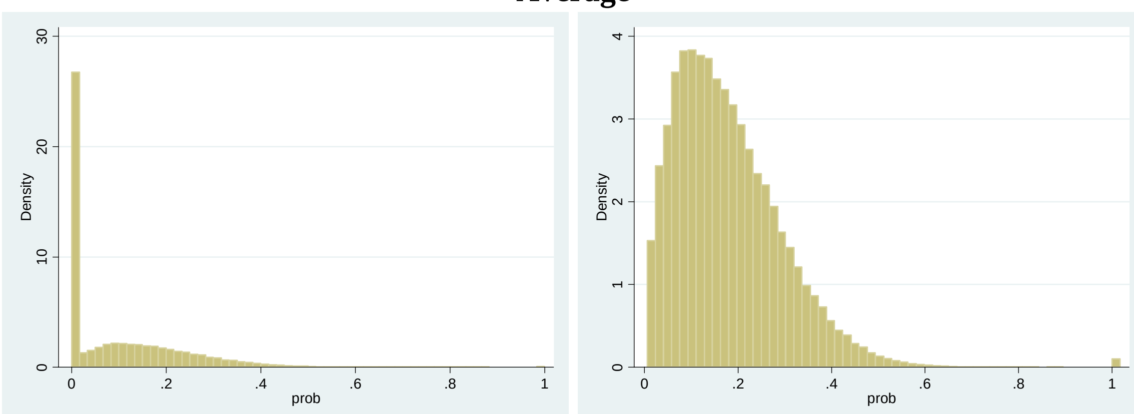
We drop all the artificial 'A' & 'X' product categories from the Feenstra dataset, leaving 1006 products. We drop any countries that reported more than 5% of their total exports in these artificial product categories. We exclude from all regressions countries with a population under 2 million in any year between 1975 and 2000.

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



Robustness Checks

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%

Alternative Definition of 'Exported'

Here, let $x_{i,c,t}$ =1 if export dollar value of good i from country c in year t is greater than \$0, and 0 otherwise.

	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.464	0.459	0.461	0.447
	(78.69)**	(78.40)**	(74.72)**	(70.68)**
density$_{i,c,t}$	0.084	0.076		
	(7.10)**	(6.46)**		
RCA_leamer $_{le,c,t}$		0.006		
		(8.58)**		
RCA_lall $_{la,c,t}$		0.009		
		(6.19)**		
(1-$x_{i,c,t}$)*density$_{i,c,t}$			0.116	
			(8.40)**	
($x_{i,c,t}$)*density$_{i,c,t}$			0.089	
			(7.82)**	
Q1				-0.381
				(15.37)**
Q2				-0.373
				(14.19)**
Q3				-0.262
				(8.32)**
Q4				0.000
				(.)
Q1* density$_{i,c,t}$				2.899
				(11.58)**
Q2* density$_{i,c,t}$				2.534
				(14.79)**
Q3* density$_{i,c,t}$				1.434
				(10.91)**
Q4* density$_{i,c,t}$				0.275
				(3.36)**
Observations	398362	389092	398362	398362
R-squared	0.58	0.59	0.58	0.59

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