Growth Lab

Structural Transformation and Patterns of Comparative Advantage in the Product Space

Ricardo Hausmann and Bailey Klinger

Growth Lab Working Paper Series
No. 5

August
2006

GROWTH LAB
HARVARD KENNEDY SCHOOL
79 JFK STREET
CAMBRIDGE, MA 02138

GROWTHLAB.HKS.HARVARD.EDU

Acknowledgments
Valuable comments have been provided by Philippe Aghion, Hunt Alcott, Laura Alfaro, Albert-Lazlo Barabasi, Elhanan Helpman, Cesar Hidalgo, Asim Khwaja, Robert Lawrence, Daniel Lederman, Roberto Rigobon, Andres Rodriguez-Clare, Dani Rodrik, & Federico Sturzenegger. All errors & omissions are our own.

Statements and views expressed in this report are solely those of the author(s) and do not imply endorsement by Harvard University, Harvard Kennedy School, or the Growth Lab.

© Copyright 2006 Hausmann, Ricardo; Klinger, Bailey; and the President and Fellows of Harvard College

This paper may be referenced as follows: Hausmann, R., and Klinger, B., (2006). “Structural Transformation and Patterns of Comparative Advantage in the Product Space.” Growth Lab Working Paper, Harvard University Kennedy School of Government.

This paper also appears in the Center for International Development Working Paper Series, No. 128.

Structural Transformation and Patterns of Comparative Advantage in the Product Space

Ricardo Hausmann and Bailey Klinger

Abstract

In this paper we examine the product space and its consequences for the process of structural transformation. We argue that the assets and capabilities needed to produce one good are imperfect substitutes for those needed to produce other goods, but the degree of asset specificity varies widely. Given this, the speed of structural transformation will depend on the density of the product space near the area where each country has developed its comparative advantage. While this space is traditionally assumed to be smooth and continuous, we find that in fact it is very heterogeneous, with some areas being very dense and others quite sparse. We develop a measure of revealed proximity between products using comparative advantage in order to map this space, and then show that its heterogeneity is not without consequence. The speed at which countries can transform their productive structure and upgrade their exports depends on having a path to nearby goods that are increasingly of higher value.

Keywords: structural transformation, discovery, technological change

JEL codes: F19, O14, O33, O40

1. Introduction

Rich countries by definition produce more output per worker than poor countries. But they also produce different, presumably more challenging products. Therefore, the process of development involves moving from simple poor-country goods to more complex rich-country goods. This process is often called *structural transformation*. Part of this transformation is related to changing factor endowments as physical, human and institutional capital is accumulated. In standard trade theory the change in the export basket is a passive consequence of changing factor endowments and does not add any policy-relevant aspects to the process[1]. However, there are many reasons why structural transformation may be more complicated than this picture suggests. Several factors may create market failures such as industry-specific learning by doing (Arrow 1962, Bardhan 1970) or industry externalities (Jaffe 1986). There may also be technological spillovers between industries (Jaffe, Trajtemberg and Henderson 1993). Alternatively, the process of finding out which of the many potential products best express a country's changing comparative advantage may create information externalities (Hausmann and Rodrik 2003) as those that identify the goods provide valuable information to other potential entrepreneurs but are not compensated for their efforts.

Hausmann, Hwang and Rodrik (2005) document the positive relationship between a country's GDP per capita and the level of income implicit in the goods that a country exports[2]. However, they also find ample variation in this relationship between countries. More importantly, they find that these variations are not without consequence: controlling for the level of development, a country's level of income implied in its exports is predictive of future growth. In their findings, countries converge to the level of sophistication of their exports. Paraphrasing Pindar, countries seem to become what they export. Hence, beyond standard measures of fundamentals, what a country exports seems to matter for its long-run income performance.

This paper attempts to dig deeper into the determinants of the evolution of the level of sophistication of a country's exports. The story we have in mind is as follows. We argue that producing new things is quite different from producing more of the same. Each product involves highly specific inputs such as knowledge, physical assets, intermediate inputs, labor training requirements, infrastructure needs, property rights, regulatory requirements or other public goods. Established industries somehow have sorted out the many potential failures involved in assuring the presence of all of these inputs, which are

[1] Even, if as pointed by Leamer (1987) countries that start with different factor endowments may move through different products as they switch cones of diversification in the process of capital accumulation, there is always a mix of products with which to express their changing endowments.

[2] To measure the level of income of a country's export package (which they call EXPY) they first define a measure of the level of income implicit in each export product (called PRODY) by calculating the weighted GDP per capita of countries exporting that good, where the weights are related to the countries' revealed comparative advantage in that good. They then calculate EXPY as the weighted average of the PRODYs of the products exported by each country. They find a very strongly positive relationship between EXPY and GDP per capita. To facilitate our exposition, we will interchangeably refer to PRODY and EXPY also as the level of 'sophistication' of the product and the export package, respectively. A related measure was developed earlier by Michaeli (1984).

then available to subsequent entrants in the industry. But firms that venture into new products will find it much harder to secure the requisite inputs. For example, they will not find workers with experience in the product in question or suppliers who regularly furnish that industry. Specific infrastructure needs such as cold storage transportation systems may be non-existent, regulatory services such as product approval and phyto-sanitary permits may be underprovided, research and development capabilities related to that industry may not be there, and so on. In short, changing products is problematic and the difficulties it involves may adversely affect the process of development.

We argue that the assets and capabilities needed to produce one good are imperfect substitutes for those needed to produce another good, but this degree of asset specificity will vary. For example, it sounds plausible to suggest that the human, physical and institutional capabilities needed to produce cotton trousers are closer to those needed to produce cotton shirts than those needed to produce computer monitors. Correspondingly, the probability that a country will develop the capability to be good at producing one good is related to its installed capability in the production of other similar, or nearby goods for which the currently existing productive capabilities can be easily adapted.

Given this varying degree of asset specificity, the speed of structural transformation will depend on the density of the product space near the area where each country has developed its productive capabilities. In theory, the space may be highly homogenous so that nearby products always exist and are at similar distances or it could be very heterogeneous, with highly dense areas in some parts of the product space and highly sparse in others.

One of our contributions in this paper is to propose a new measure of similarity between products that is outcomes-based. In essence, we measure the distance between each pair of products based on the probability that countries in the world export both[3]. This measure goes beyond more standard measures of similarity based on broad factor endowments or *a priori* notions of technological sophistication. Furthermore, we show that the proximity or distance between products is highly heterogeneous. This heterogeneity has particular implications for the speed and patterns of structural transformation, which we motivate with a simple model and examine empirically.

Our metaphor is that products are like trees, and any two trees can be close together or far apart, depending on the similarity of the needed capabilities. Firms are like monkeys, who derive their livelihood from exploiting the tree they occupy. We take the forest – the product space – as given and identical for all countries. As a measure of the productivity of each tree we use Hausmann, Hwang and Rodrik's (2006) measure of the income per capita of the product, which they call *PRODY* and which is based on the income per capita of countries with comparative advantage in that product. The process of structural transformation involves having monkeys jump from the poorer part of the forest to the richer part, but the probability of doing so successfully will depend on the expected productivity of those trees and to how close the monkeys are to unoccupied trees where

[3] For the reasons explained below, we will use the minimum of the conditional probabilities between the two goods going in either direction.

proximity is related to the usefulness of the specific assets the country has for the production of the new good.

In section 2 we motivate the role of proximity to the process of structural transformation with a simple model, which allows us to contrast our characterization of the product space with the core models of trade and structural transformation. In section 3 we develop our empirical measure of distance, and show some of its characteristics. One important feature of the data is that the distance between goods is highly variable. Some parts of the forest are very sparse while others are much denser.

In section 4 we explore how the probability that a country develops comparative advantage in a particular good is related to the density of the country's current production relative to that good, testing the contribution of our measure to the standard assumptions in trade theory. In the parlance of our metaphor, the probability of monkeys jumping to an unoccupied tree depends on the proximity between the current location of the monkeys and the good in question as well as to the upscale character of the good measured by the relationship between the *PRODY* of the good and the *EXPY* of the country. We control for unobserved country and product characteristics as well as other variables such as the level of development of the country. We find strong effects of density, *PRODY* and *EXPY* on the probability of developing comparative advantage in the good.

In section 5 we aggregate the data to the national level to study whether the structure of the product space and the current pattern of specialization affects the aggregate speed of structural transformation. We develop for each country a measure of the value of the unoccupied product space where we take account of the distance between the country's current areas of comparative advantage and each potential product. We call this variable 'open forest'. Countries differ widely in the value this variable takes. Some countries are in a sparsely populated part of the forest, while other countries are in a much more densely populated section of the forest. We then show that this variable strongly predicts the speed of structural transformation measured by the growth in the level of sophistication of exports. Section 6 discusses some of the implications of this analysis.

This paper is related to several strands in the literature. Since we are interested in understanding growth in developing countries, we do not look at the development of new products in the North but at what might be termed imitation in the South. We therefore assume that the product space is fixed and focus only on imitation. However, we stress the differential ability to imitate depending on the distance of a country's current specialization to the rest of the product space. We believe we are the first to emphasize this dimension. Much of 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 distance to a new product, measured by the fixed cost needed to develop a new variety, is independent of the type of specialization the country exhibits[4].

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

There is a literature that distinguishes between improvements in quality within a given product (vertical shifts) and shifts to different products (horizontal shifts). Young (1991) considers bounded learning by doing within a product that also generates spillovers across products. He also considers a continuum of goods and asks whether free trade will lead to a specialization in goods that have exhausted learning by doing vs. goods that still have room to learn. However, he does not look at how the heterogeneity of the product space and the location of individual countries in that space affects their cost of moving to new goods.

At the opposite extreme, Matsuyama (1991) assumes that some goods exhibit endogenous growth and others do not. Therefore, countries may be trapped if static comparative advantage makes a country specialize in the static good. In our framework, we focus not on the increases in productivity within products but at the increases arising from shifting towards new products. The effects of international trade will be related to whether it forces a country to specialize in the sparse ort the dense part of the product space.

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 technologies (horizontal shift). The degree of the similarity of the new technology to the old determines how transferable the accumulated knowledge is, with less similar technologies having a higher productivity loss. An interesting insight is generated by the fact that if a country specializes too much on a good, it may not have incentives to shift to other goods as this will imply a loss of income. Agents can be on paths where they jump early on to new technologies, or become 'stuck' in an old technology forever. However, Jovanovic and Nyarko do not focus on the possibility that the distance to nearby products may vary widely between countries. In this paper we do not focus on improvements within products, but instead concentrate on the varying distances between goods.

Another work that focuses on the difference between vertical vs. horizontal specialization is Schott (2004). He uses US data at the 10-digit level to show that there has been a trend towards reduces specialization by product, reflected in a rising share of products that are imported simultaneously by the US from countries classified as high, middle and low income. However, within each product, there is a positive relationship between the level of development of the country and the import unit value of the goods imported from that country. Our work can be interpreted as showing that this pattern is not random and that countries that enter the types of products exported by rich countries do so because these products are relatively close to the products they already export. In this respect, we will show that there is an enormous heterogeneity across countries in their distance to manufactured goods with Asia and Eastern Europe being much nearer while many countries in Africa and Latin America are quite far away.

There is a significant literature addressing the degree of similarity between products or sectors. For us, the relevant dimension is how the assets, capabilities and opportunities accumulated in the production of a good affect the productivity in the production of another good. In standard trade theory, a country develops comparative advantage in two goods if the endowments required for the production of both of goods are similar. In some sense, we are saying the same thing, but we stress the potentially very large set of endowments that are required, and their varying degree of specificity. This view is compatible with the highly differentiated patterns of specialization across of countries with apparently similar endowments, once one looks at the product space at higher levels of disaggregation, as argued by Hausmann and Rodrik (2003).

Early in the literature of development economics Albert Hirschman (1957) proposed the idea that clusters were related to forward and backward linkages, implying that the demand for inputs or the availability of a product might trigger related industries. This is a very particular reason for activities to group themselves, especially in a world of relatively open trade. Another measure of proximity involves R&D spillovers. Jaffe (1986) measures technological proximity at the firm level using the similarity of the technological classifications of patents, while Caballero and Jaffe (1993) measure the relatedness between products and technologies by tracing patent citations. However, it is unclear why these spillovers would affect the location of firms in the same country. Van Pottelsberghe de la Potterie (1997) compares such approaches to other measures of proximity, such as IO tables. Porter (1990) highlights the role of proximity in various dimensions that form the basis for 'clusters', whose benefits are discussed in Krugman 1991, Krugman & Venables 1996, Porter 1998, and Antonelli 1999. While these approaches focus on particular dimensions of similarity, our outcomes-based measure potentially incorporates many other relevant dimensions, allowing for 'heterogeneous similarity'. This allows different dimensions of proximity, such as a workforce with relevant experience, similarity of physical inputs, geographic location, production technology, distribution channels or legal framework to vary in their importance across goods and over time. In particular, our data suggests patterns of proximity which do not resemble the kinds of clusters that are often posited. We also show that our approach contains more information than categories based on broad measures of factor intensity, such as the Leamer clusters (1984).

2. A Model of Structural Transformation and the Product Space

Every product requires a particular combination of inputs, such as knowledge, physical assets, intermediate inputs, labor training, infrastructure, property rights, regulatory regimes, and so on. For example, production of asparagus may require a certain type of soil, mechanized farming equipment, agribusiness firms that produce at the efficient scale, port infrastructure to ship the product unspoiled, and connections with the small group of multinational purchasers of this product. The exact set is unique to each good, but substitutability is possible. For instance, producing artichokes may require similar infrastructure, similar corporate forms and management techniques, similar marketing relationships, but a different kind of soil and planting cycle.

So the capabilities used to produce one good are an imperfect substitute for those required to produce another. For every pair of goods in the world there is a notion of distance between them: if the goods require highly similar inputs and endowments, then they are 'closer' together, but if they require totally different capabilities, they are 'farther' apart. For example, asparagus may be close to artichokes, but farther from bananas. These distances are a characteristic of productive technology, meaning they do not change from country to country, although it may change over time.

Consider a model of overlapping generations of firms (e.g. Diamond 1989, Cabral 2000) that live for two periods and have output fixed to 1. There are two goods in the world: a standard good (with numeraire price P_1=1) and a new good with price P_2>1. The standard good has been produced in the economy previously, so the particular set of requisite capabilities exist in the country. A firm can produce the standard good and earn 1, or it can invest in the production of the new good that garners a higher price. But because this good has not been produced in the country before, the unique capabilities it requires don't yet exist. Adapting the existing capabilities for the new product creates a fixed cost C. This cost rises with distance between the two goods, δ_{12}, as it is more difficult to move to goods where the capabilities required are totally dissimilar from what currently exists in the country. But once this move is made, the capabilities developed are a public good in the sense that any other firm can now enter without having to pay the fixed cost. Therefore, returns to producing the new good in the first period are:

$$(1)\ P_2 - C(\delta_{12})$$

We assume that $P_2 < C(\delta_{12}) + 1$, meaning that the old firm would not find it profitable to jump to good 2 in period one, and will therefore remain in the standard good. The young firm could either stay in good 1 for two periods (earning 1+1), or it could jump to the higher quality good by paying the fixed cost in the first period, and earning P_2 in both the first and the second period. Therefore, the young firm would jump if:

$$(2)\ P_2 > \frac{C(\delta_{12})}{2} + 1$$

Note that a firm would only move if the good is upscale, i.e. $P_2 > P_1$. If the condition is satisfied, new firms will move to the new product in their first period and remain there in the second period. But if this inequality does not hold, then all firms will remain in the standard good. This would be sub-optimal from the point of view of a social planner with a longer time horizon, as the third generation would also be able to produce the new good without paying the fixed cost of jumping, and therefore the relevant inequality is easier to satisfy:

$$(3)\ P_2 > \frac{C(\delta_{12})}{4} + 1$$

This would be an example of an *intra*-industry spillover, as the firm moving to good 2 does not internalize the benefits it creates for subsequent entrants in that industry. Furthermore, if you extend the model to three goods, there would also be *inter*-industry spillovers, as the young firm adapting the productive capabilities in the country to good 2 does not internalize the returns from their jump allowing future generations to cut the distance to moving into good 3, if $\delta_{23} < \delta_{13.}$

We can extend this model to a continuum of goods, with each firm deciding how far to jump in this continuum in order to maximize profits. Let price rise linearly with distance, and let costs be quadratic in distance, meaning that marginal costs of moving rise linearly with distance.

$$(4)\ P = f\delta$$

$$(5)\ C(\delta) = \frac{c\delta^2}{2}$$

The maximization for the old and new firms, respectively, are:

$$(6)\ \max_{\delta_o} \Pi_o = f\delta_o - \frac{c\delta_o{}^2}{2}$$

$$(7)\ \max_{\delta_{n,1}\delta_{n,2}} \Pi_n = f\delta_{n,1} - \frac{c\delta_{n,1}{}^2}{2} + f\delta_{n,2} - \frac{c(\delta_{n,2} - \delta_{n,1})^2}{2}$$

where δ_o is how far the old firm jumps, $\delta_{n,1}$ is how far the new firm jumps the first period, and $\delta_{n,2}$ how far the new firm jumps second period when it is old. The optimal distances for firms to jump in this space are:

$$(8)\ \delta_o^* = \frac{f}{c}; \delta_{n,1}^* = 2\frac{f}{c}; \delta_{n,2}^* = 3\frac{f}{c}$$

which implies that the young firm jumps two steps sized f/c in the first period and one in the second.

Let us assume that the product space is not continuous, so that it is not necessary for goods to exist with the characteristics needed to satisfy these conditions. Then there could be stagnation in this process of structural transformation, with firms opting not to innovate because there are no goods at the right distance that are sufficiently attractive to pay for the adjustment costs. Stagnation would occur if

$$(9)\ 2\frac{f}{c} < (\delta_{n,1} - \delta_{n,0})$$

Note that a social planner with a longer time horizon might have found it profitable to move to the new product.

So, the process of structural transformation depends on distance, the cost of jumping, and the degree to which the price of the new good exceeds the current goods. Furthermore, this could be interrupted by breaks in the product space. The process of structural transformation could also be interrupted if there are local price maxima, so firms would not have the incentive to move out of that location because nearby goods are downscale (that is, nearby goods fetch a lower price than the current good) and the upscale goods are too far away. Note again that the social planner would have factored these considerations into account when choosing the path of structural transformation and might have avoided local maxima altogether.

The product space can be represented by a matrix of the pairwise distances for all n products:

$$(10)\ \Delta = \begin{bmatrix} 0 & \delta_{1,2} & \delta_{1,3} & \Lambda & \delta_{1,n} \\ & \mathrm{O} & \delta_{2,3} & \mathrm{O} & \mathrm{M} \\ & & \mathrm{O} & \mathrm{O} & \mathrm{M} \\ & & & \mathrm{O} & \delta_{n-1,n} \\ & & & & 0 \end{bmatrix}$$

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

$$(11)\ \Delta = \begin{bmatrix} 0 & c & \infty & \Lambda & \infty \\ & \mathrm{O} & c & \mathrm{O} & \mathrm{M} \\ & & \mathrm{O} & \mathrm{O} & \infty \\ & & & \mathrm{O} & c \\ & & & & 0 \end{bmatrix}$$

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. This would be represented as a product space with groupings determined by factor intensities, and firms are limited jumping to products in which the country has relative factor abundance.

$$(12)\ \Delta = \begin{bmatrix} \begin{bmatrix} 0 & l \\ l & 0 \end{bmatrix} & & & \infty \\ & & \begin{bmatrix} 0 & k \\ k & 0 \end{bmatrix} & \\ & \infty & & \begin{bmatrix} 0 & h \\ h & 0 \end{bmatrix} \end{bmatrix}$$

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. Instead of trying to simplify this matrix into a block form based on assumptions of the relative importance of certain factors of production or technological characteristics of products, we allow for each element to vary depending on the unique characteristics of the pair of goods and their relevant dimensions of similarity. We now develop a methodology for estimating this matrix directly, consider how closely it conforms to the assumptions of a factor proportions model, and evaluate its impact on the process of structural transformation controlling for the other determinants motivated by our model.

3. Data & Methodology

The principal methodological challenge in our approach is to find a measure of distance at the product level in order to map the product space. This could potentially be measured by the physical characteristics of the product as captured by most customs classifications. But it is difficult to assume that the developers of such classifications had in mind the kind of relative asset specificity that motivates our research.

More sophisticated measures of distance between goods have been developed in the literature. For example, input-output tables or R&D intensity can be used to measure the linkages between products (e.g. Ditezenbacher & Lahr 2001, Jaffe 1986). Yet these are measures of particular similarities between goods and not necessarily those that would prove dominant in practice. For example, it is not clear that being composed of similar inputs is more important than being sold to the same markets, or that being of the same R&D intensity is more important than requiring the same specific institutions or infrastructure.

We seek a measure of the revealed distance between products that avoids any priors we might have as to the root cause of similarity. Our main idea is that the similarity of capabilities (or the distance between trees) is heterogeneous, but is related to the likelihood that countries have revealed comparative advantage in both goods. To develop this measure we use product-level data of exports, which is appropriate as 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 a good it must have the right endowments and capabilities to produce that good and export it successfully. If two goods need the same capabilities, 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 the different countries.

The conditional probability *P(A|B)* would have this characteristic. However, the conditional probability is not a symmetric measure: *P(A|B)* is not equal to *P(B|A)*. Yet our notion of distance between two goods is 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, like 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, all countries exporting metal ores would also export ostrich meat, but this is not the case, and our measure captures it. In the robustness checks section of the Appendix we take the directional conditional probabilities, allowing for asymmetric distance, and all 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 positive number, but that its exports of this good be 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[5]. Since every country tends to have a very specialized basket of exports, this measure captures all its significant exports but leaves aside the noise[6]. This is a measure of revealed outcomes with no priors, and goods will only be measured as highly proximate if they indeed strongly tend to be exported together, for whatever reason.

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

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

[6] 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. 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

$$(13)\ \varphi_{i,j,t} = \min\left\{P\left(x_{i,t} \mid x_{j,t}\right), P\left(x_{j,t} \mid x_{i,t}\right)\right\}$$

where for any country *c*

$$(14)\ x_{i,c,t} = \begin{cases} 1 & if \quad RCA_{i,c,t} > 1 \\ 0 & otherwise \end{cases}$$

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[7].

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 Forest: 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 forest, and which are on the periphery by simply adding the row for that product in the matrix of proximities. We define the distance-weighted number of products around a tree *i* at time *t*.

$$(15)\ paths_{i,t} = \sum_{j} \varphi_{i,j,t}$$

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

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

Table 2.a The Fifteen Goods in the Densest Part of the Forest

Code	Product Name	Paths
6996	MISCELLANEOUS ARTICLES OF BASE METAL	208.7
6785	TUBE & PIPE FITTINGS(JOINTS,ELBOWS)OF IRON/STEEL	208.2
6921	RESERVOIRS,TANKS,VATS AND SIMILAR CONTAINERS	204.6
7449	PARTS OF THE MACHINERY OF 744.2-	200.5
6210	MATERIALS OF RUBBER(E.G.,PASTES.PLATES,SHEETS,ETC)	199.8
8935	ART.OF ELECTRIC LIGHTING OF MATERIALS OF DIV.58	199.2
8939	MISCELLANEOUS ART.OF MATERIALS OF DIV.58	198.1
5335	COLOUR.PREPTNS OF A KIND USED IN CERAMIC,ENAMELLI.	197.5
8932	SANITARY OR TOILET ART.OF MATERIALS OF DIV.58	196.2
6632	NATURAL OR ARTIFICIAL ABRASIVE POWDER OR GRAIN	195.5
7139	PARTS OF INT.COMB.PISTON ENGINES OF 713.2-/713.8-	195.1
7849	OTHER PARTS & ACCESSORIES OF MOTOR VEHICLES	194.8
6911	STRUCTURES & PARTS OF STRUC.:IRON/STEEL;PLATES	194.4
7919	RAIL&TRAMWAY TRACK FIXTURES&FITTINGS,SIGNALL.EQUI.	192.9
7868	OTHER VEHICLES,NOT MECHANICALLY PROPELLED,PARTS	192.1

Table 2.b The Fifteen Goods in the Least Dense Part of the Forest

Code	Product Name	Paths
0019	LIVE ANIMALS OF A KIND MAINLY USED FOR HUMAN FOOD	3.2
9110	POSTAL PACKAGES NOT CLASSIFIED ACCORDING TO KIND	7.3
6553	KNITTED/CROCHETED FABRICS ELASTIC OR RUBBERIZED	9.6
2655	MANILA HEMP,RAW OR PROCESSED,NOT SPUN;TOW & WASTE	12.6
4245	CASTOR OIL	25.9
2640	JUTE & OTHER TEXTILE BAST FIBRES,NES,RAW/PROCESSED	26.0
2231	COPRA	26.7
6344	WOOD-BASED PANELS,N.E.S.	28.9
2235	CASTOR OIL SEEDS	29.2
6545	FABRICS,WOVEN,OF JUTE OR OF OTHER TEXTILE BAST FIB	31.2
5723	PYROTECHNIC ARTICLES:(FIREWORK,RAILWAY FOG ETC.)	31.5
2440	CORK,NATURAL,RAW & WASTE (INCLUD.IN BLOCKS/SHEETS)	34.1
2654	SISAL & OTHER FIBRES OF AGAVE FAMILY,RAW OR PROCE.	34.5
0721	COCOA BEANS,WHOLE OR BROKEN,RAW OR ROASTED	40.3
0742	MATE	40.7

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. Additional descriptive statistics for φ, our measure of proximity, can be found in the Appendix.

Given these estimates, we can take a first cut at analyzing the proximity matrix, our representation of the product space, by considering how closely it conforms to the assumptions of trade models based on relative factor endowments as discussed in Section

2. We arrange the products into blocks based on their factor intensities, and examine the average proximity within blocks and between blocks. This is shown below for a partitioning according to Leamer's commodity clusters (1984). A factor proportions view of the world would suggest a high proximity within groups, and a low proximity (high distance) between groups.

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	Not Classified
Petroleum	**0.28**	0.10	0.11	0.11	0.09	0.09	0.09	0.11	0.07	0.10	0.04
Raw Materials		**0.11**	0.09	0.08	0.08	0.07	0.07	0.09	0.07	0.09	0.03
Forest Products			**0.19**	0.10	0.10	0.09	0.11	0.13	0.10	0.11	0.04
Tropical Agriculture				**0.15**	0.10	0.09	0.10	0.11	0.07	0.09	0.04
Animal Products					**0.12**	0.09	0.09	0.09	0.07	0.09	0.03
Cereals, etc.						**0.10**	0.08	0.09	0.07	0.09	0.03
Labor Intensive							**0.14**	0.13	0.10	0.10	0.03
Capital Intensive								**0.16**	0.11	0.12	0.03
Machinery									**0.14**	0.12	0.03
Chemical										**0.15**	0.03
Not Classified											**0.25**

Source: Author's Calculations

We see from this analysis that the assumptions of factor proportions models are not unreasonable. For each commodity cluster, the average proximity is higher within commodity clusters than between them. This is particularly true for petroleum, forest products, and capital intensive products. Therefore the factor-proportions characterization of the product space can be seen in the matrix as we have measured it. Yet we also see that such a simplification hides great heterogeneity. For example, there is nearly equivalent proximity within cereals as between cereals and chemicals. Labor intensive products and capital intensive products have high distance from raw materials, but are much closer to one another. It appears that while not unreasonable, simplifications of the product space mask a great deal of heterogeneity, which is not surprising given how varied production processes and their required capabilities are. We now go on to test our model of structural transformation more rigorously.

4. Proximity and the Speed of Structural Transformation

Armed with an outcomes-based measure of distance, we can test our model of the process of structural transformation. Furthermore, we can test if allowing for a more heterogeneous product space is important by separating the effects of broad factor intensities from individual product distances. Since our measure of proximity is constructed using cross-sectional data at a particular point in time, we can exploit the time variation to answer these questions, controlling for the other determinants of structural transformation presented in section 2 (price) as well as level of development.

Our measure of "price" in the theoretical model of section 2, in which we assumed unit outputs per firms, is really equivalent to Hausmann Hwang & Rodrik's (2005) measure of the income level of the product $PRODY_{i,t}$. This is a measure calculated as the GDP per capita of countries that produce it, weighted by their revealed comparative advantage in that product. As mentioned above, Hausmann Hwang & Rodrik use this product-level

variable to calculate the level of sophistication of a country's export basket, $EXPY_{c,t}$ as the $PRODY_{i,t}$ for each component of the country's export basket weighted by its share. Price in our model is considered relative to the numeraire, which is the price of the 'standard' good. The price of this standard good is captured by *EXPY*. Formally,

$$(16)\ PRODY_{i,t} = \sum_{c} \left[\frac{\left(xval_{c,i,t} \Big/ \sum_{i} xval_{c,i,t} \right)}{\sum_{c} \left(xval_{c,i,t} \Big/ \sum_{i} xval_{c,i,t} \right)} \times GDPpercapita_{c,t} \right]$$

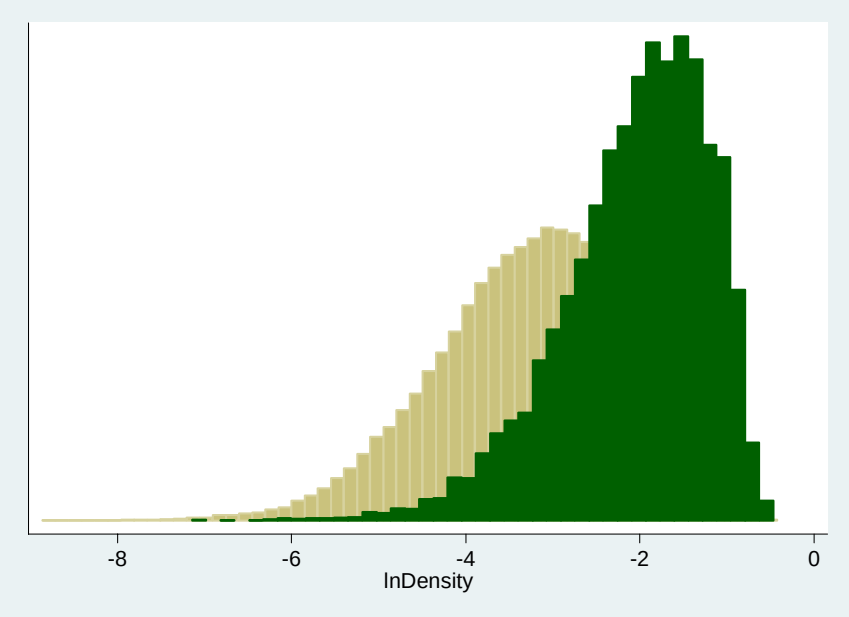
$$(17)\ EXPY_{c,t} = \sum_{i} \left(\frac{xval_{c,i,t}}{\sum_{i} xval_{c,i,t}} \times PRODY_{i,t} \right)$$

If the characteristics of product space are indeed important to the process of structural transformation, then the probability of developing revealed comparative advantage (RCA) in a particular good in the future is affected by the ease with which the current capabilities in the economy can be adapted to the new product. That is, the new product's proximity to the country's current export basket will matter. To test this, we need to use the pairwise proximity measures for each element of the country's entire export basket. We call this measure density. For each product, it measures the degree to which a country's current exports 'surround' the particular product under consideration. It 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. As such, it varies from 0 to 1, with higher values indicating that the country has monkeys in many nearby trees and therefore should be more likely to export that good in the future.

$$(18)\ density_{i,c,t} = \left(\frac{\sum_{k} \varphi_{i,k,t} x_{c,k,t}}{\sum_{k} \varphi_{i,k,t}} \right)$$

According to our model, firms are more likely to move to new products if the distance is low, which would be the case if density is high. We test this proposition by plotting a histogram of density for products in which countries did not have RCA in the current period. We split the distribution according to whether the countries still did not have RCA in the next period (brown) versus those in which it did develop that advantage (green). Figure 1 clearly shows that density is higher for trees that were subsequently 'jumped to', suggesting that structural transformation does indeed depend on distance as we have measured it.

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 model of section 2 more formally in Table 4. Column 1 shows the results of the probit estimation where the dependent variable is whether a country has RCA in the product next period controlling for whether it had it in the current period. We test the determinants of structural transformation motivated in the model: the price of the new product (*PRODY*) controlling the price of the standard goods produced in the economy (*EXPY*), as well as the proximity to the new product (*density*). We control for GDP per capita, as well as country, product, and time effects, and cluster standard errors by country.

As we are controlling for country-fixed effects, it is not surprising that GDP per capita is not highly significant. We expect the likelihood to jump as being positively affected by the difference between the income earning potential of the new good, proxied by *PRODY* and that of the current export basket, captured by *EXPY*. The coefficient on *PRODY* is positive and significant, while that of *EXPY* is negative. Most importantly, we see that density, which captures the ease with which the country's factors and skills can be adapted to the new product, does have a positive and significant effect on the process of structural transformation.

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. To address this issue, we perform the same estimation using OLS, shown in Column 3. The estimated coefficient is smaller in magnitude, but remains highly significant. In the Appendix, we run a series of robustness tests. First, the equivalent regressions are run using 5-year panels from 1975-2000. In addition, as our

model is about the process of occupying new trees rather than abandoning trees, we restrict the sample to only those products not exported in the current period, to account for the possibility of asymmetric effects of density. Restricting the sample may generate other problems of sample selection bias. Nevertheless, the results (using OLS) in the Appendix show the estimated coefficients are statistically significant and larger in magnitude than those reported here.

In addition to testing the relevance of the product space to structural transformation, we can also determine the importance of relaxing the assumption that the matrix is in block form based on factor endowments. This is accomplished by grouping the products into broad commodity clusters from Leamer (1984) and controlling for the country's revealed comparative advantage in that overall cluster. This allows us to test for the possibility that our measure of proximity is only capturing patterns of specialization in products based on factor endowments.

The results in Columns 2 and 4 show that indeed, a country's revealed comparative advantage in the broadly-defined commodity cluster affects the probability of jumping to other products in the same category. However, controlling for this block characteristic of the product space barely lowers the coefficient on density. Our measure of density remains strong and highly significant in both the probit and OLS estimation. Using the most conservative estimate of the effect of density on the probability of exporting a good, a 1 standard deviation increase in density, controlling for broad factor endowments, results in a 1.3 percentage point increase in the probability of exporting a product the following year, all else equal.

Table 4: 1985-2000 Annual Observations

	(1)	(2)	(3)	(4)
	Probit	Probit	OLS	OLS
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.678	0.673	0.801	0.799
	(55.65)**	(57.60)**	(110.82)**	(110.78)**
ln$density_{i,c,t}$	0.043	0.039	0.013	0.011
	(4.66)**	(4.44)**	(4.28)**	(4.05)**
ln$GDPpc_{c,t}$	0.019	0.017	0.005	0.004
	(2.26)*	(2.20)*	(1.08)	(0.81)
ln$EXPY_{c,t}$	-0.049	-0.042	-0.020	-0.017
	(5.55)**	(4.88)**	(4.16)**	(3.57)**
ln$PRODY_{i,t}$	0.009	0.008	0.008	0.008
	(5.34)**	(5.25)**	(7.37)**	(7.23)**
$RCA_{i,c,t}$		0.003		0.003
		(10.40)**		(7.88)**
Constant			0.082	0.105
			(1.37)	(.)
Observations	1172681	1170478	1175839	1173635
R-squared			0.70	0.70

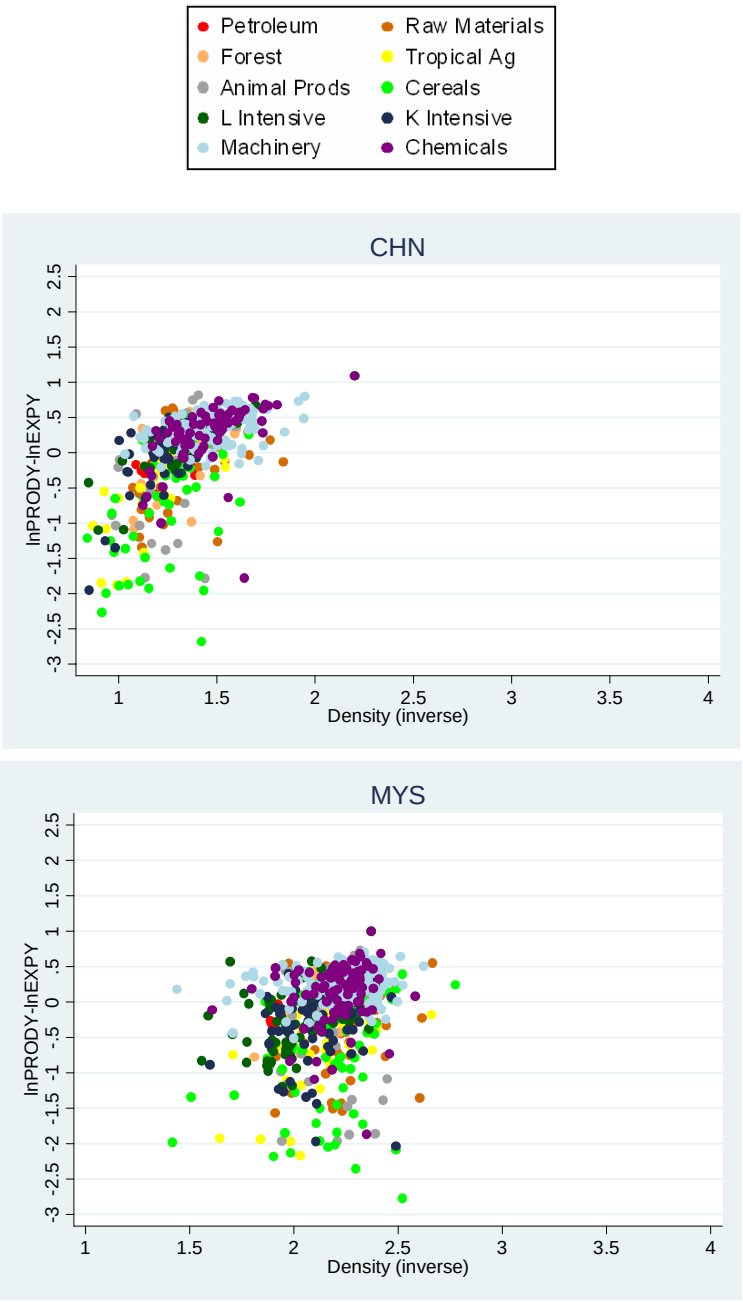
Standard errors are clustered by country. Year, country, and product dummies included in all estimations. Probit coefficients are marginal effects. Absolute value of z statistics in parentheses.
* significant at 5%; ** significant at 1%

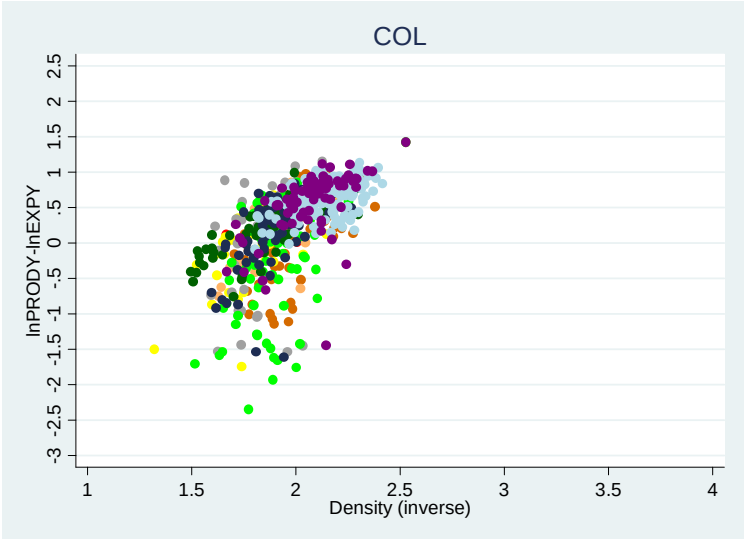
These equations confirm the intuition that the probability that a country develops comparative advantage in a certain good depends on the proximity of the country's current comparative advantage to that good and of its attractiveness in terms of a higher *PRODY*. We can now ask the question of how close and how attractive are the unoccupied trees that a country has. We explore this question by graphing for every country the difference between *PRODY* and *EXPY* against the (inverse of) density for each product. We color the points according to their Leamer group classification. We maintain the same scales in the axes so that the location of the points can be compared across countries. Note that the measure of density is in logs, so that a difference of 1 means a 100 percent difference.

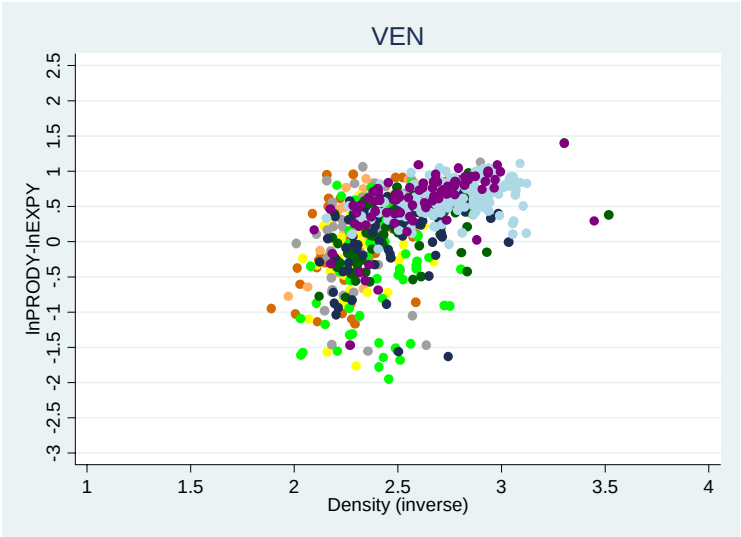
The equations shown above suggest that one should observe rapid structural transformation in countries with many upscale goods that exhibit high density. In Figure 2 we present the graphs for China, Malaysia, Colombia, Venezuela, El Salvador, and Ghana[8]. Notice how China's empty forest is so much closer than that of other countries in the figure. The nearest goods are downscale (i.e. have a lower *PRODY* than that country's *EXPY*, meaning they are less attractive than the trade-weighted average product currently exported), and are represented by labor intensive goods, cereals and tropical agriculture. However, at a density of about 1 there are many upscale products in machinery and capital intensive goods. By contrast, Ghana's nearest goods are much farther away. Between a density of 1.8 and 2.4 the only goods that appear are not upscale and are constituted by cereals and tropical agriculture. The upscale goods start at a density of 2.4 – 140 percent farther away than in China – and are constituted by animal products and labor intensive manufactures. Malaysia has many upscale goods at a density between 1.5 and 2 constituted mainly by electrical machinery and equipment, chemicals and some upscale labor intensive goods. Colombia's upscale goods are at a similar distance as Malaysia's but the nearest ones are animal products, cereals and labor intensive manufactures. Venezuela has a much lower density of nearby products: the upscale products start at distance of about 2 (100 percent lower than China) and are constituted by raw materials, animal products, chemicals and some metal products. El Salvador is somewhere between Colombia and Ghana with its upscale products starting at a density of about 2 and showing a very shallow slope.

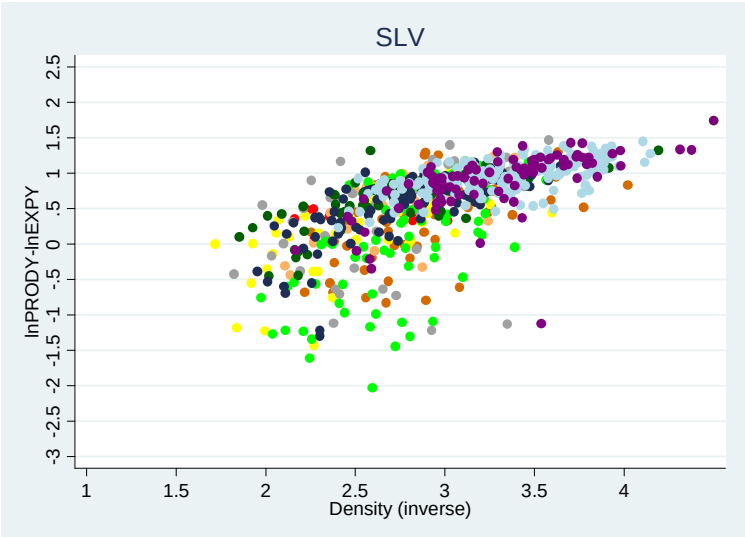
[8] We thank Albert-Laszlo Barabasi for having suggested this graph.

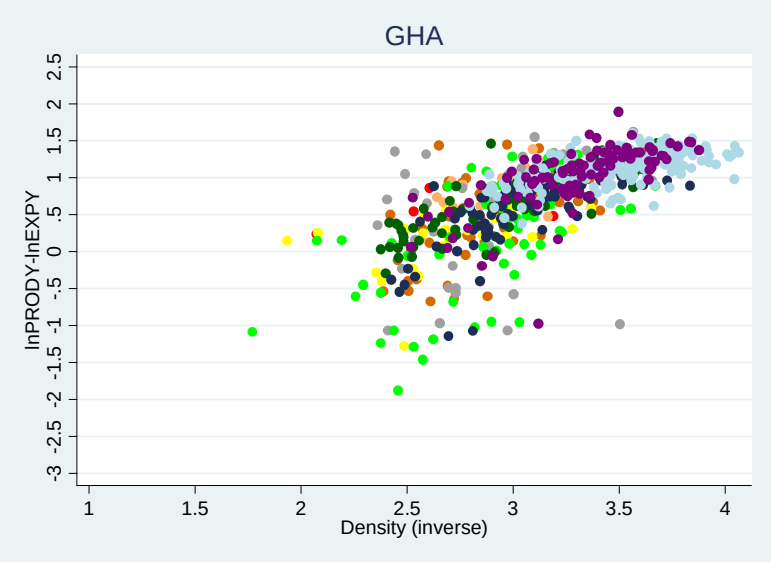
Figure 2: Visual Representation of the 'Open Forest' by Country, 1999
PRODY – EXPY (y-axis, logs) vs. Inverse of Density (x-axis, higher value means the good is further from the current export basket)




COL
lnPRODY-lnEXPY
-3 -2.5 -2 -1.5 -1 -.5 0 .5 1 1.5 2 2.5
1 1.5 2 2.5 3 3.5 4
Density (inverse)


VEN
lnPRODY-lnEXPY
-3 -2.5 -2 -1.5 -1 -.5 0 .5 1 1.5 2 2.5
1 1.5 2 2.5 3 3.5 4
Density (inverse)


SLV
lnPRODY-lnEXPY
-3 -2.5 -2 -1.5 -1 -.5 0 .5 1 1.5 2 2.5
1 1.5 2 2.5 3 3.5 4
Density (inverse)



5. The Product Space & Country Level Export Sophistication

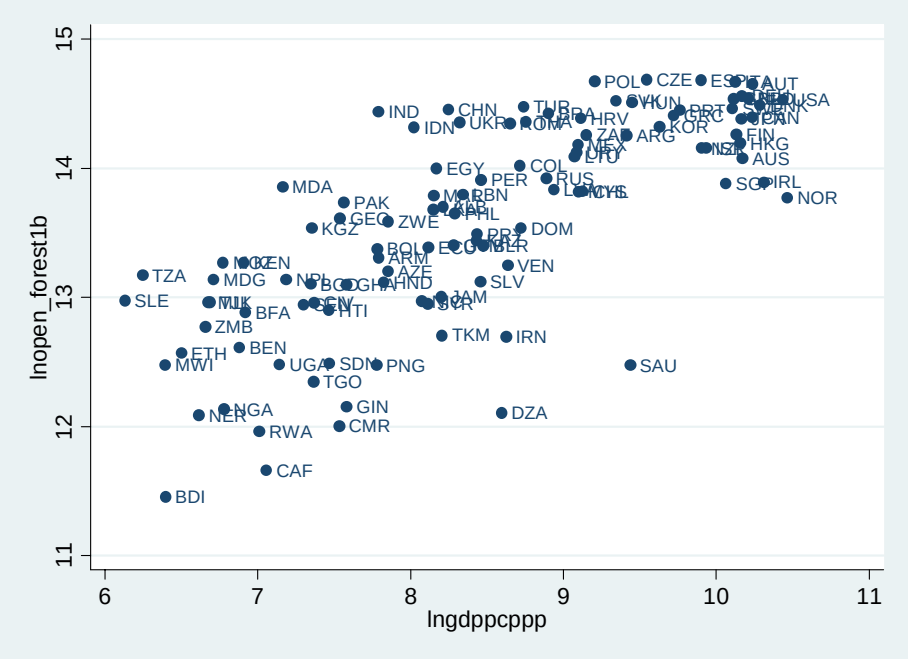
We have seen that the characteristics of the product space have consequences for the process of structural transformation at the product level. We now examine how these relationships translate into differences in structural transformation at the country level by testing if the growth of *EXPY*, i.e. the increase in the level of sophistication or the income level implicit in a country's exports, is affected by the opportunities that the current productive structure provides due to its location in the product space.

We have seen that the opportunities for future structural transformation are in part determined by what products are nearby. Using our estimation of the proximity and *PRODY* in the product space, we can measure the 'option value' of a country's unexploited opportunities. Given the set of products a country is currently producing, we can measure the 'open forest' at its doorstep as the distance-weighted value of all the products it could potentially produce where the value is captured by *PRODY*. Formally:

$$(19)\ open_forest_{c,t} = \sum_i \sum_j \left[\frac{\varphi_{i,j,t}}{\sum_i \varphi_{i,j,t}} \left(1 - x_{c,j,t}\right) x_{c,i,t} PRODY_{j,t} \right]$$

A scatter plot of open forest versus GDP per capita is shown in Figure 3. Countries differ widely in the value of their open forest. The difference between the Czech Republic and El Salvador is over 150 percent. Clearly, some countries are in a sparsely populated part of the product space, while other countries are in a much more densely populated part of the forest with many products nearby.

Figure 3: Open Forest vs. GDP Per capita (logs), 2000

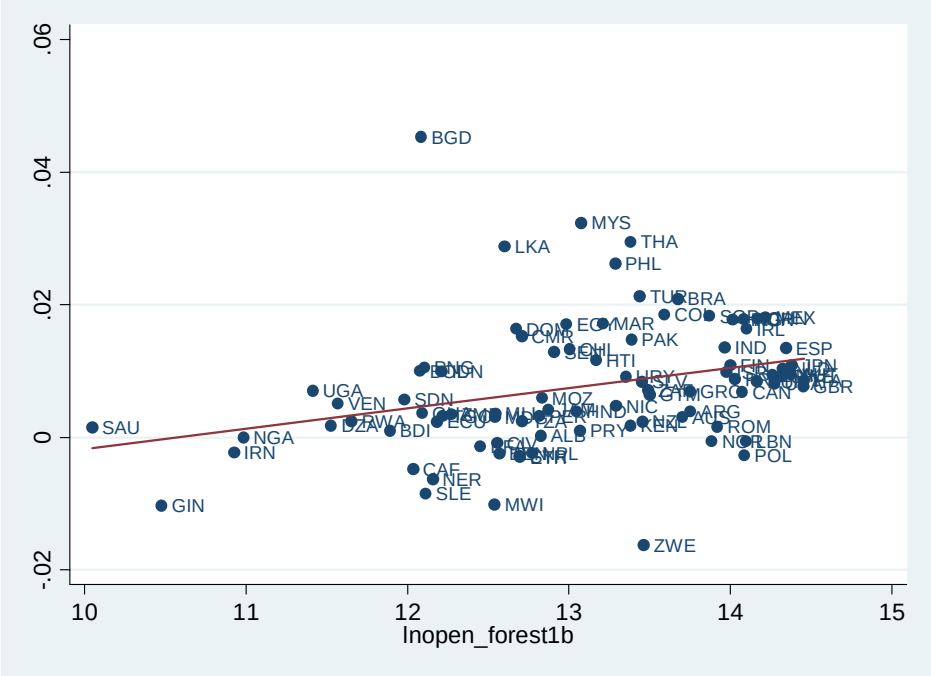


Includes all countries with population over 2 million. Source: Author's Calculations

We can see that wealthier countries tend to be in a denser part of the forest as compared to poorer countries, yet there is significant variation. Oil exporters seem to have an export basket that provides few opportunities for future structural transformation, whereas many eastern European countries, as well as China, India, and Indonesia (which also have large populations), seem to be in a denser part of the forest.

We can look at the question of whether the opportunities actually represented in 'open forest' are actually developed over time. Figure 4 shows open forest in 1975 against growth in *EXPY* between 1975 and 2000. There is a significant positive relationship, with some outliers. For example, Zimbabwe has fared much worse than the opportunities of its export basket suggested, whereas Bangladesh has done much better.

Figure 4: Open Forest in 1975 (log) vs. Growth in EXPY from 1975-2000 (p.a.)



Source: Author's Calculations

We test this relationship formally by regressing future growth in *EXPY* on current open forest, controlling for the initial level of development and level of sophistication of the export basket. Table 5 shows the results. We see that there is convergence in *EXPY* controlling for level of development, as countries with lower levels of export sophistication (column 2: random effects) or previously rapidly growing export sophistication (column 1: fixed effects) experience slower *EXPY* growth in the future. Most interestingly, we see that even when controlling for fixed country effects, changes in open forest lead to subsequent growth in *EXPY*. The proximity of new opportunities as we have measured it is a highly significant determinant of *EXPY* growth, as a 1-standard deviation in open forest is associated with higher *EXPY* growth of 1.6 percentage points per year. The equivalent regression using 5-year panels, as well as robustness checks considering population, can be found in the Appendix.

This measure of the open forest combines two potentially important factors: the distance-weighted *number* of trees, and the *value* of those trees. We can decompose open forest into these two components: the size of the open forest, and its value:

$$(20)\quad open_forest_size_{c,t} = \sum_i \sum_j \left[\frac{\varphi_{i,j,t}}{\sum_i \varphi_{i,j,t}} \left(1 - x_{c,j,t}\right) x_{c,i,t} \right]$$

$$(21)\quad open_forest_value_{c,i,t} = \frac{open_forest_{c,i,t}}{open_forest_size_{c,i,t}}$$

We see that both size and value are statistically significant when controlling for random country effects, but only value is significant when controlling for fixed effects. In terms of economic significance, the value of open forest also dominates. A one standard deviation increase in open forest size results in an increase in annual *EXPY* growth of a half percentage point, using the smallest estimated coefficient, whereas a one standard deviation in open forest value results in a 1.5 percentage point increase.

Table 5: Open_Forest and EXPY Growth, 1985-2000

	(1)	(2)	(3)	(4)
	FE	RE	FE	RE
	EXPY growth	*EXPY* growth	*EXPY* growth	*EXPY* growth
ln$EXPY_{c,t}$	-0.185	-0.059	-0.229	-0.068
	(9.36)**	(5.69)**	(10.86)**	(6.35)**
ln$GDPpc_{c,t}$	0.025	0.010	0.009	0.012
	(1.48)	(2.75)**	(0.53)	(3.22)**
ln$open_forest_{c,t}$	0.027	0.016		
	(3.67)**	(4.14)**		
ln$open_forest_size_{c,t}$			0.006	0.010
			(0.79)	(2.38)*
ln$open_forest_value_{c,t}$			0.329	0.145
			(5.95)**	(3.51)**
Constant	1.085	0.242	-1.111	-0.865
	(5.81)**	(4.99)**	(2.53)*	(2.43)*
Observations	1434	1434	1434	1434
Number of countryid	106	106	106	106
R-squared	0.06		0.09	

Growth rate is between *t* and *t+1* (annual observations)
Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

As an illustration of these magnitudes, consider the case of Argentina and South Korea. We see in Figure 5.B that the countries started from a similar level of export sophistication, but the gap between the *EXPY* of the two countries widened significantly between 1985 and 1995. How did Korea jump to higher value exports? While equivalent in initial value, *EXPY* hides two very different export baskets in terms of their opportunities for future structural transformation. That of Korea was readily adaptable to many other products with high sophistication, whereas that of Argentina used inputs and capabilities less transferable, meaning that 'open forest' was much lower as can be seen in Figure 5.C. In 1985 Korea achieved comparative advantage in certain paper products, laboratory equipment and furniture, chemicals (polypropylene, alkyds, artificial resins, and polystyrene), trucks, and trailers. These new additions to the export basket added significant opportunities in the product space. Conversely, Argentina's export basket consisted of products with less potential for productive transformation, such as horses, leather, petroleum products and soya bean oil. As such, the open forest of Korea in 1985 was over 60% higher than that of Argentina. According to our most conservative estimate, this difference resulted in over 20 percentage points of subsequent *EXPY* growth differential in Korea as compared to Argentina.

Figure 5.A

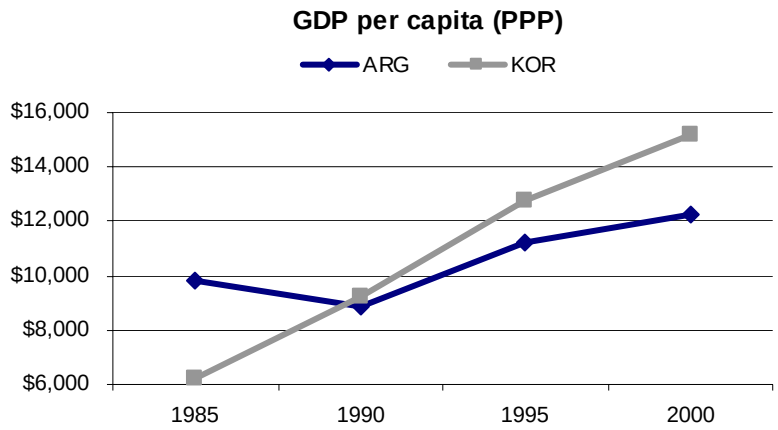


Figure 5.B

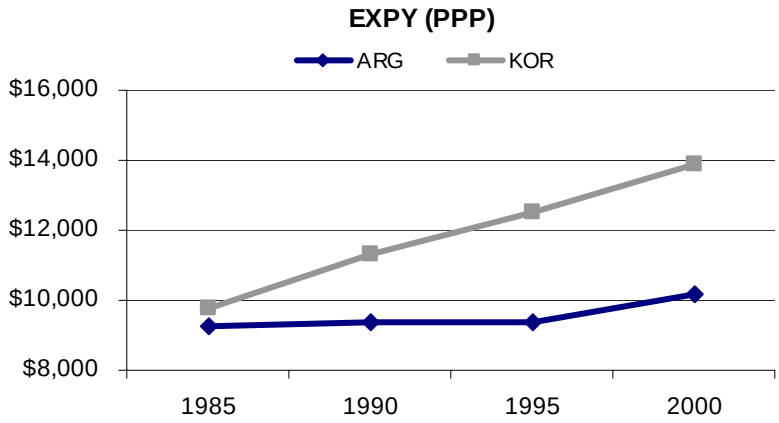
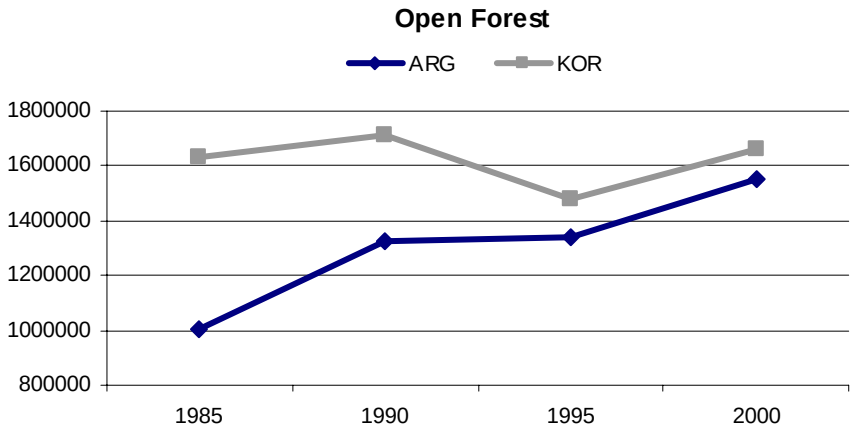


Figure 5.C



Source: Author's Calculations

6. Conclusions

Much recent theory assumes a rather homogeneous and continuous product space. This implies that imitation by developing countries of goods invented in developed countries is always equally possible, *ceteris paribus*. This paper argues that this assumption is inconsistent with the facts. Moreover, it shows that this heterogeneity is not without consequence. The speed at which countries can transform their productive structure and upgrade their exports depends on having a path of nearby goods that are increasingly of higher value. However, countries differ radically in this dimension. This adds another layer of explanation to the differing growth performance across the world. The problem in some countries is that they are specialized in goods that require assets and skills that are very specific to that product and do not prepare the country to move onto other goods. A particular case in point is the case of the oil exporting countries. As figure 2 indicates, they have an unusually low 'open forest', a product of the fact that oil requires very specific endowments and does not easily prepare a country to enter other goods. Similarly, tropical products and other raw materials also have this characteristic. By contrast, light manufactures, electronics and capital goods tend to involve skills and assets that are much closer to those required by other goods and hence facilitate the transition from one product to another.

This creates two types of externalities. If a country develops comparative advantage in a certain good, many firms can enter, producing an intra-industry spillover. In addition, these capabilities now shorten the distance to other goods, producing inter-industry spillovers. We can look into this question by asking what has been the contribution to 'open forest' of each product in which a country developed comparative advantage.

It is highly unlikely that markets under perfect competition would internalize either of these spillovers, which implies inefficiently slow structural transformation, especially in those countries with comparative advantage in a part of the forest space that is sparse.
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 greater flexibility with which the accumulated assets and capabilities could be redeployed from sector to sector.

The work in this paper can be extended in several dimensions. First, 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. Second, it would be interesting to study the evolution of the distance matrix over time. What has globalization in the last few decades done to the relative proximity of different goods? Are there more or fewer paths to structural transformation now? Third, it would be interesting to study the role of economic policy and industrial organization in achieving structural transformation. Were the transitions observed in East Asia the consequence of their predetermined position in the product space or were they related to activist government policies? Are improbable transitions more likely when foreign direct

investment is involved? Does the presence of large local conglomerates help in internalizing some of the externalities highlighted by this paper? Have the jumps to new products been followed by jumps to nearby products? Finally, our study of the proximity matrix could be enhanced by using the tools of network analysis[9].

[9] This work is currently underway with Albert-Lazlo Barabasi & 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.

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

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.

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

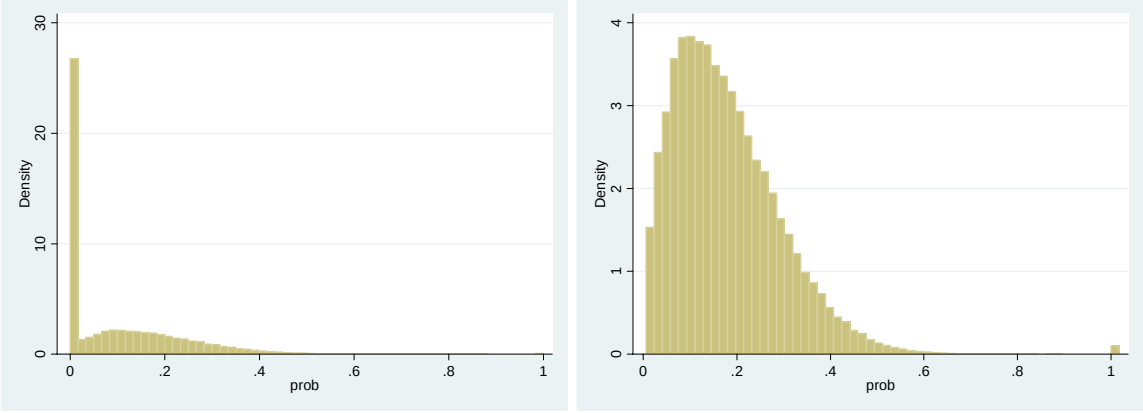
We drop all the artificial 'A' & 'X' product categories from the Feenstra dataset, leaving 1006 products. We drop any countries that reported more than 5% of their total exports in these artificial product categories. We exclude from all regressions countries with a population under 2 million.

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

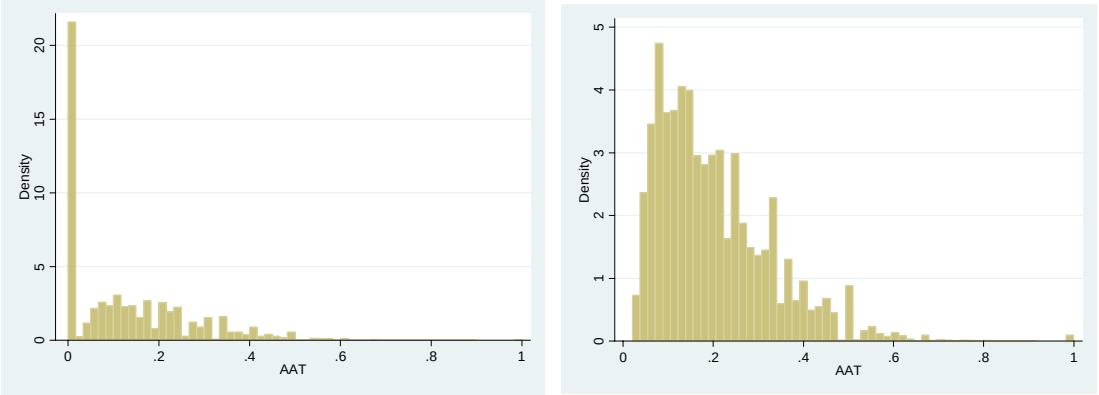


Descriptive Statistics for φ
(1985)

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

We see a similar pattern in proximity for earlier periods. The average proximity is somewhat higher, but the mode continues to be 0 quite strongly. The distribution is not as smooth, but maintains the equivalent distribution.

Histogram of proximity (left) and proximity excluding 0 values (right): 1985



Robustness Checks

Product-Level Regressions: 1975-2000 5-Year Panels

	Probit	Probit	OLS	OLS
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.558	0.548	0.704	0.701
	(87.22)**	(89.01)**	(107.24)**	(109.47)**
$\ln density_{i,c,t}$	0.035	0.031	0.009	0.007
	(10.94)**	(10.52)**	(3.06)**	(2.28)*
$\ln GDPpc_{c,t}$	-0.003	-0.002	-0.001	-0.002
	(0.51)	(0.31)	(0.22)	(0.29)
$\ln EXPY_{c,t}$	-0.038	-0.028	-0.019	-0.014
	(3.70)**	(2.95)**	(2.25)*	(1.65)
$\ln PRODY_{i,t}$	0.017	0.016	0.018	0.017
	(7.62)**	(7.39)**	(8.20)**	(7.94)**
$RCA_{i,c,t}$		0.005		0.006
		(12.51)**		(7.88)**
Constant			0.004	-0.048
			(0.06)	(0.72)
Observations	375276	374270	377823	376816
R-squared			0.57	0.58

Year, country, and product dummies included in all estimations. Probit coefficients are reported as marginal effects. Standard errors clustered by country. Absolute value of z statistics in parentheses.
* significant at 5%; ** significant at 1%

Product-Level Regressions: Restricted Sample to Goods Not Exported in Period t

	1975-2000 (5 yr panels) OLS		1985-2000 OLS	
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
ln*density*$_{i,c,t}$	0.016	0.016	0.016	0.015
	(6.08)**	(6.08)**	(5.40)**	(5.34)**
ln*GDPpc*$_{c,t}$	-0.020	-0.020	0.003	0.002
	(3.61)**	(3.61)**	(0.58)	(0.43)
ln*EXPY*$_{c,t}$	-0.026	-0.026	-0.017	-0.015
	(3.37)**	(3.37)**	(4.19)**	(3.81)**
ln*PRODY*$_{i,t}$	0.011	0.011	0.003	0.002
	(6.48)**	(6.48)**	(2.76)**	(2.55)*
RCA$_{i,c,t}$		0.004		0.002
		(6.75)**		(7.28)**
Constant	0.302	0.302	0.150	0.135
	(4.73)**	(4.73)**	(3.11)**	(2.95)**
Observations	329060	329060	1019013	1017186
R-squared	0.05	0.05	0.02	0.02

Year, country, and product dummies included in all estimations.
Standard errors clustered by country. Absolute value of z statistics in parentheses. * significant at 5%; ** significant at 1%

Product-Level Regressions with Asymmetric Distance

	Probit	Probit	OLS	OLS
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.703	0.700	0.800	0.798
	(107.01)**	(109.24)**	(110.48)**	(110.53)**
ln*density*$_{i,c,t}$	0.011	0.008	0.014	0.012
	(3.38)**	(2.62)**	(4.31)**	(4.18)**
ln*GDPpc*$_{c,t}$	-0.001	-0.001	0.006	0.004
	(0.14)	(0.22)	(1.22)	(0.96)
ln*EXPY*$_{c,t}$	-0.021	-0.015	-0.022	-0.018
	(2.37)*	(1.79)	(4.25)**	(3.69)**
ln*PRODY*$_{i,t}$	0.017	0.017	0.008	0.007
	(8.19)**	(7.98)**	(7.20)**	(7.09)**
RCA$_{i,c,t}$		0.006		0.003
		(7.84)**		(7.83)**
Constant	0.021	-0.034	0.094	0.116
	(0.30)	(0.49)	(1.48)	(.)
Observations	377823	376816	1175839	1173635
R-squared	0.57	0.58	0.70	0.70

Year, country, and product dummies included in all estimations.
Standard errors clustered by country.
Absolute value of z statistics in parentheses.
* significant at 5%; ** significant at 1%

Product-Level Regressions: Alternative Functional Form

To test for non-linearity in the relationship between density and the probability of production, we introduced a dummy variable *Cutoff* equal to 1 if lndensity was above -2.5, and 0 otherwise. The results suggest that a linear relationship between lndensity and the probability of exporting is an acceptable functional form.

	1975-2000 (5 yr panels) OLS		1985-2000 OLS	
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.704	0.700	0.801	0.799
	(106.14)**	(108.44)**	(110.61)**	(110.63)**
ln$density_{i,c,t}$	0.008	0.005	0.013	0.011
	(2.49)*	(1.78)	(4.09)**	(3.91)**
Cutoff$_{i,c,t}$	0.009	0.008	0.000	-0.000
	(2.53)*	(2.32)*	(0.11)	(0.05)
ln$GDPpc_{c,t}$	-0.002	-0.003	0.005	0.003
	(0.39)	(0.46)	(1.05)	(0.78)
ln$EXPY_{c,t}$	-0.021	-0.015	-0.020	-0.017
	(2.45)*	(1.84)	(4.26)**	(3.65)**
ln$PRODY_{i,t}$	0.018	0.017	0.008	0.008
	(8.42)**	(8.14)**	(7.38)**	(7.22)**
$RCA_{i,c,t}$		0.006		0.003
		(7.85)**		(7.81)**
Constant	0.016	-0.037	0.083	0.105
	(0.23)	(0.56)	(1.38)	(.)
Observations	377823	376816	1175839	1173635
R-squared	0.57	0.58	0.70	0.70

Year, country, and product dummies included in all estimations. Standard errors clustered by country. Robust t statistics in parentheses. * significant at 5%; ** significant at 1%

Product-Level Regressions: Excluding U.N 'Special Code' Products

	1975-2000 (5 yr panels) OLS		1985-2000 OLS	
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.731	0.726	0.804	0.801
	(103.84)**	(106.03)**	(109.44)**	(109.24)**
ln$density_{i,c,t}$	0.014	0.011	0.016	0.014
	(4.05)**	(3.39)**	(4.30)**	(4.16)**
ln$GDPpc_{c,t}$	-0.008	-0.008	0.008	0.006
	(1.21)	(1.28)	(1.52)	(1.33)
ln$EXPY_{c,t}$	-0.028	-0.022	-0.023	-0.019
	(2.82)**	(2.30)*	(4.08)**	(3.59)**
ln$PRODY_{i,t}$	0.019	0.019	0.008	0.008
	(8.95)**	(8.77)**	(7.53)**	(7.32)**
$RCA_{i,c,t}$		0.006		0.003
		(7.95)**		(7.87)**
Constant	0.199	0.142	0.119	0.088
	(2.60)*	(1.91)	(1.78)	(1.38)
Observations	334158	334158	1125073	1125073
R-squared	0.60	0.60	0.70	0.70

Year, country, and product dummies included in all estimations. Standard errors clustered by country. Robust t statistics in parentheses. * significant at 5%; ** significant at 1%.

Product-Level Regressions: Corrected for Path Similarity

It may be the case that when calculating density around good A, the proximity between goods connected to it should be considered. For example, imagine there are three goods

surrounding A, two of which are closely connected to one another and one of which not connected to those two. Each good is connected based on a similarity in requisite capabilities, so once you have occupied one of the two closely-linked trees, occupying the other does not provide as many new capabilities that would make you more likely to produce A than would occupying the third product without other close neighbors. lndensity1c includes a scalar to correct for this similarity between products surrounding the product of interest. The scalar is[10]:

$$\theta_{ij} = \frac{\sum_{k} \varphi_{ik}\varphi_{kj}}{\sum_{j}\sum_{k} \varphi_{ik}\varphi_{kj}}$$

The numerator captures the degree of similarity of j to i captured by its linkages through all other products k. The denominator is the sum of all those intermediate linkages. So this captures the degree to which j is linked with i through other channels: if high, then we want good j individually to count for less, and if low, then j is quite unique in its linkages to i and we want its density to be higher. Density1c is therefore:

$$\ln density1c_{c,i,t} = \left(\frac{\sum_{j} \varphi_{i,j,t} x_{c,j,t} \theta_{ij}^{-1}}{\sum_{j} \varphi_{i,j,t} \theta_{ij}^{-1}} \right)$$

The results are largely unaffected by this change:

	1975-2000 (5yr Panels) OLS		1985-2000 (Annual Data) OLS	
	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$	$x_{i,c,t+1}$
$x_{i,c,t}$	0.705	0.701	0.802	0.801
	(106.78)**	(109.41)**	(109.55)**	(109.65)**
lndensity1c$_{c,t}$	0.009	0.007	0.011	0.009
	(2.64)**	(1.96)†	(3.47)**	(3.21)**
ln*GDPpc*$_{c,t}$	-0.002	-0.002	0.003	0.002
	(0.29)	(0.37)	(0.71)	(0.39)
ln*EXPY*$_{c,t}$	-0.019	-0.014	-0.019	-0.015
	(2.21)*	(1.62)	(4.01)**	(3.35)**
ln*PRODY*$_{i,t}$	0.017	0.016	0.007	0.007
	(8.03)**	(7.84)**	(6.79)**	(6.67)**
RCA$_{i,c,t}$		0.006		0.003
		(7.87)**		(7.91)**
Constant	0.019	-0.036	0.081	0.104
	(0.27)	(0.52)	(1.40)	(.)
Observations	377823	376816	1175839	1173635
R-squared	0.57	0.58	0.70	0.70

Year, country, and product dummies included in all estimations
Standard errors clustered by country

[10] We thank Hunt Alcott for bringing this theoretical point to our attention, and Cesar Hidalgo for his idea of the weighting scalar.

Robust t statistics in parentheses
† significant at 10%; * significant at 5%; ** significant at 1%

Open_Forest and EXPY Growth: 1985-2000 Annual Observations

	FE	RE	FE	RE
	EXPY growth	*EXPY* growth	*EXPY* growth	*EXPY* growth
ln$EXPY_{c,t}$	-0.105	-0.050	-0.141	-0.050
	(9.50)**	(7.44)**	(10.52)**	(6.62)**
ln$GDPpc_{c,t}$	0.023	0.010	0.024	0.010
	(2.31)*	(3.73)**	(2.51)*	(3.51)**
ln$open_forest_{c,t}$	0.016	0.013		
	(3.65)**	(5.36)**		
ln$open_forest_size_{c,t}$			0.009	0.013
			(2.16)*	(5.28)**
ln$open_forest_value_{c,t}$			0.076	0.013
			(5.39)**	(1.12)
Constant	0.541	0.206	0.310	0.202
	(6.47)**	(5.76)**	(3.22)**	(2.28)*
Observations	432	432	432	432
Number of countryid	105	105	105	105
R-squared	0.25		0.30	

Growth rate between *t* and *t+1* converted to annual rate
Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

Open_Forest and EXPY Growth: Random Effects, Including Population

	1975-2000 (5yr Panels) RE		1985-2000 (Annual Data) RE	
	EXPY growth	*EXPY* growth	*EXPY* growth	*EXPY* growth
ln$EXPY_{c,t}$	-0.063	-0.060	-0.060	-0.069
	(7.50)**	(7.11)**	(5.65)**	(6.40)**
ln$GDPpc_{c,t}$	0.015	0.014	0.010	0.012
	(4.54)**	(4.40)**	(2.69)**	(3.03)**
ln$open_forest_{c,t}$	0.012		0.016	
	(4.53)**		(4.02)**	
ln$open_forest_size_{c,t}$	0.004	0.004	0.002	0.002
	(2.73)**	(2.72)**	(1.10)	(1.02)
ln$open_forest_value_{c,t}$		0.011		0.008
		(4.29)**		(1.88)
ln$population_{c,t}$		0.021		0.185
		(1.38)		(4.18)**
Constant	0.208	0.118	0.217	-1.250
	(4.82)**	(0.96)	(4.40)**	(3.24)**
Observations	432	432	1434	1434
Number of countryid	105	105	106	106

Growth rate between *t* and *t+1* converted to annual rate
Absolute value of z statistics in parentheses
* significant at 5%; ** significant at 1%