EDUCATION AND THE JOURNEY TO THE CORE:
PATH DEPENDENCE OR LEAPFROGGING?

Aashish Mehta* and Jesus Felipe^

*University of California-Santa Barbara
^Asian Development Bank

Abstract: We use changes in indices of revealed comparative advantage in 1240 products to study whether export diversification is path-dependent (i.e., history matters); whether it is more difficult to diversify into more sophisticated products; and whether education helps reduce path dependence, or helps develop comparative advantage in sophisticated products. Our main sample includes 40 developing and newly industrializing economies in 1995 and in 2010. We find that the development of comparative advantage is a path-dependent process – stepping stones products must usually be developed to approach new products. We also find strong evidence consistent with the view that education helps reduce this path dependence, and much weaker evidence that education is important for developing comparative advantage in sophisticated products. Results also suggest that education quality matters more than education quantity; that countries with high-quality basic education have a diverse export mix; and that high college attainment makes that export mix more fluid. Finally, leapfrogging into complex products is very unlikely for most developing countries.

Keywords: Core, Density, Diversification, Education, Leapfrogging, path dependence, Periphery, Revealed comparative advantage

JEL Codes: I25, O11, O14

Disclaimer: This paper reflects the views of the authors, and not those of the Asian Development Bank, those of its Board of Governors, or those of its member countries.

Acknowledgement: We are grateful to Cesar Hidalgo and Daniel Stock for clarifications and advice on the data. We also thank Zack Grossman, Jan Nederveen Pieterse, Paulina Oliva, Heather Royer, Ashish Thapliyal and Emanuel Vespa for helpful discussions of our results; and Emmanuel Andal, Reynold Galope and Caitlin Vejby for excellent research assistance.
1. Introduction

It has long been argued that a country’s development prospects are circumscribed or enhanced by the types of products that it produces (Chenery et al. 1986; Kaldor 1970; Kuznets 1965; Prebisch 1959; Seers 1962). “Core” products, such as chemicals, complex machinery and advanced scientific instruments, are thought to be intrinsically different from “peripheral” products, such as agricultural and mined commodities. In addition to offering greater learning opportunities, core products are deemed to possess higher income elasticities of demand, greater scope for product differentiation and larger returns to scale than peripheral products. They therefore offer greater opportunities for economic development.

Several recent empirical studies suggest that the argument has merit. They show that economies grow faster when they export a large variety of products (Saviotti and Frenken 2008), when exports are more sophisticated (Hausmann et al. 2007), and when they export a more complex array of products (Hausmann et al. 2011). Additionally, a well-diversified and/or sophisticated economy has been linked to employment that grows faster and is more resilient to shocks (Frenken et al. 2007), to shorter recessions (Hausmann et al. 2006), and to lower inequality (Felipe and Hidalgo 2014).

Another key finding in this literature is that the evolution of a country’s export mix is a path-dependent process, in the sense that it is easier for a country to develop a new comparative advantage in some product if it has already developed a comparative advantage in similar, or “proximate” products (Hidalgo et al. 2007). This has been interpreted to suggest that it is easier for a country to develop comparative advantage in producing a product if it is already an active producer of other products that require similar capabilities, including technological and organizational know-how, infrastructure, and institutions (Hausmann et al. 2011). If the capabilities needed to efficiently produce some target product can only be developed through learning-by-doing in the course of industrial production, then developing comparative advantage in the production of unfamiliar products will be a slow, incremental and difficult process. For example, a garment-producing nation wishing to develop comparative advantage in automobiles may need to acquire the capabilities it needs by first learning to produce, in turn, textiles, toys, small appliances, plastics, basic metal products, electrical devices, simple machinery and auto parts. Each of these stepping stones industries will need to be supported, implying spending considerable time and resources.

Unsurprisingly, governments are keen to either speed up this arduous process, or to bypass it entirely, moving directly into core products – a possibility commonly known as “leapfrogging”. Export-oriented industrialization was motivated in part, by a desire to leapfrog by learning the technologies of multinational firms (Bruton 1998). A frequent question in policy circles is whether it is possible to reduce or eliminate path dependence in export diversification.

---

1 The evidence is most convincing for developing economies. Evidence regarding higher-income economies is more mixed, insofar as some high-growth and high-income countries’ exports are natural-resource intensive (Lederman and Maloney 2003), and concentrated (Hesse 2009).

2 We will explain precisely how proximity is measured in section 3. For now, we note only that it is a measure of the overlap in the set of countries with comparative advantages in both products. Thus, trousers are proximate to coats, but neither is proximate to jet engines.
The existing analyses that reveal incremental, path-dependent export development to be the norm implicitly assume a constant degree of path dependence across countries.

In particular, this paper focuses on the role of human capital in facilitating transitions to unfamiliar products. We hypothesize that some countries may possess the conditions that reduce path dependence, or may be able to create them. A target product will be unfamiliar to a country if the capabilities required to produce it do not overlap significantly with those required by the products in which the country has already specialized (e.g., Bangladesh, being specialized in garments and leather products, would find shoes more familiar than automobiles). Assuming that education is transformed into productive knowledge through its application in the workplace and that educated workers learn faster on the job (Kim and Nelson 2000; Lall 2001), we hypothesize that a country with a more educated workforce will be able to develop stepping-stone industries and the capabilities they require more quickly than a country with a less educated workforce. The more-educated country should therefore be more likely to develop comparative advantage in unfamiliar products during some time interval, and its export diversification process should be less path-dependent.

While testing these learning-by-doing benefits of education, we will also investigate the possibility that more educated countries are more likely to develop comparative advantage in more sophisticated products. A product is sophisticated if the capabilities required to produce it are globally scarce (e.g., watch movements, MRI machines and other core products are sophisticated; plastic toys and traditional textiles are not). More educated countries would be more likely to export sophisticated product under a Heckscher-Ohlin framework in which the output elasticity of skilled relative to unskilled labor is higher in the production of more sophisticated products.

We will explore both whether education matters for learning to produce unfamiliar products, and whether it matters for becoming a producer of more sophisticated products. Differentiating between these two roles may help explain why only some countries with a highly-educated workforce have successfully diversified their exports. For example, if diversification is not path-dependent, and education directly provides the capabilities to produce sophisticated products, then Tunisia’s high college attainment rate could allow it to soon become the world’s next big supplier of jet engines. Moreover, Tunisia might be able to leapfrog directly from its current mix of simple car parts, textiles and tourism into jet-engines, without needing to support intervening, stepping-stone industries. On the other hand, if development is path-dependent, and therefore requires the incremental development of capabilities in stepping-stone industries, a Tunisian jet engine initiative will fail unless it is accompanied by policies to support those stepping-stone industries.

To investigate empirically these two potential roles of education, we use highly disaggregated export data and draw on recent studies (Hausmann et al. 2011; Hidalgo et al. 2007) to derive the metrics we need by country and product. These include a measure of how close a target product is from those products a country exports with comparative advantage (i.e. how familiar it is in that country), and measures of the product sophistication. While both the familiarity and the sophistication of a product are functions of the capabilities required to produce it, only its familiarity varies across countries (e.g., automobile engines are less familiar
for Bangladesh than for Thailand, but are equally sophisticated everywhere). We will exploit econometrically the orthogonality of these two characteristic.³ By combining our familiarity and sophistication measures with estimates of the quality and quantity of education that a country’s workforce possesses, we will ask whether more educated countries were more likely to develop comparative advantage in unfamiliar products and in sophisticated products between 1995 and 2010. We also ask: (i) was the quality or the quantity of education more important for overcoming unfamiliarity and sophistication?; (ii) which level of education (i.e., primary, secondary or college) was more important for overcoming unfamiliarity and sophistication?; (iii) did the challenges posed by unfamiliarity and sophistication differ across products of varying importance for development (a concept to be defined in the next section)?; and (iv) did the challenges posed by unfamiliarity and sophistication differ between emerging, nascent and mature industries (defined by the degree to which the country had already specialized in the products these industries produced in 1995)?

Our analysis will also shed light on the possibility of leapfrogging and on the role that education may play facilitating it. For purposes of this paper, leapfrogging is defined as the ability to develop comparative advantage in core products, regardless of the distance to these products. In other words, we will argue that leapfrogging is possible if and only if the distance between a country’s current export basket and a core product does not affect the country’s ability to develop comparative advantage in that core product in a given time frame (in our case, 15 years).

The remainder of the paper is structured as follows. Section 2 motivates the paper with reference to history, empirics and previous studies. Section 3 describes our data, econometric framework and testable hypotheses. Section 4 presents results, and section 5 concludes.

2. Motivation and Conceptual Framework

History suggests that the connections between education and export diversification require careful consideration. Starting in the 1960s, Japan, Asia’s New Industrializing Economies, and later the Southeast Asian economies and China invested significantly in education, while opening up new export markets and accelerating growth (Wang and Wei 2010; World Bank 1993). The United States’ success in diversifying and upgrading its industrial base through most of the 20th century also required significant expansions in education (Goldin and Katz 2009). The industrial success of European countries such as Germany, Finland and Switzerland is similarly credited in part to their historically solid human capital achievements (Dahlman et al. 2006; Freeman 1995; Polasek et al. 2010). And yet, several counterexamples suggest that even if education is necessary for diversifying and upgrading a country’s production and export structures, it may not suffice: The Philippines enjoyed a long-standing educational advantage over much of Southeast Asia, while many countries in the Middle East and North Africa have better educated populations than most of South Asia - and yet their production mixes do not

³ The correlations between our measure of familiarity around target products (called density and defined in Section 3) and the six measures of sophistication that we use are less than 0.004 in absolute value.
necessarily compare favorably. Or to take another example, Bangladesh’s product mix became ever more concentrated in garments as the country’s education level increased rapidly.

Figures 1a and 1b corroborate these suggestions. They reveal that, in a cross-section, the (log of the) number of products that a developing country exports with comparative advantage is correlated with the average years of schooling and with the cognitive skill of its workforce. However, both correlations are noisy (R² = 0.060 for years of schooling, 0.135 for cognitive skill), suggesting that education neither guarantees, nor is an absolute pre-requisite for diversification. What it takes to leverage educational attainment into a salubrious product mix is therefore an open empirical question.

These historical and empirical results are broadly reflective of the gap between theory and empirics in the wider literature on education and economic development. Human capital theory generally predicts strong relationships between education and growth (Galor and Moav 2004; Lucas 1988; Romer 1990). Yet, even though many empirical studies report large individual returns to education, relationships between changes in aggregate education levels and growth have been much more difficult to detect (Pritchett 2001; Pritchett 2006). 4 The literature offers at least four explanations of this gap. One approach highlights errors in measuring shifts in aggregate educational attainment (Krueger and Lindahl 2001). Another appeals to complementarities between education and experience that are suppressed when aggregating education levels across workers of different ages (Lutz et al. 2008). A third emphasizes that educational expansions do more to create human capital if the education is of a high quality (Hanushek and Woessmann 2008, 2009). Our findings will be quite consistent with this position, as comparing figure 1a with 1b already suggest.

This paper focuses on the fourth explanation for why the relationship between education and economic growth is noisy. This is that the benefits of education for economic development depend upon the range of productive activities to which educated workers may apply their skills, and upon the new activities whose introduction their skills can realistically facilitate. Specifically, we will allow for the possibility that a country’s product mix tomorrow is conditioned by its product mix today; so that countries that initially specialize in products offering opportunities to develop only a limited range of productive capabilities may benefit less from an educated workforce than those that initially specialize in products that permit the development of a wider range of capabilities. In such a path-dependent5 world, the effects of education on the export mix depend upon what a country already produces. This is in the spirit of, for example, Nelson and Winter’s (1982) evolutionary theory.

---

4 Theories predicting a relationship between the aggregate educational level and subsequent growth (e.g., Nelson and Phelps 1966) have generally fared better, but even this relationship is noisy (Krueger and Lindahl 2001).

5 Precisely speaking (see Page 2006), export diversification in our framework is a ‘state-dependent’ process - the state of the system is defined by the set of products in which a country has comparative advantage, and the distribution of likely states in the next period depends upon the state of the system in the current period. This is effectively a Markov process. Full ‘path-dependence’ is a more stringent type of history-dependent process which requires that outcomes in one period depend upon the entire history of outcomes, and the order in which those outcomes were realized matters (i.e., changing the order could change the outcome). As we observe transitions only over one time interval, 1995-2010, the distinction is moot.
This type of critique of simpler human capital theories is not conceptually new (Easterly 2001, Ch. 4, and pp.187-189; Pritchett 2001). However, the relationship between education and changes in a country’s product mix has not been adequately examined. Two exceptions are noteworthy. Ciccone and Papaioannou (2009), show that employment in skill-intensive industries grew faster in countries with more highly educated workers and in those countries that achieved the most rapid educational expansions. However, their study is not concerned with export diversification, path dependence or the role of education in ameliorating it. Agosin et al. (2012) suggest that more educated countries are better able to maintain a diversified export mix in the face of terms-of-trade shocks, but do not study the role of education in overcoming unfamiliarity or sophistication of target products. To the extent that diversifying into core products is important for development, these are important gaps to fill.

3. Data, Definitions, Framework and Hypotheses

We have so far used the terms diversification and upgrading in somewhat generic terms. Here we define them, show how we measure a country’s progress towards achieving them, and how these concepts can be used to investigate the relationship between countries’ education levels and their movements into distant and sophisticated products.

3.1 Data and variable definitions

We use data from the C’entre D’Edtudes Prospectives and d’Informations Internationales (CEPII) on most countries’ exports of 1,240 products during 1995-2010. This period covers the post-globalization era, when variations in trade policies across the countries and products in our sample are less likely to introduce biases. We collected data for up to 92 countries that were not OECD members in 1973. Analyses requiring data on the quality of education in the workforce are limited to 40 of these countries for which we have data on both the quality and quantity of education in the workforce. Regression sample sizes vary depending on data availability for other variables. With two exceptions, discussed below, our results are qualitatively robust to the enlargement of the sample to include OECD member.

We measure the quantity of education in the workforce using average years of schooling, and primary, secondary and college attainment rates in the population aged 15 and up (Barro and Lee 2010). For education quality, we use Hanushek and Woessman’s (2009) estimates of cognitive skill derived from the results of international standardized mathematics and science tests administered to 15-year olds.

Denoting exports by $X$, country $c$’s index of revealed comparative advantage in product $p$ is $RCA_{c,p} \equiv \left( \frac{X_{c,p}}{X_p} \right) / \left( \frac{X_{world,p}}{X_{world}} \right)$. We will say that country $c$ exports product $p$ with comparative advantage (CA), or that country $c$ specializes in product $p$, whenever $RCA_{c,p} \geq 1$.

---

6 http://www.cepii.com/anglaisgraph/bdd/bdd.htm
We define the indicator of revealed comparative advantage as $D_{c,p} = I\{RCA_{c,p} > 1\}$. Using this, we define the proximity $\phi_{p,q}$ between products $p$ and $q$ as the minimum of the proportion of those countries that specialize in $p$ that also specialize in $q$, and the converse proportion:

$$\phi_{p,q} = \min \left\{ \frac{\Pr(D_{c,p} = 1 \mid D_{c,q} = 1)}{\Pr(D_{c,q} = 1 \mid D_{c,p} = 1)} \right\}$$

Proximity is a measure of the overlap in the sets of countries that specialize in both products. This sheds light on which pairs of products are near each other (i.e., rely on similar capabilities) and which are distant from each other.

The density of a country’s (c) export-mix measures how close (far) a commodities not exported with comparative advantage are to the country’s export basket. It is calculated as the ratio of the sum of the proximities between the products not in the export basket and those products with RCA>1 (i.e., the country’s export basket), to the sum of the proximities between the products not exported with comparative advantage and every other product, that is:

$$\text{Density}_{c,p} = \frac{\sum_{p \neq q} D_{c,q} \phi_{p,q}}{\sum_{p \neq q} \phi_{p,q}}$$

A country will have low density around distant (unfamiliar) products and high density around nearby (familiar) products. We will use density to test whether countries tend to develop comparative advantage in nearby products. This is a key variable in our framework, as it will allow us to test our hypotheses about path dependence and leapfrogging.

Let $M$ be a $C \times P$ matrix with each element equal to $D_{c,p}$. Then the number of products in which country $c$ is specializes ($\text{Diversity}$) is:

$$\text{Diversity}_c \equiv \text{colsum}(M)$$

We will check our results for robustness using six measures of product sophistication. $M$ can be used to define two of them. A product has high Ubiquity, and is therefore presumed unsophisticated, if many countries specialize in it:

$$\text{Ubiquity}_p = \text{rowsum} (M)$$

Ubiquity is an imperfect measure of lack of sophistication because some unsophisticated products (e.g., uncut diamonds) are not ubiquitous. Hausmann et al. (2011, p. 24) note that such products tend to be exported by countries that suffer the resource curse, and that are therefore less diversified. They propose a recursive process for extracting information from $M$ that implicitly penalizes products for being exported by less diversified countries, and penalizes countries for exporting ubiquitous products. In the limit, this process yields an index of product complexity ($\text{PCI}_p$), which we take as a second measure of product sophistication.
Hausmann et al. (2007) introduced a third sophistication measure. This is the average of countries’ per capita GDPS \( y_c \) weighted by the degree to which each specializes in \( p \):

\[
ProdY_p = \frac{\sum_c RCA_{c,p} y_c}{\sum_c RCA_{c,p}}
\]

This measure of the “income content of products” reflects the idea that richer countries export sophisticated products.

Given our interest in the role that education plays in helping countries move across products, we have constructed three additional measures of the national human capital content of products, analogous to \( ProdY_p \). The measures are distinguished by the underlying national human capital measure: average years of schooling \( Sch_c \), cognitive skill \( Cog_c \) and college attainment \( Coll_c \):

\[
ProdSch_p = \frac{\sum_c RCA_{c,p} Sch_c}{\sum_c RCA_{c,p}}; \quad ProdCog_p = \frac{\sum_c RCA_{c,p} Cog_c}{\sum_c RCA_{c,p}}; \quad ProdColl_p = \frac{\sum_c RCA_{c,p} Coll_c}{\sum_c RCA_{c,p}}
\]

Finally, we define core and peripheral products. To do so, we first define the connectedness of each product as the sum of its proximities to all other products:

\[
C_p = \sum_{q \neq p} \phi_{p,q}
\]

We classify products as “core” if they are in the top tercile of the distributions of both connectedness and \( PCI_p \), “peripheral” if they are in the bottom tercile of both distributions; and the rest are characterized as “in-between”. To illustrate: most unprocessed agricultural and mined commodities, human hair, jute fibers and electric power are revealed to be peripheral; jet engines, x-ray machines, watch movements, optical devices and machine tools are core products; and paper, electric shavers, hats, copper wire and wine are in-between. In our data set of 1,240 products, 230 are core; 232 are peripheral; and the remaining 778 are in-between.

Table 1 lists our data sources and provides summary statistics. The countries in our sample differ widely in educational attainment and quality, as well as in the sophistication and diversification of their export baskets.

### 3.2 Regression Framework

Let \( S_p \) represent one of our six measures of sophistication, and \( E_c \) represent country \( c \)'s initial education level (capturing a combination of the quality and quantity of education). For ease of

---

\(^7\) Connectedness is the sum of the sum of the proximities between \( q \) and every other product \( p \), that is, the denominator of density in equation (2).
exposition, we initially treat $E_c$ as a scalar, but will eventually replace it with a vector capturing education quantity, quality and their interaction. The density, education and sophistication measures are all scaled to have zero mean and a standard deviation of one. This is done to directly obtain estimates of the beta coefficients, capturing the effect on RCA of one standard deviation differences in each independent variable. This will be helpful for discussing the relative importance of different variables and different roles of education. We estimate variants of the following specification:

$$
(7) \text{RCA}_{c,p,t+1} = \alpha_c + \beta_D \text{Density}_{c,p} + \beta_{ED} (E_c \times \text{Density}_{c,p}) + \gamma S_p + \gamma_{ED} (E_c \times S_p) + \delta \text{RCA}_{c,p,t} + e_{c,p}
$$

where, for simplicity, we omit the time subscript (measured at time $t$) for Density, Education and Sophistication. Density and the initial level of RCA are measured in 1995, education and sophistication are measured in 2000 (the year for which we have internationally comparable data on education quality), and the dependent variable is the RCA in 2010. Export shocks, often related to weather, one-time bulk orders, or shifts in terms-of-trade, tend to be short-lived. We therefore expect $0 < \delta < 1$. The lagged dependent variable introduces bias, but not inconsistency, when estimated by OLS. This model is equivalent to a model in which the dependent variable is the change in the RCA, and we will therefore sometimes interpret coefficients as capturing a variable’s relationship with the change in RCA. Finally, $\alpha_c$ denotes countries’ fixed effects.

This regression framework allows us to test four hypotheses, which together yield answers to our main question, namely, whether more educated countries were more likely to develop comparative advantage in unfamiliar products and in sophisticated products between 1995 and 2010. We state these hypotheses in terms of the alternative (HA1):

**HA1:** Changes in a country’s export mix are path-dependent: $\partial \text{RCA}_{c,p}/\partial \text{Density}_{c,p} > 0$. In other words, we say that acquiring comparative advantage in product $p$ is a path-dependent process if it requires having developed previously capabilities in products similar to product $p$ (recall that Density is measured at time 1995 and the dependent variable in 2010). Since the partial derivative $(\beta_D + \beta_{ED} E_c)$ in equation (7) varies with education, we first test whether changes in the export mix are path-dependent for countries with workforces with average levels of education (i.e., $E_c = 0$). In this case, the alternative hypothesis is simply $\beta_D > 0$. We ask the same question for countries with workforces with education levels one standard deviation above and below the mean (i.e., $\beta_D + \beta_{ED} > 0$ and $\beta_D - \beta_{ED} > 0$, respectively, under the alternative). The three respective null hypotheses (H1) (i.e., changes in a country’s export mix are not path-dependent) are: $\beta_D = 0$, $\beta_D + \beta_{ED} = 0$, and $\beta_D - \beta_{ED} = 0$.

---

8 Replacing the dependent variable with the $\Delta$RCA changes nothing except that the coefficient on $RCA_{c,p,t}$ becomes $(\delta - 1)$, as in an error-correction model.

9 We stress the point made in Section 2 that, strictly speaking and in Page’s (2006) terminology, this is a test for a ‘state-dependent’ process, as our outcome RCA at time (2010=t+1) depends only on the transition at (t=1995).
HA2: Education reduces path dependence: \( \partial^2 RCA_{c,p} / \partial E_c \partial Density_{c,p} < 0 \), i.e., \( \beta_{ED} < 0 \). This occurs if education speeds up the acquisition of missing capabilities, permitting a country to develop comparative advantages in products located at a greater distance from its existing product mix in a given time interval. Another way to state this alternative hypothesis is that education substitutes for density. This somewhat obvious possibility has not, to our knowledge, been examined in the literature. The null hypothesis (H2) is \( \beta_{ED} = 0 \).

Leapfrogging: We define leapfrogging as the ability to develop comparative advantage in core products, independent of a country’s density around that product. The test for whether leapfrogging is possible is directly given by H1 (against HA1) in a sample restricted to core products, that is, \( \partial RCA_{c,p} / \partial Density_{c,p} = 0 \). Inability to reject H1 in this sample will imply that leapfrogging is possible. Likewise, in a sample of core products, inability to reject H2 will imply that education increases a country’s chances of leapfrogging. We note that not rejecting H1 for the group of peripheral products does not constitute leapfrogging. After all, the ability to develop comparative advantage in raw diamonds is rooted in geology, not in capabilities. In-between products are a grey area in this regard.

HA3: Developing comparative advantage in a product is less likely the higher is the level of sophistication of the product: \( \partial RCA_{c,p} / \partial S_p < 0 \). Since the partial derivative in (7) \( (\gamma_S + \gamma_{ES}E_c) \) varies with education, we evaluate whether it is difficult to develop comparative advantage in sophisticated products for countries with whose workforces have an average level of education \( (i.e., E_c = 0) \). In this case, the alternative hypothesis is \( \gamma_S < 0 \). We ask the same question for countries whose workforce have education levels one standard deviation above and below the mean (i.e., \( \gamma_S + \gamma_{ES} < 0 \) and \( \gamma_S - \gamma_{ES} < 0 \), respectively, under the alternative). The null hypothesis (H3) in each case is that the estimated partial derivative is zero. Rejecting the null for a country indicates that its future path is likely to skirt the most sophisticated products.

HA4: Education facilitates the development of comparative advantage in sophisticated products: \( \partial^2 RCA_{c,p} / \partial E_c \partial S_{c,p} > 0 \), i.e., \( \gamma_{ES} > 0 \). This happens if education provides the capabilities needed to produce those products. In this case, education helps shift a country’s export diversification path so that it runs through more sophisticated products. The null hypothesis is \( \gamma_{ES} = 0 \).

To keep things simple, we have explained our testable hypotheses assuming that \( E_c \) is measured by a single variable. For estimation purposes, \( E_c \) will be a vector of variables capturing the quantity and quality of education and also their interaction. This will complicate the calculation of the marginal effects and the implementation of our strategy, but poses no conceptual challenge. For example, in our main specification, education is captured by average years of schooling \( (Y_c) \), a measure of education quality \( (Q_c) \), and their interaction. Both
variables are normalized to have a mean of zero and standard deviation of one. Thus, the extended version of equation (7) becomes:

\[
RCA_{c,p,t+1} = \alpha_c + (\beta_D + \beta_{Y0}Y_c + \beta_{Q0}Q_c + \beta_{YQ0}Y_cQ_c)\times\text{Density}_{c,p}
\]

\[
+ (\gamma_S + \gamma_{Y5}Y_c + \gamma_{Q5}Q_c + \gamma_{YQ5}Y_cQ_c)\times S_p + \delta RCA_{c,p,t} + e_{c,p}
\]

and the partial and cross-partial derivatives become functions of the education variables. Thus, for example:

\[
\frac{\partial RCA_{c,p,t+1}}{\partial \text{Density}_{c,p}} = \beta_D + \beta_{Y0}Y_c + \beta_{Q0}Q_c + \beta_{YQ0}Y_cQ_c
\]

\[
\frac{\partial^2 RCA_{c,p,t+1}}{\partial Q_c \partial \text{Density}_{c,p}} = \beta_{Q0} + \beta_{YQ0}Y_c
\]

By testing our four hypotheses using different vectors of education measures, and in different subsamples, we can also answer four subsidiary questions: (i) Is education quality or quantity more strongly associated with overcoming unfamiliarity (i.e., rejecting H2 in favor of HA2)?; and is education quality or quantity more strongly associated with overcoming sophistication (i.e., rejecting H4 in favor of HA4)?; (ii) what is the relative importance of primary, secondary and college attainment in overcoming unfamiliarity (results on H2) and sophistication (results on H4)?; (iii) Are the effects of unfamiliarity, sophistication and education different (results on H2, H3 and H4) in sub-samples of core, peripheral and in-between products? If distance does not matter for a country’s ability to develop comparative advantage in core products (i.e., H1 is not rejected in that subsample), we will conclude that this country is able to leapfrog; and (iv) Are the effects of unfamiliarity, sophistication and education different (results on H1, H2, H3 and H4) in subsamples containing nascent, emerging and mature national industries? We define nascent industries as those country-product pairs with \(RCA<0.5\) in 1995, mature industries as those with \(RCA>1.5\), and emerging industries as those with \(0.5 \leq RCA \leq 1.5\).

We emphasize that our dynamic framework helps shed light on why some countries have not grown faster despite accumulating human capital. If exporting core products helps countries grow, but countries cannot leapfrog (i.e., H1 is rejected for core products), then a country that initially specializes in peripheral products but cannot or does not build stepping stones industries would be confined to applying its human capital to relatively unproductive activities and would not grow fast (Easterly 2001, Ch. 4; Pritchett 2001; Pritchett 2006).

To confirm the robustness of the results on H3 and H4 to the choice of sophistication measure, we will estimate a version of equation (8) in which \(\gamma_S S_p\) is replaced with product fixed effects. We will also test the robustness of our results using a range of “corner-solution” models, as suggested by Wooldridge (2002, chapter 16), given that RCAs are non-negative.
Before proceeding to estimate the model, it is necessary to comment on identification. Our right-hand-side variables are all predetermined, so reverse causation is unlikely. We will also show that our results are robust to the inclusion of many country-level characteristics. However, we know of no reasonable means of quantifying industrial policies, which may have influenced the evolution of the product mix and the education level of the workforce. Therefore, rather than identifying causal effects, our regressions provide a structured description of countries’ movements into different products over the 15-year period of analysis, with a focus on how these paths differed between countries whose workers had different educational levels. This is a difference in differences approach - we study differences across products between countries whose workforces have different education levels.

4. Results

Our analysis of the relationship between human capital and export diversification proceeds in two steps. The first step (section 4.1) is an analysis of the data at the country level to study the relationship between diversity and education. The results of this exercise are merely suggestive, because the limited time-span covered by our data precludes a robust dynamic analysis at the aggregate level. The second step (section 4.2) uses data at the country-product level and specification (7) and the extended version (8) to ask how national human capital variables influence which products countries develop comparative advantage in.

4.1 Country-Level Analysis

As Figure 1 showed, the cross-sectional relationship between Diversity and years of schooling is positive but noisy. Table 2 shows that this finding is robust to the introduction of basic controls. It provides the results of regressions of \( \ln(Diversity_c) \) on human capital, population and per capita GDP. The regressions in the table examine cross-sectional relationships in 2000 (our cross-sectional measure of cognitive skill captures conditions in that year best), using our five measures of human capital: primary, secondary, and college attainment, years of schooling, and cognitive skill.

All significant control variables take on expected signs. More populous countries have a more diverse export-mix, consistent with the idea that a large domestic consumer base can support a wider array of industries. Calculations using the delta method reveal that the relationship between per capita GDP and diversity is either insignificant (regressions 1, 2 and 4), or, once we correct for cognitive skill (regressions 3 and 5), it displays the familiar inverted U-shape with a peak near the middle of the per capita GDP distribution (Imbs and Wacziarg 2003).

---

10 Lee (1996) tested the impact of a number of industrial policy variables on Korea’s growth. We do not have that wealth of information for the wide sample of countries that we use in our analysis.
11 Results do not change qualitatively if, instead of log-diversity, we use Herfindahl-Hirschmann index of export concentration, i.e., \( HHI_c = 100 \times \sum_p \left( \frac{X_{p,c}}{X_c} \right)^2 \). Diversity and the HHI are obviously related, but capture slightly different features of the national export mix. Results are available upon request.
Countries whose workers have higher average years of schooling have a more diversified export basket in both the larger and smaller sample (columns 1 and 2). When we control for both years of schooling and cognitive skill, the latter is significant, while the former is not (column 3). Regardless of whether we control for cognitive skill or not, only primary education is associated with greater export diversity (regressions 4 and 5). These results, together, indicate that in a cross section, education quantity and quality are both associated with diversity, but diversity is more strongly associated with the quality than with the quantity of education; and with primary more than with secondary or college education. Moreover, the magnitude of these partial correlations can be large: for example, for a country with average years of schooling, a one standard deviation improvement in the cognitive skills (education quality) is associated with a 27.5 percent increase in diversity (regression 3).

4.2 Country-Product Level Analysis

4.2.1 Model selection

RCAs cannot be negative, and they are zero in roughly 15% of the country-year-product observations in our dataset. Although RCAs are neither censored (all are observable) nor truncated (the zeroes are true zeroes), a linear regression model is potentially problematic because it can predict a negative RCA at an observed value of the vector of independent variables. Following Wooldridge (2002, chapter 16) we treat equation (8) as a corner-solution model, estimate it using four different methods, and see whether results vary qualitatively. The four methods are OLS, OLS on a sample restricted to those country-product pairs in which exports are positive in both years, the Tobit model, and an exponential specification estimated using NLS.\textsuperscript{12}

Table 3 provides our main regression estimates for the exports of 40 non-OECD developing countries, while Table A1 (and supplemental tables available upon request) tests their robustness to the choice of corner-solution model in the samples of all, core, in-between and peripheral products, respectively. Table 3 contains seven regressions, all variants of regression (8). The observations are country-product pairs. The regressions in columns (1)-(2) provide results for the full sample of products, and those in columns (3)-(7) provide results for subsets of core, in-between and peripheral products. Each regression involves two sets of terms of interest: density considerations, which allow us to test whether human capital substitutes for density near the target product; and sophistication considerations, which allow us to test

\textsuperscript{12} Formally, if $\beta x$ is a linear combination of our variables (including interaction terms and the lagged dependent variable). Then the NLS model is simply $RCA = \exp(\beta x) + u$; estimated using robust standard errors to permit heteroskedasticity. The NLS model does not include country fixed effects because their inclusion would make it impossible to calculate derivatives for hypothetical countries, analogous to those derived from the OLS and Tobit models. This is because country fixed effects would appear in the derivatives of the NLS conditional expectations function. Instead, we estimate a slightly less general model, including each education variable uninteracted with anything in $x$, in addition to the interactions of the education variables with density and sophistication.
whether human capital is associated with developing comparative advantage in sophisticated products. All seven regressions use multiple education measures, including educational attainment, cognitive skill, and interactions between cognitive skill and average years of schooling. Regression (2) replaces the total number of years of schooling in the regression in column (1) with primary, secondary and college attainment rates.

Beginning with the full sample, Appendix Table A1 provides the derivatives and cross partial derivatives required to test H1-H4. Results in the first column are calculated using the estimates in regression (1) in Table 3; while those in the remaining columns are calculated using the same sample and right-hand-side variables, but using different statistical models to deal with the corner solution problem. The signs of all statistically significant derivatives are the same across specifications, indicating that the choice of statistical model has no qualitative effect on the answers to our questions in the full sample. We have also run these regressions for subsamples restricted to core, in-between and peripheral products (using the same samples and right-hand-side variables as regressions (3)-(7) in Table 3). There is no sign switching in any of the statistically significant derivatives across specifications for in-between and peripheral products (results available upon request). However, one derivative switches signs in the sample of core products, and patterns of significance differ between OLS and NLS models within the subsamples of core and peripheral products (results available upon request).

Therefore, Table 3 reports only the results of the OLS model for the full sample, and for the sample of in-between products, but reports the OLS and NLS results for core and peripheral products (columns 3, 4, 6 and 7). Table 4 summarizes the hypothesis test results for our preferred models. Hypothesis tests are discussed with respect to the NLS models for core and peripheral products, and OLS for all and in-between products. Core products are the most important from a development perspective, and so we will analyze the possibility of leapfrogging using both OLS and NLS to check for robustness.

4.2.2. Results for all products taken together

We begin by discussing the results regarding density considerations in the full sample (Table 3, regressions 1 and 2):

HA1: Changes in a country’s export mix are path-dependent

The coefficient of density is positive and significant in regressions (1) and (2). Therefore, we reject H1 in favor of HA1 for a country whose workers have average levels of education quality and quantity. The derivatives in the first column of Table 4 test whether density matters for countries at the mean, and one standard deviation above or below it in terms of cognitive skill and years of schooling. The null is rejected whenever cognitive skills are at or below the global average. Given that 23 out of the 40 countries in our sample have below-average cognitive skills, and that this sample is probably biased in favor of countries with higher cognitive skill, export diversification is path-dependent for most developing countries.  

13 Cognitive skills and years of schooling are positively correlated in our sample ($\rho=0.63$), and average years of schooling are only 0.69 standard deviations below the global mean in those non-OECD countries
HA2: Education reduces path dependence

The interaction term between density and cognitive skill is negative and significant in regressions (1) and (2). Therefore, we reject H2 in favor of HA2 for cognitive skill – density matters less for countries with better cognitive skills, holding years of schooling constant at their mean level. H2 is also rejected with respect to cognitive skill for countries whose workers have years of schooling that are one standard deviation above and below the mean (Table 4, column 1). On the other hand, the interaction between years of schooling and density is insignificant holding cognitive skill constant at its mean level (Table 3, regression 1); and the cross partial derivatives with respect to density and years of schooling are insignificant, whether cognitive skills are above or below average (Table 4, column 1). Therefore, we cannot reject H2 for years of schooling – years of schooling may not reduce the importance of density. Together, these results suggest that, whatever the level of education quality or quantity, education quality is more important than education quantity for reducing path dependence.

We also examine whether particular levels of education are associated with less path dependence (Table 3, regression 2). College education does play this role. Controlling for secondary and college attainment, higher primary attainment is weakly associated with more path dependence. One possible interpretation of this result is that a country in which many workers have received primary education of a lower quality (insofar as it has not enabled them to obtain more secondary and college education), will experience more path dependence. This may be a further indication that education quality is more important than education quantity. It is worth emphasizing the complementarity between these results and those at the country-level. Together, they indicate that high levels of primary education are conducive to having a diverse export mix, but that countries with more college education are likely to change the products in the export mix more easily.

HA3: Developing comparative advantage in a product is less likely the higher is the level of sophistication of the product

Sophistication is, on its own, statistically insignificant (Table 3, regressions 1 and 2). Table 4, column 1 (results on H3) shows that countries are not less likely to develop comparative advantage in sophisticated products (i.e., null hypothesis is not rejected), so long as cognitive skills are at or above the global average. However, the sophistication of the target product is an impediment to the development of comparative advantage (i.e., H3 is rejected) for countries with below average cognitive skill.

HA4: Education facilitates the development of comparative advantage in sophisticated products

The null is largely rejected in favor of this alternate hypothesis. The interaction between sophistication and cognitive skill (Table 3, regressions 1 and 2) is positive and significant. This implies that a country with a workforce with average number of years of schooling will be more successful in taking sophisticated products if its education quality is good. However, the cross partial derivatives involving sophistication (Table 4, column 1, H4) show that cognitive skill for which we do not have measures of cognitive skill, but 0.20 above it for those in which cognitive skills are measured.
improves the chances of developing comparative in more sophisticated products only for countries whose workers possess below average years of schooling. Education quantity does not help overcome sophistication even when different levels of schooling are allowed to have separate effects (Table 3, regression 2).

Together, these results confirm the logic that education plays a role in developing competitive advantage in more sophisticated and less familiar products. However, the point estimates in Tables 3 and 4 suggest that education is far more important for overcoming lack of familiarity (i.e., low density) than it is for overcoming sophistication. Specifically, we can consider the implications of Table 3, regression 1, for a country with a workforce with average years of schooling, and a target product of sophistication one standard deviation above the mean, around which the country’s export mix has a density that is one standard deviation below the mean. In this case, a one standard deviation difference in cognitive skill is associated with an RCA that is higher by 0.702 because of the effects of cognitive skill through density. On the other hand, a one standard deviation difference in cognitive skills is associated with an expected RCA that is only 0.116 higher because of its effects through sophistication. Similarly, from Table 4, we can see that under these circumstances, a one standard deviation increase in cognitive skill in a country with a workforce with years of schooling one standard deviation below the mean, would lead to an expected increase in RCA of 0.738 through density effects, and of 0.226 through sophistication effects. This relative importance of education-density interactions is further underscored by the lack of a role for education quality in overcoming sophistication.

To see whether the limited economic significance of sophistication is a function of the way we have measured it (through $ProdCog$), Appendix Table A2 provides analogs to regression (1) using all six measures of sophistication. Regression (1) from Table 3 reappears as Regression (5) in Appendix Table A2. Density, as well as its interactions with cognitive skill, continues to be important. Sophistication on its own is never significant. Under two other sophistication measures ($ProdSch$ and $ProdColl$), we find that years of schooling is more closely associated with specialization in more sophisticated products than it is in the regressions we have analyzed so far. However, cognitive skill is not helpful for overcoming sophistication in these regressions. These results reconfirm that a country’s density around a product is a better predictor of whether it will develop a comparative advantage in that product than is the product’s level of sophistication; and that the role of education in overcoming sophistication is less important than its role in overcoming unfamiliarity.

Appendix Table A3 shows what happens when the direct effects of sophistication are captured using 2-digit product fixed effects in the OLS specifications. This analysis is motivated by the fact that we measure product sophistication indirectly, i.e., embodied in the characteristics of countries that specialize in these products. If using product fixed effects to capture the direct barrier posed by product sophistication changed the results substantially, we would argue that this might reflect conceptual problems with these sophistication measures. However, results in Tables 3 and A3 are broadly similar.

Based on these results, and generalizing across products, the answers to our overarching questions are as follows: (i) export diversification is generally path-dependent, but sophistication is only a barrier to the development of comparative advantage for countries with
less educated workers; (ii) education is more important for learning how to produce unfamiliar (i.e., distant) products than for learning how to produce sophisticated products. Also generalizing across products, we note a stronger role for education quality than quantity, and some indication that countries whose workers have more college education find it easier to alter their product mix.

We now turn to more detailed analyses of subsets of products and industries to answer our subsidiary questions.

4.2.3. Results for subsets of products

Core Products:

Regressions (3) and (4) in Table 3, and Table 4, column 2 provide the results for core products. As noted, we focus on the NLS results although the different estimation methods produce very similar outcomes. The OLS and NLS coefficients on density alone (Table 3) and the NLS derivatives with respect to density for countries with average or below-average cognitive skill levels (Table 4) are all positive. This leads to the rejection of H1 for all but the most cognitively advanced developing countries, which implies that it is difficult for countries with below-average cognitive skills to develop comparative advantage in distant core products.

In contrast to these density effects, and the same as for the whole sample, the sophistication of a target core product is not an impediment to develop comparative advantage in it, in particular for countries with education quantity and quality above the mean. It is only an impediment for a workforce with average or below-average education quantity and quality (Table 3, and tests of H3 in Table 4).

The fact that H1 is rejected for core products for education levels commonly observed in developing countries suggests that path dependence is often inescapable when attempting to specialize in core products. In other words, leapfrogging (moving directly into core products without having to first develop stepping stone industries) appears to be a rarity. Of course, it is possible that a few countries have developed sufficient education to leapfrog. To examine this possibility, we calculated, separately for each country, the derivatives of the change in RCA with respect to density, along with their standard errors, using both the OLS and NLS results. Figures 2a and 2b provide the resulting 95% confidence intervals, which test whether H1 can be rejected for core products, given individual countries’ actual human capital endowments. They show that leapfrogging is highly unlikely (i.e., zero lies outside the 95% confidence interval) for most countries. We reject the possibility of leapfrogging for all but seven countries using OLS and for all but twelve countries using NLS. There are six economies for which we cannot reject the null that leapfrogging is possible in both the OLS and NLS specifications: Singapore, China, Hong Kong, Korea, Armenia and Albania. Only Singapore’s point estimate is negative, and only in the NLS specification. In the cases of Armenia and Albania, the derivative with respect to density is quite high, and the null is not rejected because the confidence intervals are large. The four remaining economies have exceptionally high cognitive skill levels – between 0.7 (China) and 1.37 (Korea) standard deviations above the global average. These results confirm that
leapfrogging over a 15 year interval may be possible, but only in extremely well-educated developing countries.\textsuperscript{14}

Figure 3 provides confidence intervals analogous to those in Figure 2a, calculated from an OLS regression using only core products but including countries that were OECD members in 1973. Their inclusion alters our perception of the feasibility of leapfrogging - Singapore, Armenia and Albania are the only countries with sufficient education so that the possibility of leapfrogging is not rejected in Figure 3. This is because the mix of core products that the OECD countries export changed less between 1995 and 2010 than the core-product mix of the non-OECD countries, and the OECD countries were more educationally advanced. Including OECD countries therefore causes us to overestimate the degree of path dependence faced by non-OECD countries, and underestimates their chances of overcoming a lack of density around target products.\textsuperscript{15}

The finding that leapfrogging is extremely unlikely during a 15-year period for all but the best educated developing countries is important. It implies that countries that have been unable to fix their education systems but wish to develop comparative advantage in core products will first need to build a series of stepping-stones industries in order to acquire the needed capabilities.

In-between Products:

Results when we restrict the sample to in-between products (Table 3, regression 5; Table 4, column 3) are qualitatively the same as those in the full sample of all products.

Peripheral Products:

We refer to the NLS results for peripheral products, given the large numbers of country-product pairs with zero exports (Table 3, regression 7; Table 4, column 4). As with other products, we find that developing a comparative advantage in peripheral products is path-dependent (density

\textsuperscript{14} We have attempted to test for leapfrogging using a framework that allows for differences between countries’ experiences that are due to factors beyond education. To do this, we restricted the sample to core products and then regressed the RCA index in 2010 on the RCA index in 1995 (each country at a time), density around that product in 1995 and \textit{ProdCog}. Unfortunately, many countries (particularly low income countries) export very few core products, and some countries have extremely large RCAs. As a result, results using these single-country subsamples become too sensitive to modeling choices and outliers to permit meaningful analysis. Standard errors are also often large. However, amongst the 101 countries which export non-zero quantities of at least half of all core products, only five returned negative point-estimates for the derivative with respect to density using an NLS specification. This suggests, crudely, that while individual country’s export diversification paths are difficult to describe using regressions, leapfrogging is broadly unlikely.

\textsuperscript{15} We do not present results from the NLS model for similar reasons. Rich countries had largely established their comparative advantage in core products by 1995. They therefore had a much higher initial density around most core products than non-rich countries, and also experienced smaller increases in RCAs in core products. Combining rich and non-rich countries, while excluding country fixed effects, therefore provides an estimate of the effect of density around core products that is biased downward. This is readily observable from a comparison of OLS estimates with and without country fixed effects. The NLS model does not readily accommodate country fixed effects, because fixed effects have a different interpretation in an exponential specification, shifting both first and higher moments of the RCA distribution.
matters, H1 is rejected against HA1, for countries whose workforces are within one standard deviation of the global average education level). Also, as before, moving into the more sophisticated amongst the peripheral products is only more difficult for a country with low education quality and quantity (i.e., H3 is rejected in favor of HA3 only when both quality and quantity are low, Table 4).

The effects of education, however, are different in peripheral products. Education does not seem to do much to overcome unfamiliarity when learning how to produce peripheral products (H2 is never rejected). When developing comparative advantage in peripheral products, either quality or quantity is helpful for overcoming product sophistication (H4 is rejected in favor of HA4 for both quality and quantity, Table 4). These results suggest, quite reasonably, that it is possible to pick up the capabilities needed to produce peripheral products directly; and that if learning-by-doing is necessary to move into peripheral products, education does not greatly aid this process. This may be because a country’s capacity to produce peripheral products depends heavily on factors such as geography or geology, or because peripheral products have low knowledge content.

4.2.4. Nascent, Emerging and Mature Industries.

Table 5 reports the results when our two baseline specifications are estimated for subsamples of nascent, emerging and mature industries (country-product pairs), defined by the country’s initial (1995) RCA in the product: industries with 0<RCA<0.5 are nascent, industries with 0.5≤RCA<1.5 are emerging, and industries with 1.5≤RCA are mature. Industries with zero exports in 1995 are excluded from the nascent industry sample, because many of them may be zero for reasons such as geology, climate, etc., and therefore not amenable to change. Given that these subsamples are defined by industries’ initial RCAs, the magnitudes of the regression coefficients cannot be compared across columns. Unsurprisingly, the number of nascent industries with RCAs of zero in 2010 is quite large (38%). For simplicity, we therefore present Tobit estimates.

Pseudo R-squared results indicate that it is roughly five times easier to predict changes in RCAs in mature industries than it is in emerging industries, and that this task is harder still in nascent industries. One interpretation of this result is that it is fundamentally difficult to pick winners (i.e., industries with growth potential), especially among nascent industries.

Qualitatively, the role of education in advancing RCA is mostly invariant to the level of industrial development. Education quality matters greatly for overcoming unfamiliarity in all industries, while quantity plays a role only in emerging industries. College also is associated with overcoming unfamiliarity. One important distinction is the role of cognitive skill in developing comparative advantage in sophisticated products. Cognitive skill helps overcome sophistication in mature industries competing in a market in which the majority of the competition has high cognitive skill. This suggests that while getting a foothold in an industry may involve more learning-by-doing, higher degrees of success in international competition for sophisticated product markets requires high-quality education, independent of what one already produces.
4.2.5. Other Institutional determinants of export diversification.

As a final robustness check, we experimented with including in our regressions different measures of national institutional quality (Table 6). We maintained the country fixed effects, and kept all the same terms as in Table 3, regression 1. Where the institutional variables permitted, we also used exactly the same sample as in this regression. We experimented with nine different institutional measures, one at a time. For each institutional variable, we added the interactions of that variable with density and with sophistication. The variables are: roads per square kilometer in 2005 from the WDI; measures of regulatory quality and rule of law in 2000 from the World Bank’s World Governance Indicators; the number of days and the number of procedures required in 2003 to open a simple business, from the World Bank’s doing business surveys; an infrastructure index from the 2004 Global Competitiveness Index; and indices of information and communications technology infrastructure, sanctity of contract, and freedom from corruption from the 2001 Growth Competitiveness Index.

The results do not alter our earlier conclusions. The interactions between eight of these nine variables and density carry statistically insignificant coefficients. Curiously, a higher number of procedures required to open a business significantly reduces the role of density, and thus path dependence. Similarly, the interactions between eight of these variables and sophistication do not enter significantly. Again, curiously, a higher number of days required to open a business is marginally associated with a higher probability of specializing in sophisticated products. Most importantly, the signs and significance on all our education terms remain exactly as they were in Table 3, regression 1, no matter which institutional control is added.16

These results suggest that education and learning-by-doing are extremely important, and that none of our measured institutional variables sheds light on the speed at which countries develop new comparative advantages, or on the types of products in which they develop comparative advantage.

5. Conclusions

We have investigated the role that education plays in helping countries develop comparative advantage in new products. Our analysis confirms the view that development of comparative advantage is a path-dependent process – stepping stones products must usually be developed to approach new products. Having said this, there is strong evidence that education helps reduce this path dependence by facilitating more rapid incremental movement across the stepping stones of industrial development. Put differently, education helps countries overcome the

---

16 A few measures of institutional quality are statistically associated with greater path dependence when OECD countries are included. This is because having already specialized in core products by 1995, OECD countries had less dynamic export mixes; but also scored higher on these indicators of institutional development. To the extent that the reduced dynamism of OECD export mixes reflects unmeasured historical influences, comparisons of OECD with non-OECD countries provide a biased estimate of the effects of education and institutions on path dependence in non-OECD countries.
unfamiliarity of target products. In contrast, the sophistication of target products is generally only an impediment for those countries with the least educated populations.

The analysis also shows that education quality matters more than education quantity. Likewise, there is strong evidence that high quality basic education matters for export diversification, and that conditional on the quality of basic education, college and secondary attainment are important. The dynamics of export diversification are fairly similar in emerging and mature industries, while the development of comparative advantage in nascent industries is extremely difficult to predict. Finally, countries require only modest amounts of education to overcome barriers posed by the sophistication of peripheral products, while education does not help overcome distance in peripheral products. Core products pose very different problems: while product sophistication amongst core products is only an impediment to poorly educated countries, high quality education is required to significantly overcome unfamiliarity with core products. Perhaps most importantly, we find evidence that only the best educated amongst the developing and newly industrialized countries were able to leapfrog into core products.

The main implications of these results are as follows: First, countries wishing to parlay their human capital stock into industrial development will need to consider carefully the path dependence that is in store. The most difficult product markets to crack are difficult for a reason – they involve severe learning-by-doing. Stepping stone industries will therefore need to be supported, and attempts to move directly into sophisticated industries without first developing stepping stone industries may reflect excess optimism. Second, education that endows high-school graduates with strong cognitive skills could be more important for development through export diversification than ratcheting up the quantity of education. Countries committed to growth through human capital accumulation need to refocus their efforts on increasing education quality, and should think carefully about which stepping stones industries to support so that workers can acquire useful knowledge in the workplace.
References

Agosin, M.R., Alvarez, R., Bravo-Ortega, C., 2012. Determinants of export diversification around the world: 1962–2000. The World Economy 35, 295-315.

Barro, R.J., Lee, J.-W., 2010. A new data set of educational attainment in the World, 1950-2010, National Bureau of Economic Research Working Paper 15902.

Bruton, H.J., 1998. A reconsideration of import substitution. Journal of Economic Literature, 903-936.

Chenery, H., Robinson, S., Syrquin, M., 1986. Industrialization and Economic Growth. Oxford University Press, Oxford.

Ciccone, A., Papaioannou, E., 2009. Human capital, the structure of production, and growth. The Review of Economics and Statistics 91, 66-82.

Dahlman, C.J., Routti, J., Ylä-Anttila, P., 2006. Finland as a knowledge economy: elements of success and lessons learned.

Easterly, W., 2001. The elusive quest for growth: economists adventures and misadventures in the tropics. MIT Press, Cambridge, MA.

Felipe, J., Hidalgo, C.A., 2014. Economic Diversification: Implications for Kazakhstan, Manuscript.

Freeman, C., 1995. The ‘National System of Innovation’ in historical perspective. Cambridge Journal of Economics 19, 5-24.

Frenken, K., Van Oort, F., Verburg, T., 2007. Related Variety, Unrelated Variety and Regional Economic Growth. Regional Studies 41, 685-697.

Galor, O., Moav, O., 2004. From physical to human capital accumulation: Inequality and the process of development. The Review of Economic Studies 71, 1001-1026.

Goldin, C.D., Katz, L.F., 2009. The race between education and technology. Harvard University Press.

Hanushek, E.A., Woessmann, L., 2008. The role of cognitive skills in economic development. Journal of Economic Literature 46, 607-668.

Hanushek, E.A., Woessmann, L., 2009. Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. National Bureau of Economic Research.

Hausmann, R., Hidalgo, C., Bustos, S., Coscia, M., Chung, S., Jimenez, J., Simoes, A., Yildirim, M.A., 2011. The Atlas of Economic Complexity: Mapping Paths to Prosperity. Center for International Development, Harvard University., Cambridge, MA.

Hausmann, R., Hwang, J., Rodrik, D., 2007. What you export matters. Journal of Economic Growth 12, 1-25.

Hausmann, R., Rodriguez, F.R., Wagner, R.A., 2006. Growth Collapses, CID Working Paper No. 136.

Hesse, H., 2009. EXPORT DIVERSIFICATION AND ECONOMIC GROWTH, in: Newfarmer, R.S., Shaw, W., Walkenhorst, P. (Eds.), BREAKING INTO NEW MARKETS: Emerging lessons for export diversification. World Bank, Washington, D.C., pp. 55-80.
Hidalgo, C.A., Klinger, B., Barabási, A.L., Hausmann, R., 2007. The product space conditions the development of nations. Science 317, 482-487.

Imbs, J., Wacziarg, R., 2003. Stages of diversification. American Economic Review, 63-86.

Kaldor, N., 1970. THE CASE FOR REGIONAL POLICIES*. Scottish Journal of Political Economy 17, 337-348.

Kim, L., Nelson, R.R., 2000. Technology, learning, and innovation: Experiences of newly industrializing economies. Cambridge University Press.

Krueger, A.B., Lindahl, M., 2001. Education for growth: Why and for whom? Journal of Economic Literature 39, 1101-1136.

Kuznets, S., 1965. Modern economic growth: Rate, structure and spread. Yale University Press, New Haven.

Lall, S., 2001. Competitiveness, technology and skills. Elgar: Northampton, MA, USA.

Lederman, D., Maloney, W., 2003. Trade Structure and Growth. World Bank Policy Research Working Paper.

Lee, J.-W., 1996. Government interventions and productivity growth. Journal of Economic Growth 1, 391-414.

Lucas, R.E., 1988. On the mechanics of economic development. Journal of Monetary Economics 22, 3-42.

Lutz, W., Cuaresma, J.C., Sanderson, W., 2008. The Demography of Educational Attainment and Economic Growth. Science 319, 1047-1048.

Nelson, R., Winter, S., 1982. An evolutionary theory of economic change.

Nelson, R.R., Phelps, E.S., 1966. Investment in humans, technological diffusion, and economic growth. American Economic Review 56, 69-82.

Page, S.E., 2006. Path dependence. Quarterly Journal of Political Science 1, 87-115.

Polasek, W., Schwarzauer, W., Sellner, R., 2010. Human capital and regional growth in Switzerland. Reihe Ökonomie/Economics Series, Institut für Höhere Studien (IHS).

Prebisch, R., 1959. Commercial Policy in the Underdeveloped Countries. The American Economic Review 49, 251-273.

Pritchett, L., 2001. Where has all the education gone? World Bank Econ. Rev. 15, 367-391.

Pritchett, L., 2006. Does learning to add up add up? The returns to schooling in aggregate data?, in: Hanushek, E., Welch, F. (Eds.), Handbook of the Economics of Education. North-Holland, pp. 635-695.

Romer, P.M., 1990. Endogenous technological change. J. Polit. Econ., S71-S102.

Saviotti, P., Frenken, K., 2008. Export variety and the economic performance of countries. Journal of Evolutionary Economics 18, 201-218.

Seers, D., 1962. A model of comparative rates of growth in the world economy. The Economic Journal 72, 45-78.

Wang, Z., Wei, S.-J., 2010. What accounts for the rising sophistication of China's exports?, China's Growing Role in World Trade. University of Chicago Press, pp. 63-104.

23
Wooldridge, J.M., 2002. Econometric analysis of cross section and panel data. The MIT press.

World Bank, 1993. The East Asian Miracle, Washington, DC.
A country’s RCA in a particular product is that product’s share of the country’s exports, divided by the product’s share in the global exports. Sample includes 40 countries that were not OECD members in 1973, for which we have data on years of schooling and cognitive skill. Cognitive skill is measured in standard deviations from the global mean.
Figure 2: Derivative of RCA with respect to density, by country: Core products

Means and 95% confidence intervals

Note: Derivatives and confidence intervals calculated using results from Table 3, regressions (3) and (4), at the average value of density, sophistication and lagged RCA in the sample used in these regressions. Values of years of schooling and cognitive skill are as observed for the country in question. Morocco has been excluded. This is because its confidence interval is very large (but does not contain zero).
Figure 3: Derivative of RCA with respect to density, by country: Core products

OLS Estimates Including OECD Countries, Means and 95% confidence intervals

Note: Derivatives and confidence intervals calculated using results from OLS regressions, including 61 countries. Values of years of schooling and cognitive skill are as observed for the country in question. Morocco has been excluded. This is because its confidence interval is very large (but does not contain zero).
## Table 1: Summary Statistics, pre-scaling

| Source | Year | N  | Mean | Std. Dev | Min. | Max. |
|--------|------|----|------|----------|------|------|
| **Country-level variables** (Potential sample restricted to countries with education quality and quantity data) |      |    |      |          |      |      |
| Diversity (No. of products out of 1,240) | Calculated from CEPII data | 2000 | 61   | 271.5 | 121.5 | 19   | 532  |
| Herfindahl-Hirschman Index (0-100) | Calculated from CEPII data | 2000 | 61   | 6.99  | 12.45 | 0.72 | 72.98 |
| Cognitive Skill | Hanushek & Woessman (2009) | 2000 | 61   | 4.55  | 0.59  | 3.09 | 5.34 |
| Years of Schooling | Barro & Lee (2010) | 2000 | 61   | 8.90  | 2.05  | 3.89 | 12.71 |
| Primary Attainment (%) | Barro & Lee (2010) | 2000 | 61   | 81.9  | 14.8  | 41.3 | 99.6 |
| Secondary Attainment (%) | Barro & Lee (2010) | 2000 | 61   | 42.3  | 18.3  | 5.3  | 78.8 |
| College Attainment (%) | Barro & Lee (2010) | 2000 | 61   | 8.8   | 5.7   | 0.5  | 26.4 |
| Population (in millions) | World Development Indicators | 2000 | 61   | 73.7  | 209.2 | 1.4  | 1262.6 |
| Real Per capita GDP (US$) | World Development Indicators | 2000 | 60   | 17,156.0 | 12,197.1 | 43,974.8 |

| Source | Year | N  | Mean | Std. Dev | Min. | Max. |
|--------|------|----|------|----------|------|------|
| **Product-level variables** (Potential sample includes all countries with export data and with data on education or per capita GDP) |      |    |      |          |      |      |
| ProdY (US$) | Calculated from CEPII data | 1995 | 1,240 | 8,941.8 | 4,910.6 | 284.3 | 23,286.0 |
| PCI | Calculated from CEPII data | 1995 | 1,240 | 1.1 | 1.8 | -4.3 | 5.4 |
| Ubiquity | Calculated from CEPII data | 1995 | 1,240 | 14.0 | 2.8 | 6.2 | 23.1 |
| ProdYrSch (Years of schooling) | CEPII, Barro & Lee | 1995 | 1,240 | 7.8 | 1.3 | 1.4 | 10.9 |
| ProdCog | CEPII, Hanushek & Woessman | 1995 | 1,240 | 4.6 | 0.3 | 3.4 | 5.3 |
| ProdColl (%) | CEPII, Barro & Lee | 1995 | 1,240 | 6.6 | 2.3 | 0.7 | 16.8 |

| Source | Year | N  | Mean | Std. Dev | Min. | Max. |
|--------|------|----|------|----------|------|------|
| **Industry (Country-Product) level Variables** (Potential sample restricted to countries with education quality and quantity data) |      |    |      |          |      |      |
| Density (range: 0-1), All Industries, of which | Calculated from CEPII data | 1995 | 75,640 | 0.218 | 0.118 | 0.004 | 0.738 |
| Peripheral products | 14,152 | 0.217 | 0.118 | 0.006 | 0.714 |
| In-between Products | 47,458 | 0.218 | 0.118 | 0.004 | 0.738 |
| Core Products | 14,030 | 0.218 | 0.118 | 0.004 | 0.724 |
| Nascent Industries | 51,131 | 0.176 | 0.096 | 0.004 | 0.588 |
| Emerging Industries | 13,464 | 0.288 | 0.110 | 0.011 | 0.636 |
| Mature Industries | 11,045 | 0.325 | 0.113 | 0.009 | 0.738 |
| RCA, all industries, of which: | Calculated from CEPII data | 1995 | 75,640 | 1.14 | 6.06 | 0.00 | 456.51 |
| Peripheral products | 14,152 | 1.05 | 5.14 | 0.00 | 284.70 |
| In-between Products | 47,458 | 1.15 | 5.98 | 0.00 | 456.51 |
| Core Products | 14,030 | 1.18 | 7.09 | 0.00 | 377.70 |
| Nascent Industries | 51,131 | 0.11 | 0.14 | 0.00 | 0.50 |
| Emerging Industries | 13,464 | 0.89 | 0.28 | 0.50 | 1.50 |
| Mature Industries | 11,045 | 6.17 | 14.86 | 1.50 | 456.51 |
| RCA (All Industries) | Calculated from CEPII data | 2010 | 76,880 | 1.16 | 6.45 | 0.00 | 786.40 |

Note: Unless otherwise indicated, variables are unitless indices.
**Table 2: Determinants of aggregate export diversification/concentration**

| Dependent Variable: | Log-Diversification |
|---------------------|---------------------|
| (1) | (2) | (3) | (4) | (5) |
| Years of schooling | 0.561*** | 0.381** | 0.146 |
| Cognitive skill | 0.275*** | 0.252** |
| Years of schooling x Cognitive Skill | 0.086 | 0.147* |
| Primary attainment | | 0.517*** | 0.407** |
| Secondary attainment | -0.012 | -0.208 |
| College attainment | -0.073 | -0.123 |
| Log(Per capita GDP) | 0.061 | 0.453 | 0.679** | 0.099 | 0.667* |
| Log(Per capita GDP) squared | -0.040 | -0.147 | -0.228** | -0.029 | -0.215** |
| Log(Population) | 0.230*** | 0.155** | 0.121** | 0.224*** | 0.128*** |
| Constant | 7.571*** | 6.705*** | 6.369*** | 7.389*** | 6.348*** |
| Sample Size | 89 | 39 | 39 | 89 | 39 |
| R-Squared | 0.367 | 0.329 | 0.477 | 0.36 | 0.554 |

Note: * p<0.1, ** p<0.05, *** p<0.01. OLS coefficients using data from the year 2000 only. Standard errors are robust in all specifications. Each human capital variable has been scaled by its standard deviation in the largest sample for which it is available.
### Table 3: Education and Changes in Revealed Comparative Advantage (1995-2010), Regression Coefficients

| Density Considerations                      | All Products, human capital proxied by: | Product type: |
|---------------------------------------------|----------------------------------------|---------------|
|                                             | Years & quality of schooling (OLS)     | Core (OLS)    |
|                                             | Years, quality level of schooling (OLS)| Core, (exponential model, NLS) | In-Between (OLS) | Peripheral (OLS) | Peripheral (exponential model, NLS) |
|                                             | (1)                                    | (3)           | (4)           | (5)             | (6)             | (7)             |
| Density                                    | 0.911***                               | 1.148**       | 0.480***      | 0.874**         | 0.540           | 0.545***        |
|                                             | 0.682**                                |               |               |                 |                 |                 |
| Density x Years of Schooling               | 0.095                                  | -0.190        | 0.052         | 0.165           | 0.066           | -0.022          |
|                                             |                                        |               |               |                 |                 |                 |
| Density x Cognitive Skill                  | -0.702***                              | -0.680**      | -0.382***     | -0.742***       | -0.465*         | -0.015          |
|                                             |                                        |               |               |                 |                 |                 |
| Density x Years of Schooling x Cognitive Skill | 0.036                                | 0.148         | 0.568         | 0.015           | -0.051          | -0.184          |
|                                             |                                        |               |               |                 |                 |                 |
| Density x Primary Attainment               | 0.735*                                |               |               |                 |                 |                 |
| Density x Secondary Attainment             |                                        |               |               |                 |                 |                 |
| Density x College Attainment               |                                        |               |               |                 |                 |                 |

### Sophistication Considerations

| Sophistication Considerations               | All Products, human capital proxied by: | Product type: |
|---------------------------------------------|----------------------------------------|---------------|
|                                             | Years & quality of schooling (OLS)     | Core (OLS)    |
|                                             | Years, quality level of schooling (OLS)| Core, (exponential model, NLS) | In-Between (OLS) | Peripheral (OLS) | Peripheral (exponential model, NLS) |
|                                             | (1)                                    | (3)           | (4)           | (5)             | (6)             | (7)             |
| Sophistication                              | 0.008                                  | -0.037        | -0.128        | 0.034           | -0.044          | -0.029          |
|                                             | 0.013                                  |               |               |                 |                 |                 |
| Sophistication x Years of Schooling         | 0.031                                  | -0.044        | -0.039        | 0.019           | 0.145           | 0.056           |
|                                             |                                        |               |               |                 |                 |                 |
| Sophistication x Cognitive Skill            | 0.116***                               | 0.122***      | 0.130*        | 0.122*          | 0.122**         | 0.040           |
|                                             |                                        |               |               |                 |                 | 0.217**         |
| Sophistication x Years of Schooling x Cognitive Skill | -0.11                              | -0.108        | 0.048         | 0.074           | -0.125          | -0.184          |
|                                             |                                        |               |               |                 |                 | 0.009           |
| Sophistication x Primary Attainment         |                                        |               |               |                 |                 | 0.045           |
| Sophistication x Secondary Attainment       |                                        |               |               |                 |                 | -0.028          |
| Sophistication x College Attainment         |                                        |               |               |                 |                 | 0.001           |

### Controls

| Controls                                   | All Products, human capital proxied by: | Product type: |
|--------------------------------------------|----------------------------------------|---------------|
|                                             | Years & quality of schooling (OLS)     | Core (OLS)    |
|                                             | Years, quality level of schooling (OLS)| Core, (exponential model, NLS) | In-Between (OLS) | Peripheral (OLS) | Peripheral (exponential model, NLS) |
|                                             | (1)                                    | (3)           | (4)           | (5)             | (6)             | (7)             |
| Lagged RCA (1995)                          | 0.440***                               | 0.329**       | 0.014***      | 0.469***        | 0.520**         | 0.014***        |
|                                             | 0.440***                               |               |               |                 |                 |                 |
| Years of Schooling                         |                                        | -0.391*       |               |                 |                 | -0.133          |
| Cognitive Skill                            |                                        |               | 0.238         |                 |                 | -0.262*         |
| Years of Schooling x Cognitive Skill       |                                        |               | 0.394**       |                 |                 | 0.026           |

| Sample size                                | 49,600                                 | 9,200         | 31,120        | 9,280           | 9,280           | 9,280           |
| R-squared                                  | 0.195                                  | 0.196         | 0.170         | 0.171           | 0.273           | 0.106           |

**Note:** * p<0.05, ** p<0.01, *** p<0.001. All standard errors are robust. Sample of 40 developing countries is used in all cases. Product sophistication is measured as the average level of countries’ cognitive skill weighted by their RCAs in that product. All education variables have been scaled to have a mean of zero and a standard deviation of one in the largest sample for which each is available. OLS models include country fixed effects. NLS models do not.
### Table 4: Slopes with respect to key variables / Hypothesis Tests

#### Preferred Models

| Derivatives with respect to: | Test of: | Assumed level of: | Normalized Cognitive Skill | Normalized years of schooling | All Products, OLS | Core Products, OLS | In-between Products, OLS | Peripheral Products, NLS |
|-----------------------------|----------|-------------------|-----------------------------|-----------------------------|-----------------|-------------------|--------------------------|--------------------------|
| First derivatives with respect to: | | | | | (1) | (2) | (3) | (4) |
| Density | (H1 against HA1) | | | | | | | |
| -1 | -1 | 1.555 *** | 1.496 *** | 1.400 *** | 0.769 *** |
| -1 | 0 | 1.614 *** | 0.714 *** | 1.616 *** | 0.635 ** |
| -1 | 1 | 1.673 *** | 0.340 | 1.832 ** | 0.523 |
| 0 | -1 | 0.817 *** | 0.610 *** | 0.709 *** | 0.561 *** |
| 0 | 0 | 0.911 *** | 0.467 *** | 0.874 *** | 0.468 *** |
| 0 | 1 | 1.006 *** | 0.353 ** | 1.039 ** | 0.390 ** |
| 1 | -1 | 0.078 | 0.035 | 0.017 | 0.409 *** |
| 1 | 0 | 0.209 | 0.112 | 0.131 | 0.345 *** |
| 1 | 1 | 0.339 * | 0.192 ** | 0.245 | 0.290 *** |

| Sophistication | (H3 against HA3) | | | | | | | |
| -1 | -1 | -0.248 *** | -0.248 | -0.074 ** | -0.392 *** |
| -1 | 0 | -0.108 ** | -0.207 ** | -0.166 | -0.279 *** |
| -1 | 0 | 0.033 | -0.1374 *** | -0.259 | -0.191 |
| 0 | -1 | -0.022 | -0.1271 | 0.007 | -0.084 |
| 0 | 0 | 0.008 | -0.1246 | -0.037 | -0.025 |
| 0 | 1 | 0.039 | -0.1109 * | -0.081 | 0.020 |
| 1 | -1 | 0.204 | -0.0465 | 0.089 * | 0.091 |
| 1 | 0 | 0.124 ** | -0.0072 | 0.093 | 0.123 |
| 1 | 1 | 0.045 | 0.03361 * | 0.097 | 0.147 ** |

#### Cross Partial Derivatives with respect to Density and:

| Cognitive Skill | (H2 against HA2) | | | | | | | |
| 0 | -1 | -0.738 *** | -0.432 ** | -1.248 *** | -0.160 |
| 0 | 1 | -0.667 *** | -0.204 | -0.112 ** | -0.098 |
| -1 | 0 | 0.059 | -0.559 *** | -0.758 | -0.109 |
| 1 | 0 | 0.130 | -0.467 ** | 0.379 | -0.047 |

#### Cross Partial Derivatives with respect to Sophistication and:

| Cognitive Skill | (H4 against HA4) | | | | | | | |
| 0 | -1 | 0.226 ** | 0.040 | 0.081 ** | 0.229 * |
| 0 | 1 | 0.006 | 0.118 ** | 0.178 | 0.164 ** |
| -1 | 0 | 0.141 | 0.106 * | -0.092 | 0.280 * |
| 1 | 0 | -0.079 | 0.176 | 0.004 | 0.115 ** |

Note: * p<0.1, ** p<0.05, *** p<0.01. Derivatives are calculated post-estimation using specifications employing the set of explanatory variables appearing in Table 3, column (1) . Derivatives under the NLS specification are calculated at central values of initial RCA, density and sophistication in the core and peripheral subsamples respectively. Derivatives are calculated using the coefficients from the following regressions in Table 3: All Products (regression 1), Core Products (regression 4), In-between Products (regression 5), Peripheral Products, Regression 7. 
### Table 5: Nascent, Emerging and Mature Industries: Tobit Coefficients.

|                                | Nascent Industries | Emerging Industries | Mature Industries |
|--------------------------------|--------------------|---------------------|-------------------|
|                                | (1) | (2) | (3) | (4) | (5) | (6) |
| **Density Considerations**     |     |     |     |     |     |     |
| Density                        | 0.538* | 0.319*** | 0.340*** | 0.269** | 2.118** | 0.899 |
| Density x Years of Schooling   | 0.224* | 0.18 | 0.18 | 1.278 |
| Density x Cognitive Skill      | -0.366*** | -0.477*** | -0.289*** | -0.369*** | -2.319*** | -2.443*** |
| Density x Years of Schooling x Cognitive Skill | -0.267* | -0.138* | -0.376** | -0.377*** | -0.14 | 0.808 |
| Density x Primary Attainment   | 0.887 | 0.299* | 5.000** |
| Density x Secondary Attainment | -0.372 | 0.19 | -2.099 |
| Density x College Attainment   | -0.328 | -0.352*** | -2.598** |
| **Sophistication Considerations** |     |     |     |     |     |     |
| Sophistication                 | -0.029 | -0.022 | -0.002 | -0.027 | 0.045 | -0.014 |
| Sophistication x Years of Schooling | 0.032 | -0.016 | 0.052 |
| Sophistication x Cognitive Skill | 0.022 | 0.021 | 0.011 | 0.007 | 0.572** | 0.580** |
| Sophistication x Years of Schooling x Cognitive Skill | 0.012 | -0.001 | 0.004 | 0.027 | -0.515 | -0.438 |
| Sophistication x Primary Attainment | -0.007 | 0.079 | 0.356 |
| Sophistication x Secondary Attainment | 0.066 | -0.083 | -0.338 |
| Sophistication x College Attainment | -0.033 | -0.009 | 0.021 |
| **Controls**                   |     |     |     |     |     |     |
| Lagged RCA (1995)              | 1.561*** | 1.571*** | 0.494*** | 0.483*** | 0.419*** | 0.418*** |
| Sample size                    | 27,144 | 27,144 | 6,559 | 6,559 | 6,552 | 6,552 |
| McFadden's Pseudo R-squared    | 0.001 | 0.001 | 0.007 | 0.008 | 0.043 | 0.044 |

Note: * p<0.05, ** p<0.01, *** p<0.001. Tobit coefficients with robust standard errors and country fixed effects. Largest available samples are used in all cases. Product sophistication is measured as the average level of cognitive skill amongst countries that export the product with RCA>1. All education variables have been scaled to have a mean of zero and a standard deviation of one. Nascent industries are country-product pairs with a 1995 RCA that is strictly greater than zero and strictly less than 0.5. Emerging and Mature industries had RCAs in the ranges [0.5,1.5) and [1.5,∞) respectively.
Table 6: Correcting for Institutional Conditions

| Measure of Institutional Quality | Roads/km² | Regulatory Quality | Rule of Law | Days to open a business | Procedures to open a business | Physical Infrastructure | ICT Infrastructure | Sanctity of Contract | Freedom from Corruption |
|---------------------------------|-----------|--------------------|-------------|------------------------|-----------------------------|------------------------|-------------------|-------------------|-----------------------|
| **Density Considerations**      |           |                    |             |                        |                             |                        |                   |                   |                       |
| Density                          | 0.714*    | 1.028***           | 0.936***    | 0.914***               | 0.974***                   | 0.687*                 | 1.063***          | 1.107***          | 1.109***              |
| Density x Years of Schooling     | 0.108     | 0.231              | 0.133       | 0.097                  | 0.033                      | -0.007                 | 0.202             | 0.153             | 0.076                 |
| Density x Cognitive Skill        | -0.776*** | -0.638***          | -0.622***   | -0.711***              | -0.811***                  | -0.647***              | -0.674***         | -0.711***         | -0.718***             |
| Density x Years of Schooling x Cognitive Skill | 0.07     | 0.082              | 0.113       | 0.032                  | -0.027                     | 0.054                  | -0.119            | -0.125            | -0.077                |
| Density x Index of Institutional Quality | 0.001   | -0.419             | -0.349      | -0.026                 | -0.402***                  | 0.167                  | -0.092            | 0.077             | 0.127                 |
| **Sophistication Considerations**|           |                    |             |                        |                             |                        |                   |                   |                       |
| Sophistication                   | 0.005     | 0.014              | 0.013       | 0.022                  | 0.004                      | 0.008                  | -0.038            | -0.038            | -0.038                |
| Sophistication x Years of Schooling | 0.002    | 0.034              | 0.03        | 0.023                  | 0.037                      | 0.054                  | -0.006            | -0.004            | -0.004                |
| Sophistication x Cognitive skill | 0.164**   | 0.119**            | 0.121**     | 0.147***               | 0.124***                   | 0.106**                | 0.073*            | 0.075*            | 0.074*                |
| Sophistication x Years of Schooling x Cognitive Skill | -0.096 | -0.11              | -0.107      | -0.121                 | -0.108                     | -0.11                  | 0.055             | 0.052             | 0.054                 |
| Sophistication x Index of Institutional Quality | 0        | -0.018             | -0.02       | 0.098*                 | 0.035                      | 0.004                  | 0.003             | -0.005            | 0.001                 |
| **Controls**                     |           |                    |             |                        |                             |                        |                   |                   |                       |
| Lagged RCA (1995)                | -0.552*** | -0.560***          | -0.560***   | -0.560***              | -0.560***                  | -0.519***              | -0.606***         | -0.606***         | -0.606***             |
| Sample size                      | 34720     | 49600              | 49600       | 49600                  | 49600                      | 42160                  | 38440             | 38440             | 38440                 |
| R-squared                        | 0.346     | 0.252              | 0.252       | 0.252                  | 0.252                      | 0.211                  | 0.22              | 0.22              | 0.22                  |

Note: * p<0.05, ** p<0.01, *** p<0.001. OLS coefficients with robust standard errors and country fixed effects. Largest available samples are used in all cases. Product sophistication is measured as the average level of cognitive skill amongst countries that export the product with RCA>1. All education variables have been scaled to have a mean of zero and a standard deviation of one.
Table A1: Slopes with respect to key variables under alternative corner solution models

| Derivatives with respect to: | Test of: | Assumed level of: | OLS, including zeros | OLS, excluding zeros | Tobit | Exponential Specification (NLS) |
|-----------------------------|----------|-------------------|----------------------|---------------------|-------|-------------------------------|
|                             |          | Normalized Cognitive Skill | Normalized years of schooling |                     |       |                               |
|                             |          |                   |                      |                     |       |                               |
| First derivatives with respect to: |          |                   |                      |                     |       |                               |
| Density (H1 against HA1)    | -1 -1    | -0.248 *** -0.2858 *** -0.271 *** -0.196 |                     |                     |       |                               |
|                             | -1 0     | -0.108 ** -0.1123 ** -0.096 * -0.029 |                     |                     |       |                               |
|                             | -1 1     | 0.033 0.061 0.078 0.254 |                     |                     |       |                               |
|                             | 0 -1     | -0.022 -0.041 0.001 0.138 |                     |                     |       |                               |
|                             | 0 0      | 0.008 0.004 0.029 0.128 |                     |                     |       |                               |
|                             | 0 1      | 0.039 0.050 0.057 0.097 |                     |                     |       |                               |
|                             | 1 -1     | 0.204 0.203 0.272 ** 0.320 |                     |                     |       |                               |
|                             | 1 0      | 0.124 ** 0.121 * 0.154 ** 0.217 |                     |                     |       |                               |
|                             | 1 1      | 0.045 0.039 0.037 -0.007 |                     |                     |       |                               |
| Cross Partial Derivatives with respect to Density and: |          |                   |                      |                     |       |                               |
| Cognitive Skill (H2 against HA2) | 0 -1    | -0.738 *** -0.738 *** -1.050 *** -0.261 |                     |                     |       |                               |
|                             | 0 1      | -0.667 *** -0.677 *** -0.970 *** -1.094 ** |                     |                     |       |                               |
| Years of Schooling          | -1 0     | 0.059 0.069 0.354 0.375 |                     |                     |       |                               |
|                             | 1 0      | 0.130 0.130 0.435 *** -0.356 |                     |                     |       |                               |
| Cross Partial Derivatives with respect to Sophistication and: |          |                   |                      |                     |       |                               |
| Cognitive Skill (H4 against HA4) | 0 -1    | 0.226 ** 0.244 ** 0.271 *** 0.261 * |                     |                     |       |                               |
|                             | 0 1      | 0.006 -0.011 -0.021 -0.145 |                     |                     |       |                               |
| Years of Schooling          | -1 0     | 0.141 0.173 0.174 * 0.477 |                     |                     |       |                               |
|                             | 1 0      | -0.079 -0.082 -0.118 0.001 |                     |                     |       |                               |

Note: * p<0.1, ** p<0.05, *** p<0.01. Derivatives are calculated post-estimation using specifications employing the set of explanatory variables appearing in Table 3, column (1). Derivatives under the NLS specification are calculated at central values of initial RCA, density and sophistication (i.e. density and sophistication are assumed to be zero). OLS and Tobit specifications employ country fixed effects.
### Table A2: Alternative measures of product sophistication

|                     | ProdY | PCI | Ubiquity | ProdSch | ProdCog | ProdColl |
|---------------------|-------|-----|----------|---------|---------|----------|
| (1)                 | 0.897*** | 0.897*** | 0.890*** | 0.900*** | 0.911*** | 0.899*** |
| (2)                 | 0.077  | 0.077 | 0.079    | 0.083   | 0.095   | 0.083    |
| (3)                 | -0.724*** | -0.724*** | -0.727*** | -0.722*** | -0.702*** | -0.724*** |
| (4)                 | 0.092  | 0.093 | 0.096    | 0.083   | 0.036   | 0.085    |
| (5)                 | 0.015  | 0.014 | 0.032    | 0.056   | 0.008   | 0.006    |
| (6)                 | 0.012  | 0.041 | -0.016   | 0.078*  | 0.031   | 0.089**  |
| (7)                 | 0.017  | 0.007 | -0.006   | 0.001   | 0.116*** | -0.008   |
| (8)                 | -0.018 | -0.049 | 0.051    | -0.139** | -0.11   | -0.096*  |
| (9)                 | 0.441*** | 0.441*** | 0.441*** | 0.441*** | 0.440*** | 0.441*** |
| Sample size         | 49,600 | 49,600 | 49,600   | 49,600   | 49,600   | 49,600   |
| R-squared           | 0.195  | 0.195 | 0.195    | 0.195    | 0.195    | 0.195    |

**Product Space Considerations**

- Density
- Density x Years of Schooling
- Density x Cognitive Skill
- Density x Years of Schooling x Cognitive Skill

**Considerations arising from traditional human capital theory**

- Sophistication
- Sophistication x Years of Schooling
- Sophistication x Cognitive skill
- Sophistication x Years of Schooling x Cognitive Skill

**Controls**

- Lagged RCA (1995)

Note: * p<0.05, ** p<0.01, *** p<0.001. OLS coefficients with robust standard errors and country fixed effects. Largest available samples are used in all cases. All education and governance variables have been scaled to have a mean of zero and a standard deviation of one.
| Product Space Considerations | All Products, human capital proxied by: | Product type: |
|-----------------------------|--------------------------------------|--------------|
|                             | Years & quality of schooling (1) | Years, quality & level of schooling (2) | Core (3) | In-Between (4) | Peripheral (5) |
| Density                     | 0.907***                           | 0.676**      | 1.150** | 0.865*         | 0.542         |
| Density x Years of Schooling| 0.097                              | -0.202       | 0.165  | 0.094          |               |
| Density x Cognitive Skill   | -0.705***                          | -0.812***    | -0.670** | -0.748***     | -0.473*       |
| Density x Years of Schooling x Cognitive Skill | 0.044 | 0.156 | 0.564 | -0.041 | -0.188 |
| Density x Primary Attainment| 0.736*                             |              |        |                |               |
| Density x Secondary Attainment| -0.219                         |              |        |                |               |
| Density x College Attainment| -0.630***                            |              |        |                |               |
| Considerations arising from traditional human capital theory | | | | | |
| Sophistication x Years of Schooling | 0.033 | -0.052 | 0.034 | 0.117 |
| Sophistication x Cognitive skill | 0.114*** | 0.117*** | 0.136 | 0.109** | 0.062 |
| Sophistication x Years of Schooling x Cognitive Skill | -0.108 | -0.102 | 0.04 | -0.109 | -0.211 |
| Sophistication x Primary Attainment | 0.051 |              |        |                |               |
| Sophistication x Secondary Attainment | -0.029 |              |        |                |               |
| Sophistication x College Attainment | 0 |              |        |                |               |
| Controls                    |                                    |              |        |                |               |
| Lagged RCA (1995)           | -0.560***                          | -0.560***    | -0.671*** | -0.531***     | -0.480**      |
| Sample size                 | 49,600                             | 49,600       | 9,200  | 31,120         | 9,280         |
| R-squared                   | 0.254                              | 0.254        | 0.412  | 0.295          | 0.087         |

Note: * p<0.05, ** p<0.01, *** p<0.001. OLS coefficients with robust standard errors and country fixed effects. Largest available samples are used in all cases. Product sophistication is measured as the average level of countries' cognitive skill weighted by their RCAs in that product. All education variables have been scaled to have a mean of zero and a standard deviation of one.