

WPS 3872

Diversification, Innovation, and Imitation inside the Global Technological Frontier*

Bailey Klinger
Daniel Lederman

Abstract: Recent research highlights the relationship between economic development and productive diversification, which may be hindered by market failures. After identifying stages of diversification in disaggregated export data, the authors develop a metric for the flows of export “discoveries,” or inside-the-frontier innovations in developing countries. The authors then explore the empirical relationship between economic development and (1) inside-the-frontier-innovation as reflected by the introduction of new export products, (2) export diversification measured by an index of export-revenue concentration, and (3) on-the-frontier innovation as reflected in patents. The data suggest that inside-the-frontier innovation is more common among poor countries for whom export growth entails diversification, whereas on-the-frontier innovation and specialization characterize the export growth of highly developed economies. The data also suggest that the relationship between the frequency of export discoveries and economic development is not due to changes in the industrial composition of exports. The authors utilize a simple model of innovation and imitation to test the hypothesis that the threat of imitation inhibits the discovery of new exports. Econometric evidence suggests that the frequency of export discoveries across countries rises with the returns of export activities (proxied by exogenous export growth during the sample period), but the magnitude of this effect increases with barriers to entry. The count-data estimations deal with unobserved international heterogeneity, and the results are robust to various changes in the specification of the empirical model. This finding supports the hypothesis that market failures inhibit inside-the-frontier innovation.

World Bank Policy Research Working Paper 3872, April 2006

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

*This paper is an extension of previous work (Klinger and Lederman 2004). We are grateful for the valuable feedback provided by Caroline Freund, Andrés Rodríguez-Clare, Guillermo Perry, William F. Maloney, Ricardo Hausmann, Dani Rodrik, Ana-María Oviedo, and Jim Lahey. The authors are responsible for all remaining errors. This research was funded by the Office of the Chief Economist for Latin America and the Caribbean. Klinger is a Ph.D. candidate at Harvard University. Lederman is a Senior Economist with the Development Economics Research Group (DECRG) of the World Bank. All comments are welcome and should be sent to: dlederman@worldbank.org & bailey_klinger@ksgphd.harvard.edu

1. Introduction

Recent research by Imbs and Wacziarg (2003) showed that economic development is associated with increasing diversification rather than specialization in lower- and middle-income countries. This paper studies one aspect of economic diversification, namely the introduction of new export products during the process of development. In this context, diversification is related to the market failure hypothesis of Hausmann and Rodrik (2003a), which suggests that discovering that a product can be profitably produced in a particular country, where it has never been produced before, creates valuable social knowledge. Production costs and foreign demand are not knowable *ex ante*, making such a discovery an ‘inside-the-frontier innovation’. Firms that imitate the first mover derive benefits from such discoveries, yet they pay nothing for it. That is, imitation raises the social returns relative to the private returns. While imitation is clearly desirable from a social viewpoint, the first mover can not appropriate all the value created by their investments in discovery as, as this type of innovation can not be protected by the intellectual property rights regimes used to spur on-the-frontier innovation. As entrepreneurs can not appropriate all of the value they create, they underinvest in the experimentation necessary to discover new export opportunities, and consequently the process of productive diversification and private sector development stagnates.

However interesting this theory is, the market failure hypothesis has not been tested empirically. In fact, we know very little about the relationship between the emergence of new export products and economic development in general. This paper attempts to fill this void.

We develop an identification strategy for the existence of market failures affecting inside-the-frontier innovation, or discovery, using disaggregated product-level export data available for a large sample of countries. In addition, we develop a metric for discovery activity and establish empirical regularities that serve to deepen our understanding of the discovery process and its relationship to productive diversification and economic development.

The rest of the paper is organized as follows. Section 2 proposes a minimalist model of discovery and imitation by focusing on potential interactions between the returns to introducing a new product for export and the ease of imitation. Section 3 discusses the data and the methodology for identifying episodes of export discoveries. Section 4 presents various results by proceeding in stages. We first link export diversification with the process of development in the same spirit as Imbs and Wacziarg’s (2003) analysis of the relationship between employment and production diversification and development. In turn, this section contrasts empirical regularities concerning the inside-the-frontier innovation as proxied by our measure of export discoveries, export diversification, and patenting activity and their relationship with the level of development proxied by Gross Domestic Product (GDP) per capita. Likewise, we explore the view that the observed relationship between export discoveries and development is due to structural changes in factor endowments associated with development. The final set of results

concerns the market-failure hypothesis. Section 5 concludes by summarizing the main findings and discussing the benefits of imitation and low barriers to entry, as well as other policy implications.

2. A Model of Discovery and Imitation

Our simple model considers the effect that changes in returns to discovery have on the frequency of discovery, depending on barriers to entry and how they affect the appropriability of those potential profits. The frequency (D) of observed discoveries in a country can be characterized as a function of the returns from discovering a new export, π , and the first mover's ability to appropriate those returns, determined in part by barriers to entry α . On the one hand, if there are costly barriers to entry, then there will be fewer imitators to capture the value spilling over from the first mover's demonstration of the viability of the new export. On the other hand, if barriers to entry are light, then imitators will be able to copy the pioneer's discovery, giving first movers less of an incentive to experiment thus lowering the frequency of discovery.

In addition to the effect of barriers on the first mover's ability to appropriate returns from its investments in export experiments, barriers to entry also affect the first mover's profitability directly. If there are high costs to establishing a new venture and contracting employees, then the returns to the first mover are lower, no matter what proportion of those benefits the first mover will be able to appropriate. For this reason, returns to discovery are also directly a function of barriers to entry, α , with higher barriers implying lower returns. We model other determinants of returns to discovery not driven by barriers to entry as $\bar{\pi}$.

This minimalist model can be written formally as:

$$(1) \quad D = \pi \left(\begin{matrix} \bar{\pi} \\ + \\ \alpha \end{matrix} \right) \times \begin{matrix} \alpha \\ + \end{matrix}.$$

An increase in exogenous returns has an unambiguously positive relationship with discovery:

$$(2) \quad \frac{\partial D}{\partial \bar{\pi}} = \alpha \times \begin{matrix} \frac{\partial \pi(\bar{\pi}, \alpha)}{\partial \bar{\pi}} \\ + \end{matrix} > 0.$$

In contrast, an increase in barriers to entry, α , has an ambiguous relationship with discovery, as it implies both lower returns to the first mover, but also a greater ability to appropriate those returns through lower imitation:

$$(3) \quad \frac{\partial D}{\partial \alpha} = \alpha \times \begin{matrix} \frac{\partial \pi}{\partial \alpha} \\ + \\ \pi(\bar{\pi}, \alpha) \\ - \end{matrix}.$$

If market failures are not associated with imitation, then the second α in equation (1) would not have an effect on D , and barriers would have an unambiguously negative correlation with discovery. Also, the absence of market failures associated with imitation would imply that the effect of an increase in returns to discovery is not affected by barriers to entry.

If imitators can benefit from spillovers emanating from the original discovery, we would observe an ambiguous relationship between discovery and barriers to entry. Furthermore, we would expect that the relationship between returns to discovery and the frequency of discovery would be affected by variations in barriers to entry. The effect of higher returns on export discoveries would be *greater* in countries with *higher* barriers to entry, as first movers would be able to appropriate a larger share of the export returns. That is, in this scenario pioneers would have a greater incentive to experiment for a given level of expected returns. Conversely, if barriers to entry are low, then an increase in exogenous returns to exporting will have a smaller effect on discovery, as much of those returns will leak to imitators. This, then, is our test of the market failure hypothesis: how does the relationship between exogenous returns to discovery and the frequency of observed discovery vary with barriers to entry.

3. Data and Methodology

To test this hypothesis, we require a metric of inside-the-frontier innovation (when a good enters a country's productive basket for the first time), which has not yet been developed. Production data highly disaggregated at the product level would be the ideal source of such a metric, but are unavailable for most countries. As such, we exploit export data, which for reasons of customs reporting are available for a broad cross-section of countries at a high level of disaggregation by product.

The problem with using export data seems obvious: a product emerging as a new export may have been produced domestically for some time, and therefore would not represent a true inside-the-frontier innovation. However, exporting a particular good for the first time, even if it was already produced domestically, is itself an entrepreneurial act that requires discovery and can be imitated (Ibeh 2003). In other words, the act of exporting is itself a discovery worth investigating. Furthermore, the increasing importance of trade and foreign markets for developing countries, most of which are pursuing an export-led development strategy, means that discovering new products is primarily a search for new exports.

Worldwide export data are drawn from the United Nations COMTRADE database under the Harmonized Commodity Description and Coding System (HS), available beginning in the early 1990s for most countries at the 6-digit level of disaggregation (approximately 5000 products). These data have been used widely to analyze export dynamism and growth, as well as geographic patterns in export growth, but until now have not been used to study the emergence of new exports.

To identify a discovery, we require three periods: an initial period used to confirm that the good was never before exported (1994-1996), a window during which time the discovery can emerge (1997-2002) and then a final period when the discovery is confirmed to be an established export (exported for at least \$10,000 dollars in both 2002 and 2003). Given the short timeframe of available data, only a cross-sectional measure of discovery flows is possible. After eliminating microstates and countries missing more than two consecutive years of export data during the window, we are left with a sample of 73 countries representing all regions and levels of development (see Appendix for sample composition).

Within this data, there is a problem of reclassification. This occurs when, in a particular year, a country's customs body begins to report goods separately that were previously aggregated in a 'not elsewhere specified' (n.e.s.) group. This is particularly problematic from the point of view of identifying episodes of discovery, as the filter would falsely identify each of the newly disaggregated products as discoveries, even though in reality they are not new to the country's export basket. In order to systematically identify cases of customs disaggregation, we apply a filter based on the divergence from each country's temporal trend during the window. We identify 19 significant outliers that are likely cases of reclassification¹, and eliminate these country/years from our sample, leaving a more reliable measure of discovery activity by country.

4. Results

Our filter² identifies 3089 cases of discovery during 1997-2002. These are listed by country and by Leamer's (1984) commodity groups in the Appendix. Before testing the model of innovation and imitation motivated in Section 2, we discuss the necessary control variables for our test of the market failure hypothesis.

Discovery, Diversification, and Distance to the Frontier

We expect episodes of discovery to depend on the level of development for two related reasons. First, the nature of innovation changes as countries develop. As defined by Nordhaus (1969), innovation encompasses not only products that are completely new to the world, but also existing products that are new to a particular country or firm. Discovery represents an inside-the-frontier innovation: producing goods that have already been invented elsewhere. As countries become richer, they approach the global technological frontier, and the nature of innovation changes in tandem.³ Innovation on the global frontier involves the creation of products and product varieties new to the

¹ See Appendix for a description of the filter and the identified cases of reclassification.

² This metric of discovery is a revision of that developed in Klinger and Lederman (2004). It uses a more explicit definition of discovery (exports starting from 0 rather than from below a threshold), extends the length of the first period for a more stringent test of novelty of the export, and uses the reclassification filter discussed in the text.

³ Acemoglu, Aghion, and Zilibotti (2002) model the changing nature of innovation as countries approach the frontier in order to analyze firm strategies. Aghion & Howitt (2005) develop a model of growth based on distance to the frontier and its consequences for appropriate growth policies.

world. This changing nature of innovation implies that discovery activity would be more frequent at lower levels of development when countries are farther from the global technological frontier, but decline and eventually be overtaken by on-the-frontier innovation as incomes rise.

The second reason why expected levels of discovery activity would depend on the level of income is the recently-uncovered pattern of the stages of productive diversification across levels of development. Imbs and Wacziarg (2003) analyze how productive diversification behaves across income levels, and found that although there are theoretical arguments supporting both a positive and a negative monotonic relationship between diversification and growth, the evidence shows that neither view is correct. There is, in fact, a robust pattern whereby as countries develop, production is diversified until reaching a relatively high level of GDP per capita, between \$13,000 and \$14,600 1996 US dollars, after which point economies become increasingly specialized.

Although Imbs and Wacziarg (2003) used domestic production and labor data in their analysis of the stages of diversification, the same result can be found in export data. We construct a Herfindahl index (H) of exports for each country in every year with reported exports between 1992 and 2003, and estimate the following equation using a fixed-effects estimator:⁴

$$(4) \quad H = \beta_0 + \beta_1(GDPpercapita) + \beta_2(GDPpercapita)^2$$

The results, summarized in Table 1, indicate that, similar to the pattern in domestic production data, a country's export basket becomes more diversified as income rises until a relatively high level, at which point the process reverses itself and specialization occurs. This happens in the export basket at a higher transition point than that found by Imbs and Wacziarg (2003) in domestic production data (\$22,500 in 2000 US dollars, PPP adjusted, compared to \$14,600 in 1996 US dollars), but these results support the view that the pattern of economic diversification is related to patterns of international trade.

<Insert Table 1 here>

This robust pattern in both domestic production and export data suggests a particular relationship between discovery and levels of development, given the relationship between diversification and new products.⁵ We expect countries at relatively

⁴ H was calculated as $H = \sum_{i=1}^J \left(\frac{x_i}{\sum_{i=1}^J x_i} \right)^2$ where each i is an individual product and J is the total number of

products.

⁵ The connections between Imbs and Wacziarg (2003) and the process of discovery were first suggested by Hausmann and Rodrik (2003b).

low levels of development to have more frequent episodes of discovery, as they are in the process of diversifying their economies. As income rises, the frequency of these events declines, particularly at high levels of development when economies experience rising specialization.

To examine these two effects in our data, we estimate the relationships between both inside-the-frontier innovation (discoveries as identified by our filter) and on-the-frontier innovation (patent counts from Lederman and Saenz 2005) and the level of development. Table 2 shows the results from Negative Binomial estimators, which are appropriate for count data. As expected, the frequency of discovery falls as countries develop, after peaking at the lower-middle income level. Although low among the world's poorest countries, the frequency of discovery rises quickly, reaching a maximum somewhere in the neighborhood of GDP per capita of \$4000 USD as countries undergo productive diversification. As the level of development continues to rise, inside-the-frontier innovation is replaced by on-the-frontier innovation, which increases exponentially with GDP per capita.

<Insert Table 2 here>

These relationships are illustrated in Figure 1, which shows the estimated relationships from Tables 1 and 2⁶. Discovery activity peaks early in the development process, and declines as the process of diversification slows. At higher levels of development, economies begin to specialize, and innovation is driven by on-the-frontier advances that are patented. This relationship between distance to the frontier, the nature of innovation, and the stages of productive diversification reveals that we must control for the level of development (in quadratic form) when testing the model of innovation and imitation.

<Insert Figure 1 Here>

Other Controls

Discoveries may be concentrated in certain industries and thus episodes of discovery could be driven by changes in factor endowments associated with the process of development, such as increases in the capital per worker. For example, discoveries in poor countries could be concentrated in labor-intensive goods, and as growth occurs and factor endowments change, discoveries would be concentrated in capital intensive goods. In this context, discovery would be a byproduct of structural transformation.

The structural transformation hypothesis can be tested. Table 3 shows little evidence to support this view. Discovery activity in some commodity groups, such as labor intensive goods, does peak at a slightly lower income level than in others, such as chemicals. But nearly all commodities reach their maximum discovery frequency in the

⁶ Each is scaled by its maximum value for illustrative purposes.

\$3600 to \$4900 per capita range and then decline, with no commodity group peaking beyond \$7000 per capita. This evidence suggests that although discovery might be part and parcel of the process of productive diversification, it does not seem to be a byproduct of shifting comparative advantage.

<Insert Table 3 here>

One relationship that would seem important to control for in our test of the market failure hypothesis is scale. While not explicitly entering our model, it is possible that with a larger population, there would be a larger pool of entrepreneurs, leading to more experiments and a higher frequency of discovery. It is also possible that a larger pool of imitators, although increasing the social value of discovery, would reduce incentives for individual entrepreneurs to experiment, resulting in a lower frequency of discovery. Scale variables such as total population, working-age population, and number of person-years of education in the country were not found to be statistically significant in any test of our model, which is discussed in the following paragraphs.⁷

Testing the Model

As motivated in section 2, our model serves as a test of the market failure hypothesis by examining how the relationship between the frequency of discovery, D , and the returns to discovery, $\bar{\pi}$, changes with barriers to entry, α . As a measure of barriers to entry, we use a set of objective indicators rather than subjective surveys. These indicators are drawn from the World Bank's *Doing Business* database (World Bank 2004). Although only available for 2003, these are the best measures available of α , and it is unlikely that the regulatory regime changed significantly during the sample period. Rather than arbitrarily selecting a single indicator, we construct an index based on five of them, encompassing costs and delays in starting a new business, enforcing contracts, and hiring employees. The indicators for registering a new business would be most appropriate if discovery is undertaken by new entrepreneurs, whereas the measures of labor-market rigidities would be most appropriate if discovery is the work of existing firms. It is not clear which is the case, which could also vary by country and by industry. Moreover, all these regulatory measures are highly collinear. A composite index of these variables is consequently the most appropriate measure, which we construct using principal components analysis. Nevertheless, the quality of the data corresponding to the variable on the costs of starting a new business across countries is notoriously weak. Consequently we test the robustness of our results reported below to the exclusion of that factor in the composite index.⁸

Our measure of exogenous returns to export discoveries ($\bar{\pi}$) is the annual growth of exports of the Leamer commodity export basket, excluding export goods identified as discoveries, minus the annual growth rate of non-export GDP. The growth of non-export GDP is netted out of our measure so that it more directly captures the returns to exports

⁷ Nor were measures of financial system development, initial exports, institutional quality, or infrastructure.

⁸ We thank Caroline Freund for pointing this out.

as compared to other economic activities such as producing for domestic consumption or producing services, which are not captured in the merchandise trade data used to identify export discoveries. More formally, the exogenous returns to export in country c and commodity group i , $\bar{\pi}_{i,c}$ was calculated as:

$$(5) \quad \bar{\pi}_{i,c} = \left(\frac{X_{i,c,T}}{X_{i,c,T_0}} \right)^{\frac{1}{T-T_0}} - \left[\left(\frac{(y_{c,T} - x_{i,c,T})}{(y_{c,T_0} - x_{i,c,T_0})} \right)^{\frac{1}{T-T_0}} \right],$$

where subscript T_0 is the initial year in the time window in which the discoveries were counted and T is the final year of the period. X represents the real value of exports excluding the contribution of the export discoveries at time T , and lower-case x and y are merchandise exports and GDP per capita. This measure of export returns was calculated separately for each of the ten Leamer commodity categories. Furthermore, as discussed above, it is possible that barriers will have a differential impact across industries. Therefore, our analysis conducted at the country/Leamer commodity group level examines the relationship between the number of discoveries in a particular country/commodity group and export growth in that country/commodity group, as well as how that relationship changes with barriers to entry across countries. In the sample used in the regression analysis only Malaysia had exports that exceeded the value of GDP at the end of the period in 2003, which yielded a very high positive number for this proxy of export profitability, as shown in the table with descriptive statistics in the Appendix. The regression results discussed below, however, were qualitatively similar when this country's observations were excluded from the sample, but the results regarding the presence of market failures were stronger than those reported below.

In addition to these variables of interest, there are some necessary controls to be included in our test of the model. As discussed above, the process of discovery is closely linked to stages of productive diversification and the changing nature of innovation as countries approach the global technological frontier. We therefore control for this relationship by including GDP per capita and GDP per capita squared. We also allow for the possibility that discovery is driven in part by factor endowments by controlling for endowments as measured by historical net exports per capita in that particular commodity group (Leamer 1984). In addition, to control for differences across commodity groups and the different number of products composing each commodity cluster, we include dummy variables for each Leamer commodity group.

Finally, we control for historical discoveries by country/commodity group. This is accomplished using export data at a higher level of aggregation⁹, which is available for a longer time period but only identifies broad, sectoral-level discoveries rather than disaggregated product-level discoveries. This measure of historical pre-sample discoveries controls for time-invariant and unobserved characteristics driving discoveries

⁹ SITC Revision 1 at the 3 digit level. See Appendix for data and filter description.

at the country/commodity group level, as in other count-data applications (Blundell et al. 2002).

As mentioned, because our dependent variable is the number of discoveries (a positive integer) with a substantial number of zeros, we estimate the following model with an exponential functional form using a negative binomial estimator:¹⁰

$$(6) \quad D_{i,c} = \exp\left(\bar{\pi}_{i,c}^{\beta_1} (\bar{\pi}_{i,c} \times \alpha_c)^{\beta_2} \alpha_c^{\beta_3} X_{i,c}^\gamma + \lambda D_i\right) \eta_{i,c}.$$

Subscripts i and c correspond to Leamer's commodity groups and countries, respectively. $\eta_{i,c}$ is the aforementioned commodity-group and country fixed effect, which is unobserved and captured by the pre-sample number of export discoveries. α_c is the index of barriers to entry measured at the country level. X is the vector of the other control variables discussed above, and D_i is a dummy variable for each Leamer commodity group. The latter controls for unobserved sector-specific effects but also captures the number of product lines in each category, which affects the number of observed discoveries across commodity groups.

As is commonplace in count-data models, the explanatory variables were transformed into their natural logarithms or were included in growth rates (in case of the exogenous export growth variable), which then allows for the estimation of the relevant elasticities. The exceptions were the commodity-group dummies, the indicator of comparative advantage of each country in each of the commodity groups (proxied by net exports per capita), and the barriers index. The latter was calculated as the first principal component of the five regulatory indicators discussed above, after they were normalized to have means equal to zero and standard deviations equal to one. The sign and significance of β_2 in (6) encompass our test of the market failure hypothesis.

Table 4 shows the basic estimation results under the first column, and the remaining columns show the results of additional regressions that test their robustness. The second column shows the basic specification but using the barriers index that excludes the costs of starting a new business, which was based on questionable data. Column 3 contains estimates based on the same composite index of barriers, but also includes the interaction between barriers and (log) GDP per capita, which is a rather strong robustness test of whether the key estimated effect of the interaction between barriers and export profitability is not due to an interaction with the level of development. Column 4 shows the estimated coefficients after controlling for unobserved regional characteristics. The interpretation of the corresponding regional dummy variables needs to be done with care, since the model already controls for the pre-sample discovery counts by country/commodity groups. They reflect any additional impact emanating from time invariant regional characteristics. Finally, column 5 shows the results after adding interactive variables between the regional dummies and the barriers index to help us ascertain that the interaction that matters is the one with export profitability rather than some other regional factor.

¹⁰ We began with a Poisson estimator, but the likelihood-ratio test indicated that the data are overdispersed.

<Insert Table 4 here>

The expected inverted-U relationship between discoveries and GDP per capita persists and is highly significant across all specifications. In addition, historical discoveries enter as positive and significant, signaling that we are effectively correcting for fixed country/commodity effects leading to discovery. As suggested by the similarity in maximum points across Leamer categories shown in Table 3, factor endowments are not significant, a result that persists without controlling for historical discoveries (not reported).

The returns to discovery, measured as the growth rate of the Leamer commodity cluster less non-export GDP growth, enters as positive and jointly significant with the interaction term in all specifications, as predicted by our model. Given that we tested a variety of additional control variables that were mentioned above, these results suggest that the predictions of the market-failure model are robustly supported by the data. That is, export growth has a positive and significant effect on the frequency of export discoveries, but the magnitude of this effect seems to rise with barriers to entry. In addition, the direct effect of barriers to entry on discovery frequency is largely insignificant and not robust, which was also predicted by our model, where barriers only affect discovery either through their impact on export profitability as in equation (3) or through its interaction with export returns as in equation (2). This result does not imply, however, that barriers to entry do not have an impact on discovery through its negative effect on export growth, given their offsetting stimulus to discovery through increased appropriability and drag on discovery through higher costs for the first mover. Furthermore, the results suggest that barriers to entry raise the magnitude of the effect of export profitability on discovery counts, but this comes at the cost of reducing the social gains from imitation.

One potential weakness of our theoretical and empirical models is that they assume a linear relationship between the marginal effect of export returns on the frequency of discovery and barriers to entry. There could be a threshold below which barriers to entry have no effect on deterring discovery, but after which there is an effect.

To test for such a relationship, we ranked countries from low barriers to high barriers based on their composite-barriers index, and then apply rolling regressions with a window of 20 countries, sequentially adding the country with the next-highest barriers to entry and dropping the one with the lowest barriers. We estimated equation 6 without including barriers to entry or the interaction term and observed how the coefficient on export returns changes as the sample window moves from low to high-barrier countries.

The result is illustrated in Figure 2, which suggests a threshold effect. For low-barrier countries, the marginal effect on discovery of an increase in returns is negligible. For mid- to high-barrier countries the estimated effect becomes larger and statistically significant in spite of the low degrees of freedom due to the small window and large

number of control variables. This relationship is quite robust, persisting across different ranges of the moving window and different specifications of the model, and offers further support to the market failure hypothesis.¹¹

<Insert Figure 2 Here>

5. Conclusions and Policy Implications

Recent research has highlighted the importance of productive diversification for developing countries: a process that may be hindered by spillovers in the discovery of new products for export. After extending this finding from domestic production data to export data, we use highly disaggregated export data to develop a measure of this inside-the-frontier innovation, and characterize its relationship to on-the-frontier innovation, productive diversification, and development. We then go on to use this discovery metric to test a highly stylized model of innovation and imitation.

According to the data and in the context of our model, imitation of first-movers may be leading to fewer cases of inside-the-frontier innovation than otherwise would be the case. Even after controlling for the quadratic relationship with GDP per capita, the amount of discoveries in past years, factor endowments, plus a plethora of additional explanatory variables, discovery has a larger response to increased profitability in countries with higher barriers to entry. While the Hausmann and Rodrik (2003a) attribute this imitation specifically to the learning of production costs which are not predictable *ex ante*, our test makes no determination as to the importance of this particular channel. The spillover may not be from learning production costs, but instead from learning the characteristics of foreign demand (Vettas 2002), free-riding on investments to cultivate foreign demand (Bhagwati 1968, Mayer 1984), learning the redesigns needed to meet foreign safety standards (Granslandt and Markusen 2000), or some other unidentified channel.

Regardless of the particular channel, this finding suggests that there are indeed spillovers from demonstrating the viability of a new product for export. While this indicates that public support for experimentation in new sectors and activities may be warranted, it is important to note that it does not mean that increasing barriers to entry is an advisable way to increase discovery. Indeed, imitation is desirable as this is the channel through which the returns of inside-the-frontier innovation are socialized. Moreover, such barriers are directly attributed to lower levels of private sector development (World Bank 2005), and according to Hausmann and Rodrik's (2003a) model, would lead to underspecialization of the economy, as widespread imitation leads to the efficient focusing of resources in the most profitable sectors. Indeed, imitation of profitable new export sectors increases economic efficiency, and therefore should be encouraged rather than hindered. Furthermore, supporting new production with barriers to entry protects beneficiaries from market discipline, which would be repeating the errors of import-substituting industrialization policies by not allowing the market to

¹¹ A changing relationship between returns and discovery could manifest itself through changes in the other coefficients or the constant, but estimates of all other coefficients were stable across the moving window.

eventually ‘pick the winners’. Support mechanisms that do not build up inefficient barriers insulating firms from competition, and instead balance government and market failures, represent a more productive way forward¹². Furthermore, such policies are not themselves completely new. From Lesotho’s Pioneer Industries Bill of 1967 to China’s National New Product Program of 1988, various mechanisms focused specifically on new products have already been deployed. Evaluating the results of such programs would provide an even better test of the market failure hypothesis, and represents a promising avenue for future research.

¹² For a more detailed discussion, see Rodrik (2004)

6. References

- Aghion, P. and P. Howitt. 2005. "Appropriate Growth Policy, a Unifying Framework." The 2005 Joseph Schumpeter Lecture, presented to the 20th Annual Congress of the European Economic Association. Amsterdam, August 25, 2005.
- Bhagwati, J. 1968. "The Theory and Practice of Commercial Policy: Departures from Unified Exchange Rates." *Special Papers in International Economics*. New Jersey: Princeton University Press.
- Blundell, R., R. Griffith, and F. Windmeijer. 2002. "Individual Effects and Dynamics in Count Data Models." *Journal of Econometrics* 108: 113-31.
- Ganslandt, M, and J.R. Markusen. 2000. "Standards and Related Regulations in International Trade: A Modeling Approach." In *Quantifying the Impact of Technical Barriers to Trade: Can It Be Done?*, pp. 95-135. Studies in International Economics, Ann Arbor: University of Michigan Press.
- Hausmann, R. and D. Rodrik. 2003a. "Economic Development as Self-Discovery." *Journal of Development Economics* 72: 603-633.
- Hausmann, R. and D. Rodrik. 2003b. "Discovering El Salvador's Production Potential." Mimeographed. Kennedy School of Government, Harvard University, Cambridge, Massachusetts.
- Ibeh, Kevin I.N. 2003. "Toward a Contingency Framework of Export Entrepreneurship: Conceptualizations and Empirical Evidence." *Small Business Economics*. 20: 49-68.
- Imbs, J. and Wacziarg, R. 2003. "Stages of diversification." *American Economic Review* 93(1): 63-86.
- Klinger, Bailey, and Daniel Lederman. 2004. "Discovery and Development: An Empirical Exploration of 'New' Products." World Bank Policy Research Working Paper 3450, The World Bank, Washington, DC.
- Leamer, Edward E. 1984. *Sources of Comparative Advantage: Theory and Evidence*. Cambridge MA: The MIT Press.
- Lederman, Daniel and Laura Saenz. 2005. "Innovation around the World, 1960-2000." World Bank Policy Research Working Paper 3774, The World Bank, Washington DC.
- Mayer, Wolfgang. 1984. "The Infant-Export Industry Argument." *Canadian Journal of Economics* 17(2): 249-269.
- Nordhaus, William D. 1969. *Invention, Growth and Welfare*. Cambridge, Massachusetts: The MIT Press.

Rodrik, Dani. 2004. "Industrial Policy for the Twenty-First Century." Harvard University Kennedy School of Government Faculty Research Working Paper Series RWP04-047, Cambridge, MA.

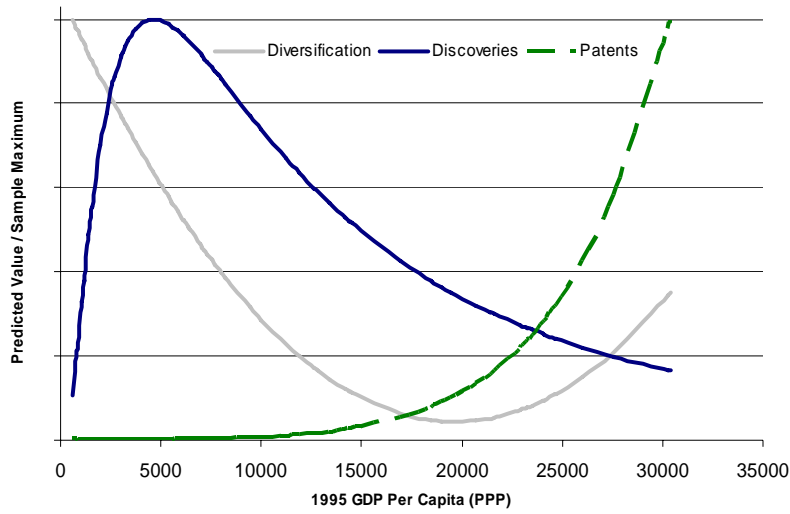
Vettas, Nikolaos. 2002. "Investment dynamics in markets with endogenous demand." *The Journal of Industrial Economics* 48(2): 189-203.

World Bank. 2004. *Doing Business in 2004: Understanding Regulation*. Washington, DC: The World Bank.

World Bank. 2005. *World Development Report 2005: A Better Investment Climate for Everyone*. Washington, DC: World Bank & Oxford University Press.

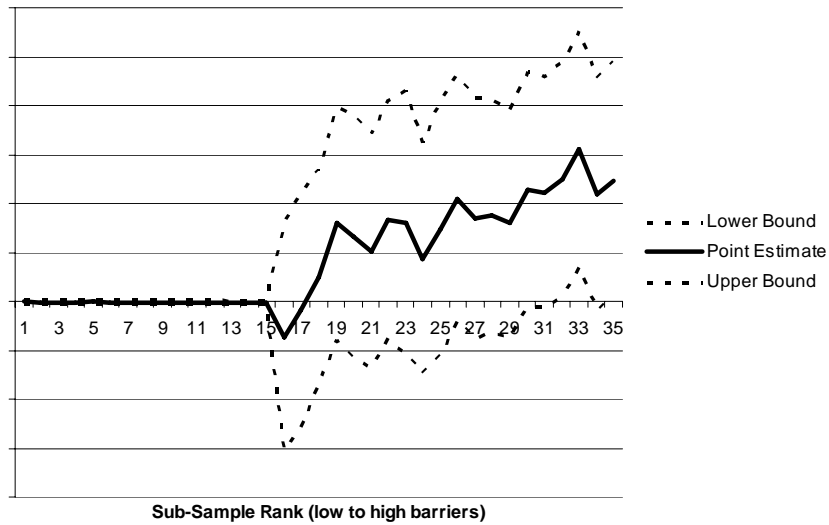
Figures & Tables

Figure 1: Diversification & Innovation



Source: Author's Calculations

Figure 2
Effects of Returns on Discovery: Non-Linear Effect of Barriers



Source: Author's Calculations. Dashed lines indicate 90% confidence interval

Table 1: Stages of Export Diversification
(Country FEs estimation; data from 1992-2003)

	Herfindahl Index
GDP per capita	-.0000117 (2.41)**
GDP per capita ²	2.60e-10 (2.46)**
Diversification/Specialization Transition Point	\$22,500
Observations	1140
Number of Countries	130

Absolute value of t statistics in parentheses

** significant at 5%

Table 2: The Changing Nature of Innovation

	Inside-the-Frontier: Discoveries	On-the-Frontier: Patents
ln(GDP per capita)	8.667 (6.53)***	-16.237 (2.96)***
ln(GDP per capita) ²	-0.514 (6.67)***	1.059 (3.43)***
Constant	-32.289 (5.71)***	65.872 (2.74)***
Observations	73	68

Absolute value of z statistics in parentheses

*** significant at 1%

Table 3: Discovery & GDP by Leamer (1984) Commodity Groups

Leamer Commodity Group	Number of Discoveries ^{N1}	Discovery Curve Maximum Point (GDP per capita)
Petroleum	1.17	2052**
Raw Materials	0.85	4901***
Forest Products	0.73	4416***
Tropical Agriculture	0.49	4486***
Animal Products	0.73	4109***
Cereals, etc.	0.72	4055***
Labor Intensive	0.38	3626***
Capital Intensive	0.56	4546***
Machinery	0.61	4578***
Chemical	0.78	6838***

N1: normalized by the number of lines in the HS 1989/1992 nomenclature composing that category. ** significant at 5%; *** at 1%. Source: Author's Calculations.

**Table 4: Negative-Binomial Estimation Results:
Determinants of Export Discovery Counts during 1994-2003**

	(1)	(2)	(3)	(4)	(5)
Ln(GDP per capita)	8.436	8.245	8.310	6.657	13.927
	(9.09)***	(8.83)***	(8.14)***	(3.76)***	(3.88)***
Ln(GDP per capita) ²	-0.507	-0.495	-0.498	-0.404	-0.812
	(9.24)***	(8.91)***	(8.32)***	(3.90)***	(4.04)***
Ln(historical discoveries)	0.305	0.305	0.304	0.249	0.257
	(3.39)***	(3.39)***	(3.37)***	(2.87)***	(2.88)***
Factor Endowments	0.349	0.354	0.355	0.397	0.383
	(0.98)	(0.99)	(1.00)	(1.08)	(0.98)
π	0.368	0.501	0.493	0.434	0.523
	(1.71)*	(2.08)**	(2.04)**	(1.75)*	(2.06)***
$\pi \times barriers$	0.674	0.748	0.736	0.652	0.774
	(1.77)*	(2.13)**	(2.09)**	(1.80)*	(2.09)**
<i>barriers</i>	-0.003	0.026	0.237	0.583	4.733
	(0.04)	(0.25)	(0.28)	(0.56)	(1.82)*
D_Raw Materials	2.357	2.380	2.382	2.382	2.366
	(5.40)***	(5.41)***	(5.43)***	(5.06)***	(5.00)***
D_Forest Products	1.922	1.943	1.943	1.948	1.949
	(4.41)***	(4.43)***	(4.44)***	(4.15)***	(4.14)***
D_Tropical Agriculture	1.728	1.755	1.754	1.704	1.717
	(3.85)***	(3.88)***	(3.88)***	(3.57)***	(3.57)***
D_Animal Products	2.291	2.314	2.314	2.309	2.317
	(5.28)***	(5.29)***	(5.30)***	(4.96)***	(4.92)***
D_Cereals	1.934	1.953	1.953	1.969	1.983
	(4.52)***	(4.53)***	(4.53)***	(4.28)***	(4.27)***
D_Labor Intensive	2.709	2.719	2.721	2.708	2.678
	(6.19)***	(6.18)***	(6.21)***	(5.88)***	(5.80)***
D_Capital Intensive	3.260	3.284	3.284	3.278	3.268
	(7.49)***	(7.48)***	(7.50)***	(7.05)***	(7.00)***
D_Machinery	3.024	3.037	3.038	3.048	3.039
	(6.94)***	(6.93)***	(6.96)***	(6.56)***	(6.50)***
D_Chemicals	3.669	3.686	3.685	3.657	3.654
	(8.25)***	(8.23)***	(8.24)***	(7.72)***	(7.66)***
$\pi \times \ln(GDPpc)$			-0.022	-0.050	-0.476
			(0.24)	(0.45)	(1.72)*
Africa				-1.102	1.838
				(3.13)***	(1.41)
East Asia				-0.337	-0.259
				(1.22)	(0.40)
Eastern Europe				-0.114	-0.607
				(0.33)	(1.01)
High-Income				-0.584	-0.133

				(1.75)	(0.20)
Latin Am.				-0.811	-0.447
				(3.04)***	(0.70)
Middle East				-0.338	-0.064
				(1.08)	(0.11)
<i>Barriers*Latin America</i>					-0.674
					(1.09)
<i>Barriers_Africa</i>					-3.201
					(2.34)**
<i>Barriers_High Inc.</i>					-0.078
					(0.16)
<i>Barriers*East Asia</i>					-0.118
					(0.20)
<i>Barriers*Middle East</i>					-0.794
					(1.43)
<i>Barriers*Eastern Europe</i>					1.655
					(2.79)***
Constant	-35.753	-35.043	-35.369	-27.627	-60.129
	(9.24)***	(9.08)***	(8.23)***	(3.71)***	(3.81)***
Observations	510	510	510	510	510

Robust z statistics in parentheses

* significant at 5%; ** significant at 1%

Note: See text and Appendix for definitions of variables. The excluded regional dummy variable in model (5) corresponds to South Asia, which in this sample is India. The excluded Leamer commodity group in all specifications is mining and petroleum.

Appendix

Composition of Sample

We take all countries that report exports in COMTRADE during at least two years during the 1994-1996 pre-window period, as well as in the 2002-2003 post-window period, eliminate those missing at least two consecutive years of data during the 1997-2002 window (necessary for purposes of identifying reclassifications), and dropping microstates (countries with a 1995 population less than 500,000 according to the World Bank WDI), we the cross-section of 73 countries in our sample listed in Table A.III below.

Identifying Cases of Reclassification

As discussed in the text, the discovery filter would misidentify cases of commodity reclassification (when a country begins to report a particular group of goods separately that were previously aggregated in a 'not elsewhere specified' product line) as discoveries. While it is not possible to identify each case of reclassification directly, we have developed a systematic and objective filter by recording discoveries by year between 1997 and 2002 (based on the year the export good first emerges) and identifying extreme outliers.

We would not necessarily expect the same number of discoveries in each year from 1997 to 2002, because with each passing year the test for novelty becomes more stringent and the incubation period between when the export first emerges and the years 2002/2003 when it must be established (exported for more than \$10,000 in both years) is shorter. Therefore, after recording country discoveries for each year from 1997-2002, we regress the pooled observations on year dummies as well as total country discoveries. The regression results are shown below in Table A.I. Based on this regression, we have for a given overall level of discovery activity the expected temporal profile of discovery for each of the six years in the window.

Table A.I: Estimation Results for Reclassification Filter

	Discoveries in Country-Year
year1997_dummy	1.457
	(12.36)**
year1998_dummy	0.944
	(7.83)**
year1999_dummy	0.334
	(2.66)**
year2000_dummy	0.146
	(1.15)
year2001_dummy	-0.073
	(0.56)
ln(totalcountrydiscoveries)	1.005
	(25.24)**
Constant	-2.450
	(13.47)**

Observations	438
--------------	-----

Absolute value of z statistics in parentheses

* significant at 5%; ** significant at 1%

Source: Author's Calculations

When a customs agency begins to report a group of products previously grouped in a n.e.s line, there will be an uncharacteristic jump in registered discoveries in one particular year. As such, we flag country/years when the standard errors of these estimates are greater than one standard deviation (as calculated with the pooled data). This identifies outliers, but is biased towards large-discovery countries. Therefore, we also flag country/years when the standard errors of the estimates, normalized by total country discoveries, are greater than one standard deviation of the pooled normalized standard errors. These estimates are biased towards low discovery-activity countries. Therefore, the common set is restricted to extreme outliers, given a country's total discovery activity and temporal effects.

Using this filter, we identify 19 cases when there is a hugely uncharacteristic jump in new exports given the year and overall discovery activity. We treat all identified discoveries during these country/years as reclassifications and drop them from the total country discovery counts. Table A.II shows the country/years discarded.

Table A.II: Country-Years Discarded as Reclassifications

Country	Year
Costa Rica	1997
Estonia	1997
Israel	1997
Netherlands	1997
El Salvador	1997
Slovak Republic	1997
Switzerland	1998
Peru	1998
Paraguay	1998
Sweden	1998
Indonesia	1999
Jordan	1999
Madagascar	2000
Oman	2000
Paraguay	2000
Egypt, Arab Rep.	2001
United Kingdom	2001
Denmark	2002
Egypt, Arab Rep.	2002

Source: Author's Calculations

In addition to reclassifications within countries, we also check the data for reclassifications within products. While the consistent use of the 1988/92 revision should prevent the data from including reclassifications through the addition of a product line to the nomenclature that pre-existed, there is one significant outlier in the data: product

271000 (Oils petroleum, bituminous, distillates, except crude). This particular product was a ‘discovery’ in 14 countries in 2002, almost double the instances of the next most frequent discovery in all years of the window combined. This obvious product/year outlier was therefore dropped from the dataset.

Identified Cases of Inside-The-Frontier Innovation

Table A.III: Discoveries by Country

Algeria	38	Greece	46	Niger	10
Argentina	29	Guatemala	106	Norway	9
Australia	34	Honduras	59	Oman	115
Austria	5	Hong Kong, China	46	Panama	51
Bolivia	88	Hungary	90	Paraguay	37
Brazil	44	India	94	Peru	49
Canada	19	Indonesia	119	Poland	221
Central African Republic	3	Ireland	66	Portugal	27
Chile	42	Israel	78	Romania	114
China	10	Italy	5	Singapore	2
Colombia	73	Japan	11	Slovak Republic	22
Costa Rica	42	Jordan	102	Slovenia	43
Cote d'Ivoire	33	Korea, Rep.	30	Spain	5
Croatia	47	Latvia	68	Sudan	15
Cyprus	26	Macedonia, FYR	42	Sweden	4
Czech Republic	8	Madagascar	19	Switzerland	2
Denmark	30	Malawi	14	Togo	34
Ecuador	57	Malaysia	41	Turkey	30
Egypt, Arab Rep.	32	Mauritius	98	Uganda	17
El Salvador	46	Mexico	13	United Kingdom	31
Estonia	53	Moldova	33	United States	1
Finland	8	Morocco	81	Uruguay	62
France	4	Netherlands	2	Venezuela	38
Gabon	22	New Zealand	20		
Germany	20	Nicaragua	54		

Table A.IV: Discoveries by Leamer Commodity Cluster

Leamer Commodity Group	Number of Discoveries Worldwide
Petroleum	41
Raw Materials	253
Forest Products	147
Tropical Agriculture	129
Animal Products	220
Cereals, etc.	165
Labor Intensive	360
Capital Intensive	587
Machinery	541
Chemical	646

Data Definitions and Sources

Variable Name	Description	Units	Year(s) Used	Transformation	Source
ln(GDP Per Capita)	Natural log of real GDP per capita (PPP)	2000 PPP Constant Prices	1995 (Table 1 uses all years)	log	World Bank WDI
Population	Population	Count	1995	None	World Bank WDI
$\bar{\pi}$	Growth of non-discovery exports (at the Leamer commodity group level) minus non-export GDP per capita growth	decimal form	1994-2003	None	COMTRADE & World Bank WDI
ln(historical discoveries)	Historical discovery counts*	Total counts between 1984 and 1993	1984-1993	log**	COMTRADE
Factor Endowments	Average value of net exports per capita between 1989 & 1993	Current dollars per capita	1989-1993	Net exports for each year divided by that year's population.	COMTRADE & World Bank WDI
Cost of starting a business	Cost of obtaining legal status to operate a firm.	2000 US dollars (PPP)	2004	Original in % of GDP per-capita. Multiplied by 2003 GDP per capita (PPP) to obtain dollar figure. log**	Cost measure: World Bank (2004). GDP per capita (PPP): World Bank WDI.
Days to start a business	Number of days to obtain legal status to operate a firm	Count	2004	log**	World Bank (2004).
Cost to enforce a contract	Cost to enforce a payment dispute in the courts	Percent of the debt	2004	log**	Cost measure: World Bank (2004).
Days to enforce a contract	Number of days to enforce a payment dispute in the courts	Count		log**	World Bank (2004).
Labor market rigidity index	Index of labor market rigidity	index	2004	log**	World Bank (2004).
Barriers	Index of the above five measures of barriers, calculated using principal component analysis (one factor)	index	2004	none (index of logs)	Author's calculations based on World Bank (2004)
Ln(Patents)	Patents in US and EU (US: patents in EU only, EU: patents in US only)	Sum of Counts	1994-1999	log**	Lederman and Saenz (2005)

*Historical counts are identified using export data from 1970 onwards at the SITCr1 3-digit level. The filter identifies a discovery in the year it first appears as an export greater than 0. The period 1974-1983 is used

to create baseline of existing exports, the period 1984-1993 to generate counts of discoveries. The filter drops countries from the sample missing more than 7 years of data in the 1974-1983 period (to ensure at least three years of data exist to identify existing exports) and more than 5 years of data in the 1984-1993 period.

**Before taking logs, 1 added to each to keep observations of 0 in the sample

Descriptive statistics of data used in regressions reported in Table 4 and others discussed in the text.

Variable	Mean	Standard Deviation	Minimum	Maximum
<i>Full Sample (N=510)</i>				
Discoveries by Leamer Commodity Groups	4.42	6.83	0.00	48.00
Log (GDP per capita) in 1995	9.11	0.98	6.33	10.32
Log (historical discoveries by Leamer Groups)	0.29	0.51	0.00	2.64
Factor Endowments [Average net exports per person by Leamer Groups, 1989-1993]	0.00	0.28	-0.90	4.39
$\bar{\pi}$ [Export growth net of non-export GDP]	0.72	4.98	-0.53	35.98
Barriers Index with Cost of Starting a Business	-0.12	0.91	-2.29	1.18
Barriers Index without Cost of Starting a Business	-0.13	0.88	-2.29	1.19
<i>Sample without Malaysia (N=500)</i>				
Discoveries by Leamer Commodity Groups	4.43	6.86	0.00	48.00
Log (GDP per capita) in 1995	9.12	0.99	6.33	10.32
Log (historical discoveries by Leamer Groups)	0.30	0.52	0.00	2.64
Factor Endowments [Average net exports per person by Leamer Groups, 1989-1993]	0.00	0.28	-0.90	4.39
$\bar{\pi}$ [Export growth net of non-export GDP]	0.01	0.13	-0.53	0.78
Barriers Index with Cost of Starting a Business	-0.11	0.92	-2.29	1.18
Barriers Index without Cost of Starting a Business	-0.12	0.88	-2.29	1.19