Discovery and Development: An Empirical Exploration of 'New' Products*

Bailey Klinger

Daniel Lederman

Office of the Chief Economist
Latin America and the Caribbean
World Bank

Abstract: *This paper uses disaggregated export data to explore the relationship between economic discovery and economic development. We find that discoveries, or episodes when countries begin exporting a new product, are not limited to so-called 'dynamic' industries, rather they also occur in traditional sectors such as agriculture. In addition, the data suggest discovery is a component of the stages of productive diversification that occur with development, following a consistent pattern: discovery activity peaks at the lower-middle income level and then declines. Based on this pattern, we show that discovery in the 1990s occurred with a higher than expected frequency in Eastern Europe and Central Asia, and lower than expected frequency in Sub-Saharan Africa. Discovery is not found to be a product of structural transformation based on changing factor endowments across income levels. Beyond export growth, population, and development, there are no significant and positive relationships between the expected drivers of entrepreneurship and the frequency of discovery. Combined with the finding that higher absorptive capacity and lower barriers to entry are associated with a reduction in discovery, this suggests that market failures arising from imitation and free-riding may be inhibiting the emergence of new export products in developing countries.*

1. Introduction

The recent economic performance of developing countries, particularly the Latin American and Caribbean economies, has left many policy makers puzzled. New industries (that is, new to the country, not necessarily new to the world) have not been emerging with the frequency expected before the reforms of the 1990s. One explanation for this experience is that market failures are inhibiting the emergence of 'new' production in developing countries, a theory suggested in numerous articles[1].

While interesting from a theoretical standpoint, most of these models have been light on evidence. They provide little empirical support for the market failures hypothesis that they suggest. In fact, we know very little about the empirical relationship between 'new' production and economic development. This paper attempts to fill this void.

We refer to the emergence of a new product (i.e. the successful production of a good by a country that did not produce it before) as an instance of 'economic discovery', a term used by Hausmann and Rodrik (2003a). Our results reveal some robust relationships that serve to deepen our understanding of the discovery process. First, we find a consistent pattern of discovery activity across income levels, which are strikingly consistent with recent empirical findings on productive diversification and development. We do not find that discoveries across industries and product types are functions of the level of development, as suggested by a factor-endowments view of structural transformation and economic growth. Finally, we find preliminary evidence in support of the hypothesis that market failures associated with free-riding and imitation are in fact inhibitors of discovery.

---

[1] Hausmann and Rodrik (2003a), Vettas (2002), Mayer (1984), for example.

In the following section, we examine these three theoretical perspectives of the phenomenon of discovery: discovery as a component of productive diversification, discovery as a result of structural transformation, and the lack of discovery as a consequence of market failures. In addition, we highlight relationships these theoretical perspectives suggest we will find in the data. After a discussion of the data and methodology used to identify discoveries, as well as some stylized facts on the frequency of discovery across countries and across industries, we subject the theoretical predictions to empirical testing.

2. Theories of Discovery and Development

This section reviews three theoretical perspectives related to discovery. Each perspective implies certain relationships, which are subjected to empirical testing in subsequent sections.

1) Discovery and development: a component of productive diversification

Consider output as simply the sum of income per good across all goods produced:

$$(1) \quad Y = \sum_{i=1}^{J} x_i .$$

When considering production in this light, there are basically two channels for a country to increase national income, Y: (a) increasing the value of production of goods (x_i's) already produced in a given economy, and/or (b) increasing the number of varieties (J) by adding new x_i's (that is, discovery). While discovery is a potential source of growth, we do not know if or how its relative importance changes over the process of development. However, there is a related phenomenon that links discovery and growth which has been studied in the literature: the process of productive diversification.

Imbs and Wacziarg (2003) analyze the process of diversification, considering how it behaves across income levels. They summarize the theoretical support for both positive and negative monotonic relationships between diversification and growth. After examining the data, these authors find that neither are 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, after which point economies become increasingly specialized.

This perspective has implications for the relationship between growth and discovery, given the connection between discovery and diversification. Consider the measure of diversification that we apply in our empirical treatment of the topic in section 3: the Herfindahl index (H), where each i is an individual product and J is the total number of products:

$$(2) \quad H = \sum_{i=1}^{J} \left(\frac{x_i}{\sum_{i=1}^{J} x_i} \right)^2 .$$

Diversification can be increased by either adding another i, thereby increasing J (discovery) or by equalizing the x's for a fixed J: producing more evenly across a given set of goods. Discovery is therefore one of two channels through which diversification can occur, and theories predicting a certain relationship between growth and diversification by extension predict a relationship between discovery and growth. The Imbs and Wacziarg U-shaped relationship between diversification and development would lead us to expect a particular relationship between discovery and development: an inverted U-shaped relationship. The level of income at which the discovery curve (the curve relating discovery activity to income per capita) would peak depends

4

on the relative importance of the two channels of increasing diversification across levels of development. Section 4 analyzes the empirical relationship between income, discovery, and diversification.

2) Discovery and development: a part of structural change based on factor endowments

Rather than equation (1), consider the fundamental framework for analyzing output and growth in economics, namely the aggregate production function. Stated in its most basic, generalized form, national income (Y) depends on an economy's factor endowments, such as physical capital (K), human capital (H), and unskilled labor (L):

$$(3)\, Y = F(K, H, L).$$

When countries grow, their relative endowments of factors of production change, which in turn determines patterns of production across income levels. As pointed out by Leamer (1984), less developed countries with a relative abundance of L will specialize in traditional labor-intensive goods. However, as they grow, their stock of capital (both human and physical) increases, causing them to shift production towards more capital-intensive goods.

This theoretical perspective has its own implications for what we would expect to find in our discovery data. Discoveries may be capturing the structural transformation that occurs with growth, in that discoveries will be concentrated in certain industries at low levels of development and in others at higher levels of development. If this factor-endowments story accurately characterizes discovery patterns across income levels, we would expect the relative frequency of

discoveries to change across industries as growth occurs. This theoretical prediction is tested in section 5.

3) Discovery and development: market failures

The third theoretical perspective relates to discovery and the effect of potential market failures. There have been many models in the economics literature that suggest that market failures inhibit the discovery process, thereby harming development. One, which has already appeared in policy documents (FUSADES 2004), is Hausmann and Rodrik's model of "Economic Development as Self-Discovery" (2003a). This model suggests that while factor endowments explain broad patterns of production across countries, production functions for goods at a disaggregated level are not known or predictable *a priori* by entrepreneurs, who must experiment in order to determine what can be produced in a given national context. However, once an entrepreneur has an experiment that pays off and they 'discover' a profitable product, others can easily imitate their success, free-riding on the initial investments in experimentation and driving down the entrepreneur's profits. The result is a market failure, whereby entrepreneurs are not able to reap the full benefits of their discovery investment and will consequently under-invest in experimentation. As there is social value in discovering what can be produced in each country setting, and yet competition can lead to underinvestment in the experimentation required to make these discoveries, there is scope for public intervention.

Vettas (2002) suggests another model, where the uncertainty is not of production costs but of foreign demand. In this model, the characteristics of foreign demand for a new product are unknown initially, and must be discovered. Furthermore, foreign demand for new products is

endogenous, in that it is an increasing function of past sales due to learning on the part of consumers (up to a maximum point, which is not predictable *a priori*). However, the initial investment required to penetrate a new market, stimulate demand, and learn the market's potential size will suffer the same appropriability problem: imitators can free ride, leading to underinvestment by entrepreneurs. Both the uncertainty and endogeneity of foreign demand would justify subsidizing entry into new markets.

Based on a similar argument of free-riding on market-cultivating expenditures, originally advanced by Bhagwati (1968), Mayer (1984) constructs a model of foreign-market cultivation which assumes actual consumption experiences are required to learn about a commodity's qualities. The model indicates that subsidization of infant-exporters is a first-best policy. Another extension relates to foreign standards (Granslandt and Markusen 2000). When attempting to export a good to a foreign market, the first entrant will have to make the initial investments in product and process redesign to meet foreign product-safety standards. However, market failures will arise if redesigns are non-excludable, as free-riding will reduce returns of the first entrant.

While interesting, these models have not been subjected to systematic empirical testing. This is likely due in part to a lack of disaggregated worldwide production data, combined with no obvious method of testing for the presence of these market failures. As described in section 3, we will use disaggregated export data, which unlike domestic production data, is widely available at a highly disaggregated level. In addition, we suggest the following framework to evaluate the importance of market failures in the discovery process.

Entrepreneurs deciding whether or not to invest in a new activity will base their decision on the experiment's expected profitability π. As suggested by the models mentioned above, a portion of this profitability is likely to be unknown and product-specific, and can therefore only be determined through experimentation and investment. This unknown component of expected profits could be production costs à la Hausmann and Rodrik (2003a), the characteristics of foreign demand à la Vettas (2002) and Bhagwati (1968), or the redesign needed to meet foreign trade regulations à la Ganslandt and Markusen (2000). In models such as Hausmann and Rodrik's (2003a), this parameter is entirely a property of individual goods, and not of entrepreneurs or countries. However, it is reasonable to suggest that part of profitability is *not* product-specific, and rather is based on national productivity. Country characteristics such as a sophisticated financial system and an educated workforce affect the profitability of all industries in an economy (creating absolute advantage), and a country's relative endowments of the factors of production also grant a comparative advantage in certain sectors and activities. We therefore view π as encompassing this entire range of the determinants of profits at the product level, from unknown and product-specific to the drivers of national comparative and absolute advantage. This approach is consistent with Neary (2003), who constructs a general equilibrium model that incorporates both comparative and absolute advantage, both of which are shown to be relevant to our considerations of discovery.[2]

As a particular product's expected profitability π (which encompasses product-specific and economy-wide profitability) rises, the likelihood of a first mover experimenting, and therefore the probability of observing a discovery of that good ($P[D]$), would also rise. $P(D)$ is

---

[2] Neary (2003) finds that although comparative advantage determines directions of trade, both comparative and absolute advantage have an impact on resource allocation, trade patterns, and trade volumes.

therefore a positive function of the particular activity's expected profitability π. If there were no concerns regarding imitation, then $P(D) = D(\pi)$, and $\dfrac{\partial P(D)}{\partial \pi} = \dfrac{\partial D}{\partial \pi} \geq 0$. However, if the first mover knows that they will only be able to appropriate a certain proportion of the profits arising from their discovery, then the probability of observing a discovery of that good is

$$(4) \quad P(D) = D(\pi \times q),$$

where the appropriability parameter q represents the proportion of the entrepreneur's profits that they are able to appropriate, and is therefore between 0 and 1.

The probability of other entrepreneurs deciding to imitate the first mover also depends on expected profitability. Huge profits will draw a stampede of imitation, decreasing the first mover's ability to appropriate their profits. Therefore, the appropriability parameter q is a negative function of π. In addition to the *desire* to imitate a first mover, which is captured by introducing π as an argument in the appropriability parameter's function, imitation may also be hindered or helped by the *ability* to imitate. To capture this, we add a second term (α) to the appropriability parameter that represents the ease with which potential entrepreneurs can imitate the first mover. Therefore, appropriability for the first mover (q) is a negative function of both profitability and ease of imitation:

$$(5) \quad q = q(\underset{-}{\pi}, \underset{-}{\alpha}).$$

Combining (4) and (5) gives a simple framework with which we can evaluate the importance of market failures:

$$(6)\ P(D) = D(\pi \times q(\pi, \alpha)).$$

If there are no market failures associated with imitation and the first mover is able to appropriate all private profits, then q is equal to 1. In this case, from (6), $\frac{\partial P(D)}{\partial \pi} = \frac{\partial D}{\partial \pi}$, which is ≥ 0. However, if imitation is hindering experimentation, then q is no longer a constant, and $q(\pi, \alpha)$ becomes important in the decisions of the first mover. We see from (6) that:

$$(7)\ \frac{\partial P(D)}{\partial \pi} = \frac{\partial D}{\partial \pi}\left[q(\pi,\alpha) + \pi \times \frac{\partial q}{\partial \pi}\right].$$

We assume that $\frac{\partial D}{\partial \pi}$ is positive but decreasing in π. By definition, $0 \leq q \leq 1$ (that is, the first mover can appropriate at best all profits, and at worse nothing). In addition, we restrict our analysis to the interesting case in which expected profits π are positive. As $\frac{\partial q}{\partial \pi} \leq 0$, $\frac{\partial P(D)}{\partial \pi}$ is less than it was in the case of no imitation. In fact, if market failures due to imitation are particularly acute, it is possible that $\frac{\partial P(D)}{\partial \pi}$ could be negative (see Appendix I for graphical representation). That is, the negative effect of π via q could outweigh the positive effects of π for the first mover, meaning $\frac{\partial P(D)}{\partial \pi} \leq 0$. The sign of $\frac{\partial P(D)}{\partial \pi}$ is therefore a measure of the importance of market failures.

In addition, from (6), we see that if there are no market failures associated with imitation, q is a constant and $\frac{\partial P(D)}{\partial \alpha} = 0$. However, in the presence of imitation, it follows from (6) that

10

$$(8) \quad \frac{\partial P(D)}{\partial \alpha} = \frac{\partial D}{\partial q} \times \frac{\partial q}{\partial \alpha},$$

which by definition is negative. Therefore, in addition to $\frac{\partial P(D)}{\partial \pi}$, $\frac{\partial P(D)}{\partial \alpha}$ is also of interest. Although we do not observe $P(D)$ directly, we do observe the total number of discoveries, which is an increasing function of the probability of discovery of a particular good $P(D)$. The relationship between total discoveries and variables that affect π and α is therefore the object of study. If a significant negative relationship is found, this supports the market failures hypothesis: q is affecting $P(D)$, meaning that imitation is hindering discovery.

In section 6 we apply this framework to our discovery data, suggesting variables that affect π and α and determining their relationship with discovery. However, we first describe the data and methodology used to identify discoveries, followed by some stylized facts based on the results.

3. Data, Methodology, and Stylized Facts

Data

In the search for economic discoveries, domestic production data would be the first choice. However, as production data are not available at a highly disaggregated level, we use export data. The problem with using export data is obvious: a product emerging as a new export may have been produced domestically for some time, and therefore would not represent an economic discovery. However, the main advantage is that export data are recorded at highly disaggregated levels for customs purposes. In addition, exporting a particular good for the first

time, even if it was already produced domestically, is itself an entrepreneurial act that requires discovery (Ibeh 2003). Using export data allows us to consider all of the market failures associated with this entrepreneurial act, including those discussed in section 1 that are specific to exports.

Worldwide export data are taken from the United Nations COMTRADE[3] database, which contains global exports by country under multiple classification systems. These data are used in many publications to analyze export dynamism and growth as well as geographic patterns in export growth (UNCTAD 2003, Mayer, Butkevicius and Kadri 2002, Lall 1998 & 2000), but have not been used to study the emergence of new products.

Export data at a highly disaggregated level are available from COMTRADE under the Harmonized Commodity Description and Coding System (HS) beginning in the early 1990s. There are over 55 countries reporting export data under the first revision of the Harmonized System (HS 1988/1992) by 1992, and 70 countries reporting by 1993 (see Appendix II for sample composition).

Data are available at multiple levels of disaggregation under the Harmonized System, and it is not obvious what the level of disaggregation for the analysis of discovery events should be. Greater disaggregation allows for the study of more specific products, and it is at this individual product level where uncertainty of production costs and market demand may require experimentation. However, there is a tradeoff when using higher levels of disaggregation, as at a

---

[3] United Nations Department of Economic and Social Affairs (UN/DESA) Commodity Trade Statistics Database, accessed via the World Bank World Integrated Trade Solution (WITS) tool, July 2004.

certain level the differences between products may not be meaningful from a discovery standpoint. For example, within the broad category of textiles, it is possible that discovering that a country can profitably produce shirts is different from discovering that it can profitably produce hats or bedsheets. However, there may not be a difference between discovering that a country can profitably produce long sleeve shirts as opposed to short sleeve shirts. With overly-disaggregated data, the filter will identify a product that is new to the export basket from an accounting viewpoint, but not from a discovery viewpoint[4]. As the appropriate level of disaggregation is not clear, we will use data for both the HS 4-digit level (approximately 1200 commodity groups) and the HS 6-digit level (approximately 5000 commodity groups). The level of disaggregation does bias the results in favor of certain industries over others, an issue discussed below.

In order to evaluate the robustness of some of our results, we also consider export data under SITC revision 1 system. Data under this classification system are available for a much longer time period than the HS data. However, the consistent lowest common denominator for this data over time is at the 3-digit level, which is highly aggregated and includes only around 175 commodity groups. Therefore, this level of aggregation may be too high to capture discoveries at the level where market failures arise, and therefore would not be suitable for our purposes. However, these data are available as far back as the 1970s, and allow us to exploit more robust time-series estimation techniques. As such, we use this time-series data to verify some of the results of the more disaggregated 1990s HS data.

---

[4] The issue of the proper level of disaggregation is also problematic in the literature on intra-industry trade (Grubel and Lloyd (1975).

Methodology: Defining a Discovery

With the 1990s HS data, we define a discovery as a product which was not sold abroad in a large amount at the beginning of the 1990s (exports were less than $10,000 in 1992, or in 1993 if 1992 data from the country were not reported), but by the end of the 1990s was consistently exported in a large quantity (exports over $1,000,000 in 2000, 2001 and 2002 or in 1999, 2000 and 2001 if 2002 data were not yet reported). These cutoff values are arbitrary, but the results were not sensitive to the choice. Nominal amounts are used so that, from a product level, the filter applies to each country in the same manner.

It is possible that increases in exports for some goods may be due largely to price effects rather than increased production. However, we cannot net out or deflate price effects with the data. Furthermore, as part of the valuable information that must be discovered is demand, new goods whose prices quickly rise represent more valuable discoveries than new goods whose prices stagnate.

In order to verify that this definition (which we will call Filter 1) does not capture goods that were exported in large quantities prior to 1992 and then fell temporarily in that year, we also employ a second filter (Filter 2) that only considers goods that were not exported for more than $10,000 in 1992, 1993 or 1994, but topped $1,000,000 in 2000 and 2001. Though more restrictive (it lowers the count of discovery events as it only considers discoveries in the second half of the 1990s), this filter rules out false identification of new products. All of the findings discussed below are found to be consistent across filters.

With the SITC time-series data, we correct for inflation to make discoveries comparable across time periods (using US producer PPP data from the US Federal Reserve), and use a moving window that defines a discovery for a particular year as a category for which exports were never more than $1,000,000 (1985 US dollars) before that year, crossed the $1,000,000 mark that year, and subsequently were exported for more than $10,000,000 ten, eleven and twelve years later. The dollar amounts used by this filter obviously have to be higher than with the 1990s HS data because the time series data are highly aggregated. This moving window identifies discoveries in each year from 1973 to 1990. In addition, any good that was not exported for more than $1,000,000 before 1991, but by the end of the 1990s was exported for more than $10,000,000 (specifically in 2000, 2001 and 2002) is recorded as a discovery in the 1990s.

The countries that reported data in the requisite years and are included in the sample are listed in Appendix II. We will now discuss some stylized facts based on the discoveries identified using this methodology.

Stylized Facts

Using these filters to search for discovery events in the UN COMTRADE export data, we identify 332 instances of discovery in the HS 4-digit data, and 1710 instances in the more disaggregated HS 6-digit data. Discoveries by industry are shown in Table 1.

Table 1: Discoveries by Industry

	4-Digit Level		6-Digit Level	
	Count	Percentage	Count	Percentage
Anmial & Animal Products	12	4%	99	6%
Vegetable Products	32	10%	86	5%
Foodstuffs	22	7%	67	4%
Minteral Products	45	14%	91	5%
Chemicals & Allied Industries	55	17%	310	18%
Plastics / Rubbers	6	2%	69	4%
Raw Hides, Skins, Leather & Furs	0	0%	7	0%
Wood & Wood Products	27	8%	115	7%
Textiles	23	7%	145	8%
Footwear & Headwear	2	1%	6	0%
Stone & Glass	13	4%	57	3%
Metals	37	11%	249	15%
Machinery & Electrical	15	5%	205	12%
Transportation	25	8%	87	5%
Miscellaneous	18	5%	117	7%
Services	0	0%	0	0%
Total	332	100%	1710	100%

Source: Author's calculations

Notice how the results change when going from 4 to 6 digits. The relative frequency of discoveries falls for agricultural and mineral products (vegetable products from 10% to 5%, mineral products from 14% to 5%) and rises for machinery and electrical goods (from 5% to 12%). This change can be explained by the fact that higher levels of disaggregation affect sectors differently. With commodities like agricultural goods, which face natural barriers to product diversification, greater levels of disaggregation lead to few additional products[5]. For example, there is no greater disaggregation of bananas from the four digit to six digit level (0803 at the four-digit level, which is only comprised of one category at the six-digit level: 080300). Grapes (0806) are only disaggregated into two groups: fresh and dried. However, higher levels of disaggregation lead to a larger number of additional manufactured goods. For instance, the 4-digit group 8511 (all types of electrical ignition, generators, parts) is disaggregated into seven different products (spark plugs, ignition magnetos, distributors and ignition coils, starter motors, generators and alternators, glow plugs and other ignition or starter equipment, and parts of electrical ignition or starting equipment).

---

[5] Even genetically-modified crops are not listed as different varities.

If the higher level of disaggregation better reflects the true range of different products that countries could produce, then the 6-digit data are most appropriate. However, as discussed above, it may also be the case that the differences among products at the 6-digit level are not meaningful from a discovery standpoint. If this is true, and the higher level of disaggregation does not reflect the range of different products countries could produce, then going from 4 to 6 digits would bias the results towards those sectors that, for accounting reasons, are decomposed into a greater number of subcategories. It is not clear which is the case, therefore results for both levels of disaggregation will be reported throughout the paper.

Notwithstanding this issue, the results shown in Table 1 are quite interesting. When discussing areas of rapid export growth, researchers identify a relatively narrow range of dynamic products, such as clothing or electronics (Butkevicius, Kadri & Mayer 2002). However, results in Table 1 show that economic discoveries in the 1990s were not highly concentrated in certain 'modern' sectors. In fact, sectors considered to be more 'traditional', like foodstuffs & agriculture, chemicals, and metals, were are also important sources of discoveries. Chemicals and allied industries contain the highest share of discovery activity at either level of disaggregation. This result lends support to a more broad vision of discovery that is not focused on a certain group of manufactured products.

Considering discovery activity by country also gives some interesting results. Discoveries by country, using both HS 4-digit and 6-digit data, are shown in Table 2.

Table 2: Discoveries by Country

	HS 4-Digit		HS 6-Digit	
	Discovery Count	Percent	Discovery Count	Percent
Argentina	5	2%	32	2%
Australia	5	2%	22	1%
Burundi	0	0%	0	0%
Bolivia	7	2%	15	1%
Brazil	10	3%	44	3%
Central African Rep.	0	0%	1	0%
Canada	2	1%	19	1%
Switzerland	2	1%	19	1%
Chile	8	2%	31	2%
China	8	2%	39	2%
Colombia	17	5%	43	3%
Cyprus	4	1%	5	0%
Czech Republic	17	5%	58	3%
Germany	0	0%	17	1%
Denmark	1	0%	10	1%
Ecuador	11	3%	30	2%
Spain	1	0%	24	1%
Finland	5	2%	30	2%
United Kingdom	4	1%	24	1%
Greece	2	1%	18	1%
Guatemala	4	1%	9	1%
Hong Kong, China	2	1%	7	0%
Croatia	4	1%	11	1%
Hungary	5	2%	92	5%
Indonesia	29	9%	160	9%
India	8	2%	53	3%
Ireland	1	0%	19	1%
Iceland	3	1%	5	0%
Japan	1	0%	4	0%
Korea, Rep.	9	3%	51	3%
Macao	0	0%	5	0%
Morocco	10	3%	19	1%
Mexico	10	3%	66	4%
Mauritius	1	0%	4	0%
Malaysia	2	1%	42	2%
Nicaragua	8	2%	12	1%
Netherlands	2	1%	19	1%
Norway	1	0%	13	1%
New Zealand	4	1%	10	1%
Oman	15	5%	47	3%
Peru	20	6%	66	4%
Portugal	6	2%	25	1%
Paraguay	7	2%	9	1%
Romania	26	8%	102	6%
Saudi Arabia	3	1%	18	1%
Singapore	1	0%	11	1%
Sweden	0	0%	24	1%
Thailand	8	2%	63	4%
Trinidad and Tobago	5	2%	28	2%
Turkey	13	4%	135	8%
Taiwan, China	4	1%	58	3%
United States	0	0%	3	0%
South Africa	11	3%	39	2%
Total	332	100%	1710	100%

Source: Author's calculations

The instances of economic discovery in the 1990s were not evenly spread across all countries.

However, the distribution does not seem to be random. There is a pattern between discovery

activity and the level of development. Specifically, we see that discovery activity is low among the poorest countries, but interestingly is also low among wealthy industrialized countries. At the 4-digit level, there were no discoveries within both the Central African Republic and the United States, while there were 26 discoveries in Romania and 29 in Indonesia. The frequency of discoveries appears to be a nonlinear function of the level of development. This apparent relationship may be the result of economic discoveries being driven by broad economic changes that occur as countries become richer. One potential explanation for this observation was identified in section 2, namely the process of productive diversification, which we now evaluate.

4. Discovery and the process of productive diversification

As discussed in section 2, recent work has found a robust relationship between economy-wide diversification and levels of development. Imbs and Wacziarg (2003) find a persistent pattern of increasing diversification until relatively high levels of development, (GDP per capita between $9,000 and $10,000 1985 US dollars) followed by increased specialization.

Before investigating if discovery is driven by these stages of diversification, we first expand on the findings of Imbs and Wacziarg (2003) by considering diversification of the export basket. This is because Imbs and Wacziarg (2003) used labor data in their analysis of the stages of diversification, while we use export data to identify discoveries. It may not be true that the export basket follows a similar pattern of diversification as the national production basket. To determine whether these same stages of diversification exist in export data, we construct a

Herfindahl index (H) of exports for each country using the HS 4-digit, HS 6-digit, and SITC 3-digit export data, and estimate the following equation[6]:

$$(9)\ H = \beta_0 + \beta_1(GDPpercapita) + \beta_2(GDPpercapita)^2$$

We use GDP per capita rather than the log of GDP per capita to remain consistent with the approach of Imbs and Wacziarg (2003). The results, summarized in Table 3, indicate that, similar to total production, 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 result is also found when employing a fixed-effects OLS regression on the SITC data from 1972-2002.

Table 3: Stages of Export Diversification

Variable	HS 4-Digit 1995	HS 6-Digit 1995	SITC 3-Digit 1972-2002 (FE)
Coefficient on GDP per capita	-.0000333 (-4.55)	-.0000183 (-3.07)	-.0000173 (-7.01)
Coefficient on GDP per capita squared	9.53e-10 (3.30)	5.25e-10 (2.44)	4.98e-10 (6.71)
Minimum Point (highest level of diversification)	$17,471	$17,429	$17,369
Adjusted R-Squared (OLS) / F-Statistic (FE)	.2683	.2047	60.79
Number of Observations / Groups	100	53	146

Note: parentheses indicate t-statistics. Source: Author's calculations

All three data types yield a level of GDP per capita at which economies switch from diversification to specialization between $17,350 and $17,500. Using domestic production data, Imbs and Wacziarg (2003) find a switching point in the $13,150 to $14,600 range (in equivalent 1996 dollars), which is slightly lower than our findings using export data. These results support

---

[6] Measures of GDP per capita are 1996 PPP values from PWT 6.1 (Aten, Heston and Summers 2002). See (3) in section 2 for the definition of the Herfindahl index. The value of the Herfindahl index ranges from 0 to 1, with lower values indicating greater diversification of export earnings.

the contention that the pattern of economic diversification observed by Imbs and Wacziarg is probably driven by patterns of international trade flows. In sum, economies engage in diversification until a relatively high level of development, after which a process of economic specialization takes hold.

This pattern of trade-driven economic diversification may explain the apparent relationship between the frequency of discoveries and the level of economic development depicted in Table 2. We expect countries at relatively low levels of development to have more frequent incidents of economic discovery, as they are in the process of diversifying their economies. However, as income rises, the frequency of these events declines, particularly at high levels of development when economies experience rising specialization. The point at which the number of discoveries reaches its maximum depends on the relative importance of the two channels of increasing diversification (i.e., new goods or more even production).

To analyze this issue, we turn to the empirical relationship between discovery frequency and the level of development. Because our dependant variable is count data with a substantial number of zeros, we apply the Poisson-distribution model in order to estimate the relationship between the number of discoveries and GDP per capita:

$$(10) \quad \lambda = e^{\beta_0 + \beta_1 (\ln GDP percapita) + \beta_2 (\ln GDP percapita)^2}$$

where λ is the number of discoveries per period[7].

---

[7] We began with a Poisson regression, however the likelihood-ratio test indicated that the data are overdispersed.

We use both discovery filters described in section 3 on the HS 4-digit and HS 6-digit data during the 1990s. In addition, we examine the SITC data in the same manner, first with a cross-section of discoveries in the 1990s, and with a conditional fixed-effects negative-binomial regression on discoveries since 1972 (including year dummies)[8]. The results are shown in Table 4.

Table 4: Stages of Discovery

	HS 4-Digit 1990s		HS 6-Digit 1990s		SITC 3-Digit	
	Filter 1	Filter 2	Filter 1	Filter 2	1990s	1972-1991 (FE)
Coefficient: ln GDP per capita	15.65175 (5.68)	13.95415 (3.72)	15.72022 (5.89)	12.06436 (4.57)	22.38211 (4.38)	19.53553 (7.47)
Coefficient: ln GDP per capita squared	-.9280328 (-5.92)	-.821797 (-3.85)	-.9114158 (-6.03)	-.695039 (-4.65)	-1.32140 (-4.50)	-1.16821 (-7.64)
Maximum Point	$4595	$4866	$5564	$5878	$4765	$4278
Pseudo R-Squared	.1515	.0980	.0719	.0549	.1487	
Discovery Count	332	150	1710	865	93	1114
Sample Size	50	49	50	49	67	76

Note: brackets indicate z-statistics. Source: Author's calculations

According to the data, discovery activity is low among the poorest countries, but rises quickly and reaches a maximum when countries earn between $4200 and $5500 per capita. After that point, discovery activity tends to fall, and is low as countries reach a relatively high level of development. Notice that the coefficients fall when moving from filter 1 to filter 2. This is expected, as the more restrictive filter has fewer discovery counts. What matters, however, is that the relationship (that is, signs of the coefficients and the level of GDP per capita at which the expected discovery count curve reaches a maximum) is consistent.

---

[8] On the fixed effects negative binomial estimator, see Hausman et. al. (1984)

This relationship between discoveries and development is illustrated in Figures 1 and 2, which show both a scatter plot of discoveries against GDP per capita, and the estimates of equation (10): the discovery curve.

Figure 1: Discovery Events, HS 4-Digit **Figure 2: Discovery Events, HS 6-Digit**
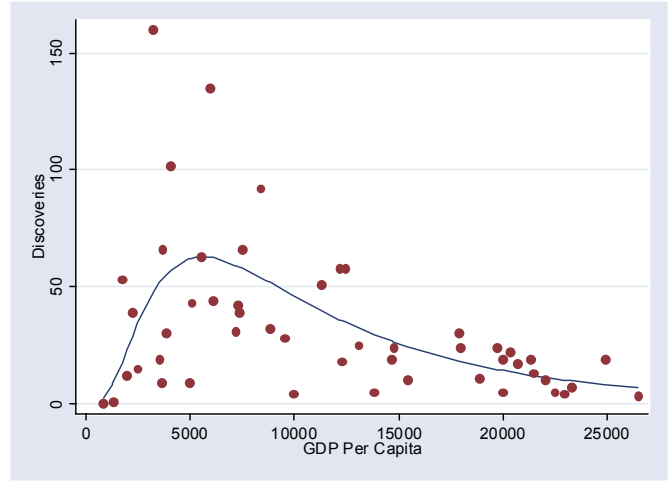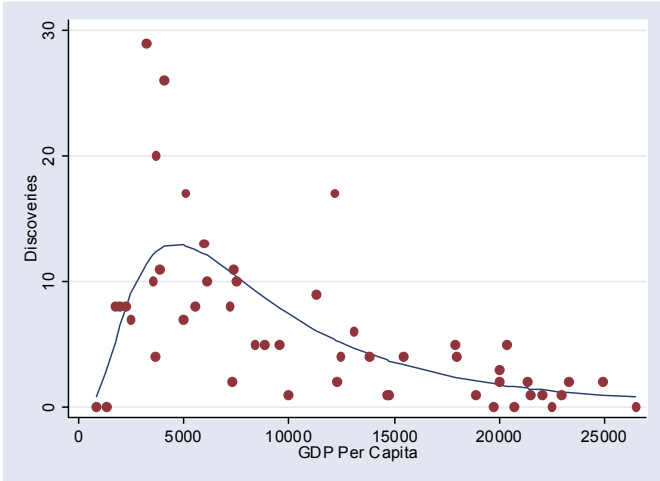
Source: Author's calculations

These results suggest the following pattern. The initial stages of the diversification process tend to be driven by the introduction of new products (discoveries). However in later stages of the diversification process, when discovery activity declines, productive diversification is driven by more even production among the goods the country already produces. Finally, at high levels of income, discovery activity falls, and the diversification process is reversed as production becomes more specialized.

Based on this robust relationship between discovery and development, we now further refine the stylized facts from section 3. Specifically, there is a certain level of discovery activity that we would expect to find in an economy given its level of development. However, as Figure

1 and 2 show, there are significant fluctuations around this expected relationship. Given levels of development, there are some countries that are over-performing in expected discovery activity, and others that are under-performing.

To further study the pattern of discoveries across countries and regions, we utilize a more flexible functional form. To this point, we have used GDP per capita in order to analyze the relationship between discovery and development, and specifically to determine the income level at which the discovery curve peaks. However, we would expect that country size would affect the number of economic discoveries. That is, a country of 300 million has a larger number of entrepreneurs and businesses that could discover new products for export when compared to a country of 300 thousand. When considering the absolute number of discoveries in an economy, population should be accounted for separately from the effect of the scale of each national economy. Therefore, we add population to the model, which enters as statistically significant and positive using the HS 4- and 6-digit well as SITC 3-digit data (see estimation results in Appendix III). In order to allow for flexibility in the effects of population and wealth, we switch from GDP per capita and GDP per capita squared to total GDP and total GDP squared when population is included in the model. We will refer to (11) as the basic model of discovery:

$$(11) \quad \lambda = e^{\beta_0 + \beta_1 (\ln GDP) + \beta_2 (\ln GDP)^2 + \beta_3 (\ln Population)}$$

We estimate this equation using a negative binomial regression (see Appendix III for estimation results), calculate the residuals, and then scale them by dividing each by the standard deviation of the residuals. These standardized residuals are measures of over- or under-

performance in discovery activity, with positive values indicating that the frequency of economic discovery is higher than expected, given a particular population and income level.

We perform t-tests on the means of these standardized residuals, grouped by region. Rejecting the hypothesis that the mean is greater than or equal to zero suggests under-performance, and rejecting the hypothesis that the mean is less than or equal to zero suggests over-performance. Note that this is over- and under-performance relative to conditional world averages, not to an 'optimal' level of discovery activity. Determining a theoretical optimal frequency of discovery is outside the scope of this paper. The results of these tests, as well as the relevant significance levels, are reported in Table 5.

Table 5: Over- and Under-Performance in Discoveries by Region

	Latin America & Caribbean	Eastern Europe and Central Asia	Sub-Saharan Africa	Middle East & North Africa	South Asia	East Asia & the Pacific
HS 4-Digit 1990s	Over-perform 15% level	Over-perform 10% level	Under-perform 20% level			
HS 6-Digit 1990s		Over-perform 5% level	Under-perform 5% level			Over-perform 20% level
SITC 1990s		Over-perform 15% level	Under-perform 5% level			
SITC 1980s	Under-perform 1% level	Over-perform 5% level	Over-perform 1% level	Under-perform 1% level	Over-perform 1% level	Over-perform 1% level
SITC 1970s	Under-perform 1% level			Under-perform 1% level	Over-perform 5% level	Over-perform 1% level

Note: Blank indicates no conclusion can be drawn, either because no statistically significant relationship, or because the sample does not include at least three countries from the region. Source: Author's calculations

On the whole, the evidence suggests that based on their populations GDPs, there was under-performance in discovery activity in Africa during the 1990s, and over-performance in Eastern Europe and Central Asia. In addition, contrary to their over-performance during the

1970s and 1980s, the Asia and the Pacific regions have not been systematically over-performing in discovery activity during the 1990s. While Latin America, the Middle East and North Africa did under-perform during the 1970s and 1980s, there is no evidence in support of systematic under-performance in these regions during the 1990s. These patterns are different than those observed in overall export growth, which has been highest in Asia, followed by Latin America, Africa, and the Middle East (Lall 1998).

We now turn to the second theoretical perspective of discovery described in section 2 and analyze the relationship between discovery activity and the structural transformation of economies as they develop, from traditional labor-intensive or natural resource-intensive goods to more capital-intensive goods.

5. Economic Discovery and Structural Transformation

The factor-endowments theory of production patterns and development suggests that discovery could be driven in part by the structural transformation of economies as they grow. If this were true, then we would find that discoveries in 'traditional' labor-intensive sectors peak at lower levels of development, and then fall as they are replaced by discoveries in 'modern' sectors. In order to test for this relationship, we perform a fixed-effects negative binomial regression to estimate (10), with discoveries disaggregated and grouped into 16 industry panels. In order to test whether the relationships with income per capita and the corresponding maximum point of the discovery curve are different from those estimated in Table 4, we revert back to the model including GDP per capita, rather than GDP and population. This approach provides estimates of the average relationship between GDP per capita and discovery counts across

countries but within industries. If industry characteristics affect the results, then the panel and pooled estimates will be quite different, and the factor endowments perspective will be shown to contribute to our understanding of discovery. The results using both HS 4-digit and HS 6-Digit data are shown in Table 6.

Table 6: Testing for Structural Transformation

	HS 4-Digit		HS 6-Digit	
	Discoveries by Industry	Discoveries Pooled	Discoveries by Industry	Discoveries Pooled
Coefficient on ln GDP per capita	13.5666 (7.02)	15.6518 (5.68)	11.9700 (9.56)	15.7203 (5.89)
95% Confidence Interval	9.7769 to 17.3562	10.2503 to 21.0532	9.5152 to 14.4248	10.4869 to 20.9535
Coefficient on ln GDP per capita squared	-.8109 (-7.29)	-.92803 (-5.92)	-.6970 (-9.83)	-.9114 (-6.03)
95% Confidence Interval	-1.02 to -.5929	-1.2353 to -.6208	-.836 to -.5581	-1.2074 to -.6154
Maximum Point	$4295	$4595	$5358	$5564
Countries in Sample	50	50	50	50

Note: parentheses indicate z-statistics. Source: Author's calculations

The coefficient estimates of the data grouped by industry fall within the 95% confidence intervals for the pooled data at both the 4- and 6-digit levels. Furthermore, the maximum points occur at very similar levels. That is, the data suggest that the observed relationship between discovery activity and income per capita is not significantly different across industries. This is illustrated in Figure 3, which shows the estimated relationship between expected discoveries and growth (the discovery curves) estimated individually for each of the 16 HS industry groups with discoveries (using the 4-digit data).

Figure 3: Predicted Discoveries by Industry
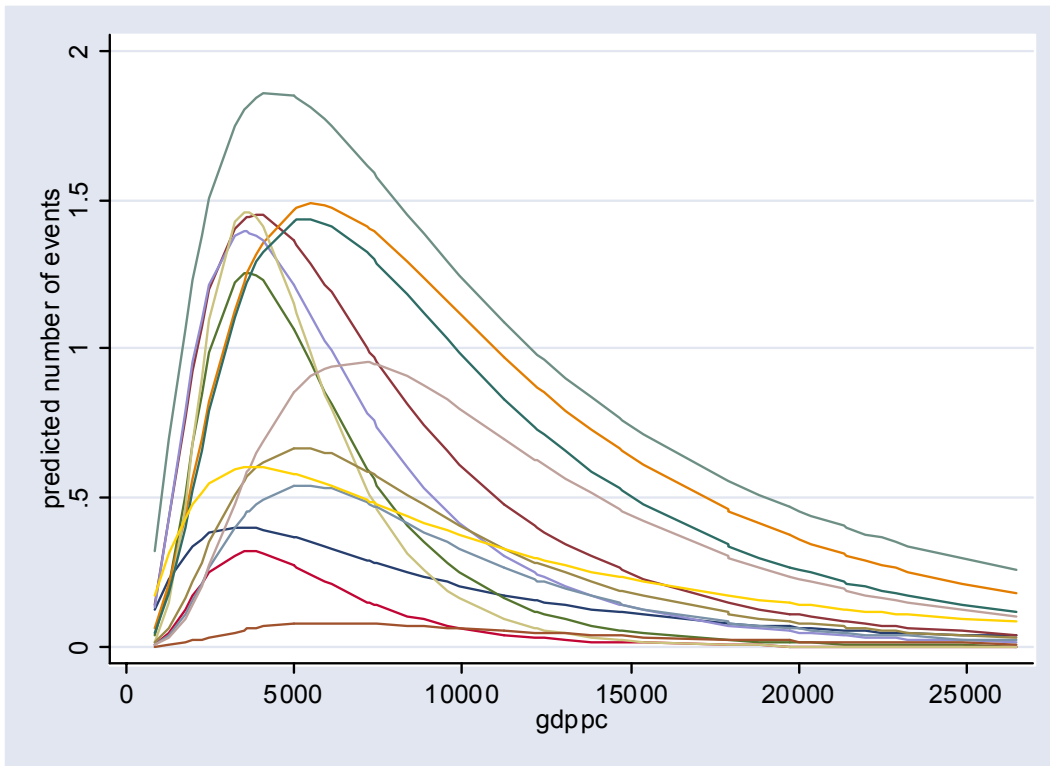
Source: Author's calculations

If discovery activity was driven by the process of structural transformation, the curves for some sectors would peak earlier than those in more capital-intensive sectors. However, we see that this is not the case. While the estimated curves are not uniform, each reaches its peak early in the development process (GDP per capita of $3000 - $6000) and then declines.

One could argue that export data grouped by industry do not clearly indicate the level of technological complexity nor the stage of the production process that countries are involved in, which would be necessary for a fair evaluation of the factor-endowment hypothesis. Export data for manufacturing industries in some cases may simply capture the labor-intensive nature of the assembly processes that are performed in developing countries due to the fragmentation of the production process (Lall 1998, Jones 2000).

To consider this possibility, we performed similar panel regressions, but with goods grouped at their highest level of disaggregation: 4- and 6-digits. Even at the highest level of disaggregation, which resulted in 1232 commodity groups and 65315 observations in the fixed-effects negative binomial panel regression, the coefficient estimates for lnGDP per capita and lnGDP per capita squared were 13.22 and –0.77, which are well within the confidence intervals from the pooled regression. We also performed the same tests with the SITC data, performing a fixed effects negative binomial regression with industry panels at the highest level of disaggregation (3-digits, 64 products per country). Both the 1990s cross-section and the pooled 1970s, 1980s and 1990s data had tipping points extremely similar to those of the pooled estimations reported in Table 5 ($4656 compared to $4765, and $3478 compared to $4278, respectively). These findings are quite robust across data types and time periods.

Therefore, unless one can assert that even at these highly disaggregated levels, the different commodity classifications do not represent goods requiring significantly different factors of production, the conclusion holds: the factor-endowments analysis of production across income levels is not closely related with discovery, and developing countries are not limited to discoveries in certain sectors based on their level of development. We now turn to the third theoretical perspective on the discovery process-- the role of market failures.

6. Discovery and the Role of Market Failures

As discussed in section 2, we have proposed a framework for evaluating the importance of market failures in the discovery process. We will add to the basic model variables that are

expected to directly increase both expected profitability π as well as the ease of imitation α, and evaluate the effect on the frequency of discovery.

This analysis is useful beyond the evaluation of the market failures hypothesis in that policy makers wishing to stimulate discovery activity must have some idea of what policy levers are effective. The relationship between discovery and development discussed in section 4 can be used to determine if a country is over- or under-performing in discovery activity relative to expectations, but it does not offer any policy guidance. For that, we must consider empirical support for the potential drivers of discovery.

Before discussing the variables we use to test for the presence of market failures, we again add to the basic model, this time by including export data, specifically the natural logarithm of 1993 exports and the average annual growth rate of exports between 1993 and 2001. We include these trade measures because, as discussed in section 3, penetrating a foreign market is itself an entrepreneurial event. Therefore, they are potential explanatory variables showing how closely related discovery and overall export growth are, and by extension how similar a discovery-promotion strategy and an export-promotion strategy might be. However, if one does not agree with the viewpoint that the penetration of a foreign market is itself a discovery, then including these export variables remains useful in that it corrects for the fact that we are not using domestic production data.

We consider four groups of explanatory variables to add to the basic model: education, absorptive capacity, ease of entry, and financial system development. Data definitions and

sources can be found in Appendix IV. We add each variable individually, as adding them all to the model simultaneously reduces the sample size to only 36 countries, resulting in extremely few degrees of freedom and a sample largely composed of developed countries.

Education

Higher levels of education lead to a more productive and entrepreneurial workforce, which may increase discovery activity (as well as imitation) over and above the effects of education on GDP. We use measures of average rates of enrollment in tertiary education as well as average educational attainment to evaluate this relationship.

Innovative and Absorptive Capacity

It is believed that national learning and absorptive capacity are functions of spending on R&D (Baumol, Nelson and Wolf 1994). Absorptive capacity makes countries more knowledgeable of what foreign goods they could potentially produce and more able to adapt production to the local context. However, it may also make them more adept at imitating the successes of their fellow nationals. We use a measure of the quantity of scientific and technical articles published in major journals by researchers residing in each country to capture basic scientific and research capacity, as well as the number of patents granted by the U.S. and E.U. agencies, weighted by the amount of commerce directed to these two markets (see Appendix IV for details).

Ease of Entry

Any factor that affects the ease of entering a new business activity would affect profitability π, and also directly affect α, the ease and speed with which copycats can imitate a discovery. We use the World Bank's measure of how difficult it is to start a new business, based on the number of procedures required to complete the process. However, these data are only available for January, 2003. Therefore, this measure is only useful to the extent that these barriers have been relatively persistent over the past decade, which we consider reasonable.

The Financial System

There have been many studies linking financial system development and economic growth (e.g. Beck, Levine and Loayza 2000). However, beyond these economy-wide effects, the financial system directly affects the profitability of discoveries because the cost of financing directly affects the costs of both experimentation and imitation. To consider this relationship, we introduce 1995 private sector credit as a percentage of GDP. As one of the areas singled out by Hausmann and Rodrik (2003b) when considering methods to overcome market failures for economic discovery in El Salvador is through government support of high-risk finance, the relationship between discovery activity and the financial system is even more interesting.

Results

The estimation results using the HS 4-digit and HS 6-digit data are shown in Table 7[9]. This same estimation was performed on the 4-digit and 6-digit data pooled by industry to give

---

[9] Most of the explanatory variables of interest are not available for earlier time periods, therefore we will not use the SITC data in this section. Note that to this point, the SITC data have not behaved differently from the HS data.

another chance to the factor endowments perspective discussed in section 2, but as before, the

results were unaffected by unobserved industry heterogeneity.

Table 7: Investigating Market Failures- GDP Included

	Model I	Model II var1: lnAvgEduc var2: TertEnroll	Model III var1: lnJournals var2: lnPatents	Model IV var1: lnBusStart	Model V var1: PSCredit
7a: HS 4-Digit					
lnGDP	8.45497***	6.93744***	6.12783***	.98569***	8.63670***
	(4.51)	(3.54)	(3.46)	(4.18)	(4.40)
lnGDP squared	-.16187***	-.136567***	-.11368***	-.19695***	-.16363***
	(-4.51)	(-3.70)	(-3.42)	(-4.27)	(-4.37)
lnPopulation	.38661*	.53342**	.27843	.68491***	.38050*
	(1.87)	(2.02)	(1.07)	(2.86)	(1.84)
lnInitialExports	-.48945**	-.43369**	-.34634**	-.32430*	-.60389***
	(-2.55)	(-2.27)	(-2.01)	(-1.65)	(-2.83)
ExportGrowth	5.82809**	2.32391	4.31084	2.32404	6.74057***
	(2.49)	(0.88)	(1.50)	(0.85)	(2.59)
var1		.40268	-.50299**	.10416	.00450
		(0.96)	(-2.01)	(0.88)	(1.32)
var2		.79929	-.50299		
		(-1.06)	(-0.33)		
F-Test: var1 and var2		.4497	.0836*		
Sample Size	49	44	44	42	47
7b: HS 6-Digit					
	Model I	Model II var1: lnAvgEduc var2: TertEnroll	Model III var1: lnJournals var2: lnPatents	Model IV var1: lnBusStart	Model V var1: PSCredit
lnGDP	7.81598***	8.12994***	7.88057***	11.0404***	7.61322***
	(5.84)	(5.99)	(5.37)	(6.98)	(5.61)
lnGDP squared	-.15779***	-.16112***	-.15431***	-.22291***	-.15189***
	(-6.37)	(-6.48)	(-5.88)	(-7.38)	(-6.07)
lnPopulation	.49978***	.42693**	.38255*	.66011***	.44651***
	(3.26)	(2.14)	(1.69)	(4.07)	(2.93)
lnInitialExports	.02533	-.03281	.04210	.12792	-.02375
	(0.19)	(-0.25)	(0.31)	(1.05)	(-0.15)
ExportGrowth	9.26602***	8.66202***	9.69250***	7.30837***	9.29990***
	(5.23)	(4.63)	(4.27)	(3.97)	(4.77)
var1		.32125	-.13946***	.15472**	-.00022
		(0.23)	(4.27)	(2.01)	(-0.09)
var2		-.83852	.81854		
		(-1.57)	(0.77)		
F-Test: var1 and var2		.2816	.2088		
Sample Size	49	44	44	42	47

Note: Parentheses indicate z-statistics. Significant at 10% level: *, 5% level: **, 1% level: *** See Appendix II for sample composition of each estimation. Source: Author's Calculations.

We make five observations based on these results. First, export growth enters in the majority of estimations as positive and significant. This is interesting in that it suggests that a discovery-promotion strategy may have much in common with an export-promotion strategy, if one believes that the penetration of a foreign market is itself a discovery (and if not, including this variable is a necessary correction for the fact that we are using export data rather than domestic production data). This finding could contradict the argument that trade liberalization implies more potential imitators and therefore lower appropriability (q), it is harmful for discovery. To the extent that liberalization increases export activity, our data suggest that it may have a positive relationship with discovery.

The second observation concerns the absorptive capacity variables. Using the 4-digit data, they enter as negative and jointly significant. In addition, the journals measure enters as negative and individually significant with the 6-digit data, although not jointly significant when tested with the patent data. This negative relationship is quite surprising, considering that absorptive capacity was expected to have a positive effect on a country's ability to discover new products, although it might also affect the ease of imitation, thereby reducing discoveries.

The third observation, consistent with the results of the absorptive capacity variables, relates to the barriers to entry variable in the 6-digit data, which enters as positive and significant. That is, higher barriers to entry are associated with an increase in discovery activity. Given that more procedures in the process of starting a new business increases the costs of a first mover, we would expect that in absence of market failures this variable should have a negative relationship with discovery activity. This result, combined with that relating to absorptive

capacity, is evidence in support of the market failures hypothesis. Absorptive capacity should have a positive relationship with π, and barriers to entry a negative relationship. In terms of the model discussed in section 2, we are finding that $\frac{\partial P(D)}{\partial \pi} \leq 0$ and $\frac{\partial P(D)}{\partial \alpha} \leq 0$, which suggests that the appropriability parameter (q) is causing a reduction in discovery activity. That is, as it becomes easier and less complicated for entrepreneurs to start a new business, and as absorptive capacity increases, the effect on the speed and profitability of imitation ($\frac{\partial q}{\partial \pi}$ and $\frac{\partial q}{\partial \alpha}$) dominates the positive effect of π on the first mover's expected profits, and discovery activity *declines*.

The fourth observation relates to the changes in the absorptive capacity variables when moving from 4-digit to 6-digit data. As discussed in section 3, going to a higher level of disaggregation places more importance on manufactured goods. Notice that, going from 4-digits to 6-digits, coefficients on both journal articles and patents (the absorptive, or innovative, capacity variables) rise. This suggests that for manufactured goods, discovery may be positively related to innovative capacity.

Finally, notice that education and private sector credit are not significant in either the 4- or 6-digit data. We would have expected these two variables to have a direct and positive effect on discovery via π. Therefore, this result suggests that either the positive effect on π in the entrepreneur's expected profits is offset by the negative effect of π in the appropriability parameter q (that is, the market failures hypothesis), or statistical limitations are preventing us from estimating the true relationship, likely due to multicolinearity between these two variables and GDP (see Appendix IV for correlation coefficients). These variables may have indirect

effects on discovery activity via overall development, or they may be drivers of both development and discovery.

We analyzed this issue further by testing how education, journals, patents, and barriers to entry entered into a model that included only initial exports and export growth. Due to the quadratic relationship between discoveries and GDP per capita, and the high correlations between these variables and development, we also added the square of each term. The results are shown in Table 8.

Table 8: Investigating Market Failures- GDP Per Capita Excluded

8a: HS 4-Digit				
	Model VI newvar=lnAvgEduc	Model VII newvar=lnJournals	Model VIII newvar=lnPatents	Model IX newvar=lnBusStart
lnInitial Exports	-.09085 (-1.13)	.03394 (0.28)	-.03446 (-0.44)	-.20165** (-2.43)
Export Growth	7.54603** (2.34)	5.88501* (1.77)	7.15425** (2.17)	6.86730** (2.00)
newvar	2.32678 (0.55)	.45839* (1.68)	-16.87737*** (-2.82)	-.32835 (-0.65)
newvar squared	-.90031 (-0.80)	-.05229*** (-2.57)	45.94293** (2.06)	.11093 (1.40)
F-Test: newvar and newvar squared	.0120**	.0019***	.0009***	.0039***
Sample Size	46	47	46	45
8b: HS 6-Digit				
	Model VI newvar=lnAvgEduc	Model VII newvar=lnJournals	Model VIII newvar=lnPatents	Model IX newvar=lnBusStart
lnInitial Exports	.19302*** (2.74)	.27729*** (2.80)	.21564*** (2.79)	.09708 (1.29)
Export Growth	12.29052*** (4.69)	10.85532*** (3.93)	13.33104*** (4.39)	10.53574*** (3.71)
newvar	3.56511 (1.03)	.60275*** (2.98)	-10.42530** (-2.05)	.08300 (0.21)
newvar squared	-1.22028 (-1.31)	-.05993 (-4.13)	28.91349 (1.48)	.04576 (0.72)
F-Test: newvar and newvar squared	.0013***	.0000***	.0165**	.0006***
Sample Size	46	47	46	45

Note: Parentheses indicate z-statistics. Significant at 10% level: *, 5% level: **, 1% level: *** See Appendix II for sample composition of each estimation. Source: Author's Calculations

The education and journals variables follow the same pattern as GDP, positive at very low levels of development, peaking at a level of educational attainment/journal publication rate typical of a country with GDP per capita around $4000, and then declining. The patent and barriers to entry variables are more interesting. For developing countries in the sample (levels of income per capita less than $14,000), the patents variable has a negative relationship with discovery, and the barriers to entry has a positive relationship. This is consistent with the results above, in that even when GDP is excluded, variables that increase both π and α are having the opposite effect on discovery activity, suggesting that their negative effects via the appropriability parameter are outweighting their positive effects on first-mover profits.

The main conclusions arising from these estimations are therefore that discovery activity appears to have a positive relationship with export growth, and by extension policies that increase exports, that discovery in differentiated products may have a positive relationship with the traditional concepts of technological innovation, and that market failures arising from imitation may inhibit the frequency of discovery.

7. Summary and Directions for Future Research

We began this investigation from the viewpoint that discussions of the emergence of new production, or discoveries, are not informed by systematic empirical analysis. We attempt to fill this void by studying the empirical properties of discovery activity through three theoretical perspectives. This is accomplished by using worldwide disaggregated export data.

We found that the discoveries of successful new exports have not been confined to those sectors that researchers label as 'dynamic' (Lall 1998, Butkevicius Kadri and Mayer 2002), but are also common in sectors such as agriculture. Discoveries occur in the context of economy-wide diversification, perhaps driven by international trade as argued by Imbs and Wacziarg, and there is therefore a certain level of discovery activity associated with the level of development of any economy. Expanding on this, we analyzed patterns of over- and under-performance in discovery activity as compared to this expectation. We also considered discovery in the process of the structural transformation of economies based on their factor-endowments, but this theoretical lens turned out to be less useful for analyzing economic discovery. Finally, we considered possible drivers of discovery, and find support for the hypothesis that market failures are inhibitors of discovery.

Policy makers now have some empirical evidence to combine with theoretical models of market failures and the emergence of new production. Given the size and sophistication of a particular economy, the basic model can be used to determine if there is over- or under-performance in discovery activity. Furthermore, the evidence of discoveries in a broad range of sectors suggests that policy makers seeking to increase the level of discovery activity in their economies need not target a narrow set of fad sectors (consistent with a similar recommendation by Hausmann and Rodrik 2003b). However, our findings do not suggest an obvious channel through which governments can stimulate discovery activity, beyond a positive relationship with export growth. Furthermore, it appears that imitation could be inhibiting the discovery process, which if true supports policies to either reduce the costs of experimentation or increase the appropriability of successful discoveries.

It is clear that, while the exploration above does solidify our empirical understanding of economic discovery, further research is necessary. The importance of free-riding in the discovery process must be examined further. It would also be useful to know through which channels the free-riding problem is most acute: the discovery of production costs, the investments to penetrate foreign markets, or elsewhere. Finally, the link between a higher frequency of discovery activity and subsequent growth has not been shown empirically, and is outside the scope of this paper. We do not know if discovery activity simply occurs with economic growth, or if it is a driver of subsequent growth. This connection is of obvious importance, and merits further study.

10. References

Aten, B., A. Heston and R. Summers. 2002. Penn World Table Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP).

Barro, Robert J. and Jong-Wha Lee. 2001. "International Data on Educational Attainment: Updates and Implications." *Oxford Economic Papers* 2001; 53: 541-563.

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.

Baumol, W.J., R.R. Nelson and E.N. Wolff. 1994. The Convergence of Productivity, Its Significance and Its Varied Connotations. Oxford: Oxford University Press.

Beck, T., R. Levine and N. Loayza. 2000. "Finance and the sources of growth." *Journal of Financial Economics*. 58(1-2): 261-300.

Butkevicius, A., A. Kadri and J. Mayer. 2002. "Dynamic Products in World Exports." United Nations Conference on Trade and Development Discussion Paper No. 159. Geneva: UNCTAD.

FUSADES (Fundación Salvadoreña para el Desarrollo Económico y Social). 2004. "Cuarta estrategia económica y social 2004-2009 oportunidades, seguridad y legitimidad: bases para el desarrollo." El Salvador: FUSADES.

Ganslandt, M and J.R. Markusen. 2000. Standards and related regulations in international trade: A modeling approach. *Quantifying the impact of technical barriers to trade: Can it be done?* 95-135. Studies in International Economics, Ann Arbor: University of Michigan Press.

Grubel, Herbert G. and P.J. Lloyd. 1975. Intra-Industry Trade: The Theory and Measurement of International Trade in Differentiated Products. New York: John Wiley & Sons.

Gwartney, James and Robert Lawson (2004). "Economic freedom of the world: 2004 annual report." Vancouver: The Fraser Institute. Data retrieved from www.freetheworld.com.

Hausman, Jerry, Bronwyn H. Hall, and Zvi Griliches. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica* 52(4): 909-938.

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." Mimeo. Kennedy School of Government, Harvard University.

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." *The American Economic Review*. 93(1): 63-86.

Jones, Ronald W. 2000. Globalization and the Theory of Input Trade. MIT Press: Cambridge, MA.

Lall, S. 1998. "Exports of manufactures by developing countries: Emerging patterns of trade and location." *Oxford Review of Economic Policy*. 14(2): 54-73.

Lall, S. 2000. "The technological structure and performance of developing country manufactured exports, 1985-1998." *Oxford Development Studies*. 28(3): 337-369.

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

Lederman, Daniel and Laura Saenz. 2003. "Innovation around the World: A Cross-Country Data Base of Innovation Indicators." Mimeographed. Office of the Chief Economist for Latin America and the Caribbean, World Bank, Washington DC.

Mayer, W. 1984. "The infant-export industry argument." *Canadian Journal of Economics*. 17(2): 249-269.

Neary, J. Peter. 2003. "Competitive versus comparative advantage." *The World Economy*. 26(4): 457-470.

UNCTAD. 2003. "UNCTAD handbook of statistics." TD/STAT.28. Geneva: UNCTAD.

Vettas, Nikolaos. 2000. "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." Rapid Response Unit, World Bank, Washington D.C.

Appendix I

Graphical Representation of Analytical Framework of Section 2

As described in the text, the basic model is as follows:

$$(6)\ P(D) = D(\pi \times q(\pi, \alpha)).$$
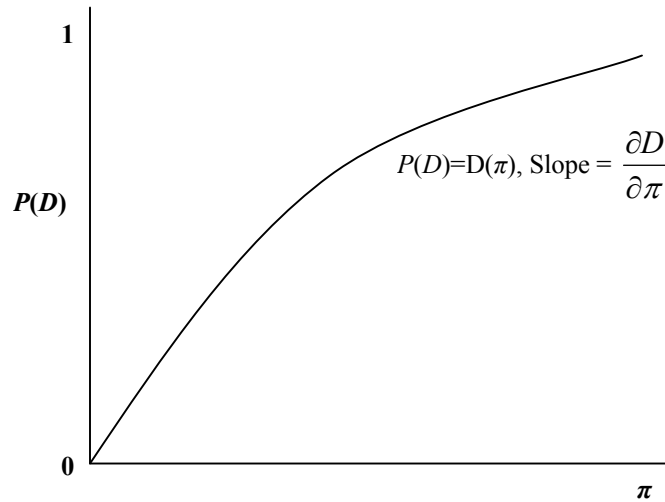$$\quad\quad\quad\quad\quad\quad {}_{+}\quad {}_{-}\ {}_{-}$$

The derivatives with respect to the two parameters we consider are therefore

$$(7)\ \frac{\partial P(D)}{\partial \pi} = \frac{\partial D}{\partial \pi}\left[q(\pi, \alpha) + \pi \times \frac{\partial q}{\partial \pi}\right] \text{ and}$$

$$(8)\ \frac{\partial P(D)}{\partial \alpha} = \frac{\partial D}{\partial q} \times \frac{\partial q}{\partial \alpha},$$

where by construction $\dfrac{\partial q}{\partial \pi} \leq 0,\ \dfrac{\partial q}{\partial \alpha} \leq 0$, and $0 \leq q \leq 1$.

In the absence of market failures, $q = 1$, and the relationship between π and the probability of observing a discovery is illustrated below:

However, now consider that there is a threat of imitation. First, we consider the case where appropriability is only based on the ease of imitation α, and is not a function of expected profits. The relationship between π and the probability of observing a discovery, where the introduction of α causes q to be less than 1, is illustrated below:

42

$$P(D) = D(\pi \times q(\alpha)), \text{ Slope} = \frac{\partial D}{\partial \pi} \times q(\alpha), \ q(\alpha) \prec 1$$
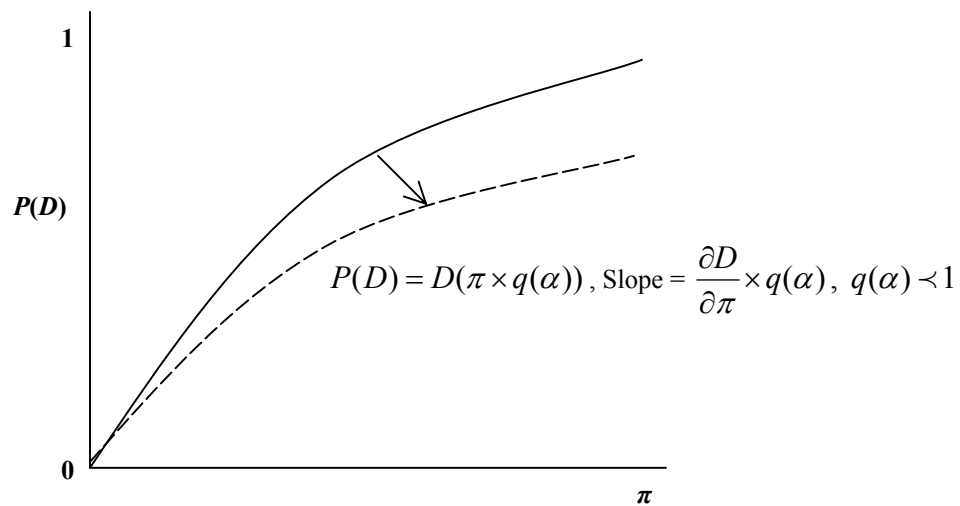
Now consider the full model, where appropriability a function of expected profits π and ease of imitation α. It is now possible to observe a negative relationship between expected profitability and the probability of observing a discovery of a particular good. The relationship between π and the probability of observing a discovery is illustrated below:

$$P(D) = D(\pi \times q(\alpha, \pi)), \text{ Slope} = \frac{\partial D}{\partial \pi}\left[q(\pi, \alpha) + \pi \times \frac{\partial q}{\partial \pi} \right]$$
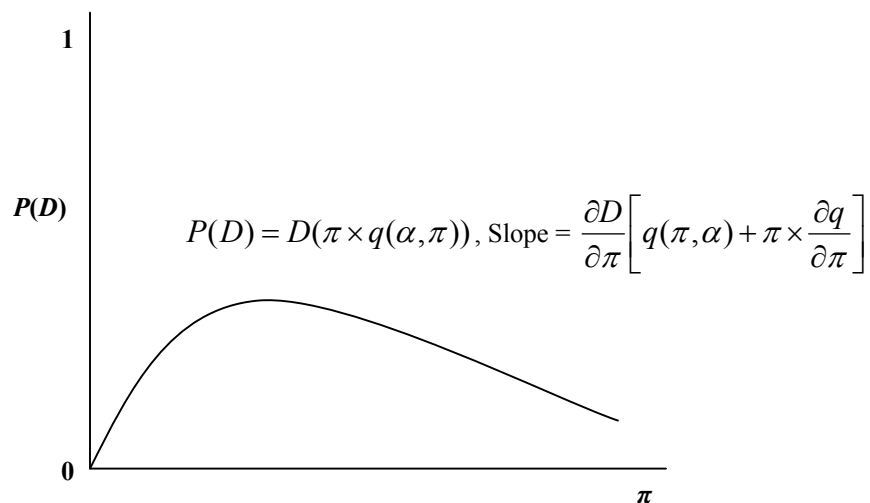
Appendix II

Sample Composition

We used all countries that reported the data in the years required by the filters, with two exceptions. First, the United Arab Emirates and Tunisia had to be dropped from the HS samples, as their HS reported data implied export growth in the 1990s significantly different from national accounts. This was not the case for all other countries in the sample. Second, we dropped states from both the HS and SITC sample with populations less than 200,000 in 1992, as the lessons for these microstates are not generalizable to most countries. This left us with 53 countries in the HS samples and 99 (for various time periods) in the SITC sample.

Countries in discovery sample: 1990s HS Data

Argentina	United Kingdom	Netherlands
Australia	Greece	Norway
Burundi[3,4,5,7,8,9,10]	Guatemala	New Zealand
Bolivia	Hong Kong, China	Oman[1,2,3,4,5,6,7]
Brazil	Croatia[1,2,3,4,5,6,8,9]	Peru[2,3,4,5,6,7,8,9,10]
Central African Republic[3,4,5,7,8,9,10]	Hungary	Portugal
Canada	Indonesia	Paraguay
Switzerland	India	Romania[6]
Chile	Ireland	Saudi Arabia[1,2,3,4,5,6,7]
China	Iceland[5,10]	Singapore
Colombia	Japan	Sweden
Cyprus[5,10]	Korea, Rep.	Thailand
Czech Republic[4,8,9]	Macao[3,4,5,7,8,9,10]	Trinidad and Tobago[5,10]
Germany	Morocco[3,7]	Turkey
Denmark	Mexico	Taiwan, China[3,6]
Ecuador	Mauritius[5,10]	United States
Spain	Malaysia	South Africa[4,9]
Finland	Nicaragua	

Composition of Samples in Estimations

Table 4: Countries not entering sample with Filter 1 denoted by 1, Filter 2 denoted by 2.
Table 5: Countries not entering sample denoted by 1.
Table 6: Countries not entering sample denoted by 1.
Table 7: Countries not entering Model I denoted by 2, Model II denoted by 3, Model III denoted by 4, Model IV denoted by 5, Model V denoted by 6.
Table 8: Countries not entering Model VI denoted by 7, Model VII denoted by 8, Model VIII denoted by 9, Model IX denoted by 10.

Countries in discovery sample: SITC Data

Country	Periods	Country	Periods
Algeria	1974-1988	Madagascar	1973-1974, 1980-1986, 1991-2002
Argentina	1973-2002	Malawi	1973-1979, 1984-1989
Australia	1973-2002	Malaysia	1973-2002
Austria	1973-2002	Malta	1973-1989
Bahamas, The	1975-1976, 1987-1988	Martinique	1973-1983
Bahrain	1973-1984, 1990-2002	Mauritius	1973-1978, 1981-2002
Bangladesh	1978-1986	Mexico	1973-2002

Barbados	1973-2002	Morocco	1973-2002
Belgium-Luxembourg	1973-1986	Nepal	1975-1988
Bolivia	1973-2002	Netherlands	1973-2002
Brazil	1973-2002	New Zealand	1973-2002
Brunei	1973-1982	Nicaragua	1973-1974, 1978-1986, 1988-2002
Cameroon	1990-2002	Nigeria	1973-1975, 1986-1987
Canada	1973-2002	Norway	1973-2002
Chile	1973-2002	Oman	1980-2002
China	1988-2002	Pakistan	1973-2002
Colombia	1973-2002	Panama	1973-2002
Costa Rica	1973-2002	Papua New Guinea	1973-1976
Cyprus	1973-2002	Paraguay	1973-2002
Denmark	1973-2002	Peru	1973-2002
Dominican Republic	1975-1976, 1982-1983, 1986-1988	Philippines	1973-2002
Ecuador	1973-2002	Poland	1981-2002
Egypt, Arab Rep.	1973-2002	Portugal	1973-2002
El Salvador	1973-2002	Qatar	1979, 1990-2002
Ethiopia	1973-1980	Reunion	1973-1983
Fiji	1973-1982, 1990-2002	Romania	1990-2002
Finland	1973-2002	Saudi Arabia	1978-1982, 1989-2002
France	1973-2002	Senegal	1979-1981, 1987, 1990-2002
Germany	1973-2002	Singapore	1973-2002
Greece	1973-2002	South Africa	1982-1984
Greenland	1977-2002	Spain	1973-2002
Guadeloupe	1973-1983	Sri Lanka	1975-1982
Guatemala	1973-2002	Sudan	1982, 1985
Honduras	1973-2002	Sweden	1973-2002
Hong Kong, China	1973-2002	Switzerland	1973-2002
Hungary	1982-1987	Syrian Arab Republic	1975, 1985-1987
Iceland	1973-2002	Taiwan, China	1973-2002
India	1973-2002	Thailand	1973-1989
Indonesia	1973-2002	Togo	1976-1979, 1987-2002
Ireland	1973-2002	Trinidad and Tobago	1973-2002
Israel	1973-2002	Tunisia	1973-2002
Italy	1973-2002	Turkey	1973-2002
Jamaica	1973-1988	United Arab Emirates	1979, 1989
Japan	1973-2002	United Kingdom	1973-2002
Jordan	1973-1983, 1987-2002	United States	1973-2002
Kenya	1980-1988	Uruguay	1975-2002
Korea, Rep.	1973-2002	Venezuela	1973-2002
Kuwait	1976-1984, 1987	Yugoslavia, FR	1973-1978
Libya	1973-1979	Zimbabwe	1985, 1990-2002
Macao	1974-2002		

Appendix III

The Basic Model

The following are the negative-binomial regression results for the basic model used in section 4 to analyze regional performance in discovery activity:

$$(12) \quad \lambda = e^{\beta_0 + \beta_1(\ln GDP) + \beta_2(\ln GDP)^2 + \beta_3(\ln Population)}$$

	HS 4-Digit 1990s		HS 6-Digit 1990s		SITC 3-Digit	
	Filter 1	**Filter 2**	**Filter 1**	**Filter 2**	**1990s**	**1972-1991 (fe)**
Coefficient: ln(GDP)	8.09771 (4.39)	6.2732 (2.64)	9.20869 (6.85)	8.40898 (6.21)	9.90686 (2.82)	5.73080 (2.68)
Coefficient: ln(GDP) squared	-.17242 (-4.67)	-.13250 (-2.78)	-1.8515 (-6.99)	-.16796 (-6.25)	-.21151 (-3.01)	-.11761 (-2.80)
Coefficient: ln(Population)	.85035 (6.15)	.18406 (3.44)	.54233 (4.78)	.40452 (3.81)	.71287 (3.00)	.43466 (2.09)
Pseudo R-Squared	.1238	.0741	.1113	.1120	.0906	
Discovery Count	332	150	1710	865	93	1114
Sample Size	50	49	50	49	67	76

Note: brackets indicate z-statistics. Composition of HS samples is equivalent to Table 4 (see Appendix II). Source: Author's calculations

We see that estimates using the second filter are well within the 95% confidence intervals of estimates using the first filter. In addition, estimates are quite similar across data sources.

Appendix IV

Data Definitions, Sources, and Correlations

Variable Name	Description	Units	Year(s) Used	Transformation	Source
ln(GDP Per Capita)	Natural log of real GDP per capita (PPP)	1996 PPP Constant Prices (Chain Series)	1992	log	PWT 6.1 (Aten Heston and Summers 2002)
ln(Population)	Natural log of population	Thousands of People	1992	log	PWT 6.1 (Aten Heston and Summers 2002)
ln(GDP)	Natural log of real GDP (PPP)	1996 PPP Constant Prices (Chain Series)	1992	Calculated from PWT as GDP Per Capita * Population * 1000	PWT 6.1 (Aten Heston and Summers 2002)
ln(AvgEduc)	Average years of education of the populatioin.	Years	1995	log	Barro & Lee (2001)
TertEnroll	Average tertiary enrollment rate	Ratio	1992-1996 Average	none	World Development Indicators (World Bank)
ln(Journals)	Natural log of Scientific Journal Articles Published by Nationals	Count	1992	log	Lederman and Saenz (2003)
ln(Patents)	Natural log of trade-weighted patents in US and EU	Patents per million dollars in related commerce	1992	*	Patent counts: Lederman and Saenz (2003). Weighted counts: Author's Calculations
ln(BusStart)	Number of procedures required to legally start a new business.	Procedure count	2003	log	World Bank (2004)
ln(Initial Exports)	Natural log of Total Exports	Current US$	1993	log	COMTRADE
Export Growth	Average annual growth rate of exports	Percent (in decimal form)	1993-2001	[ln(total exports 2001)-ln(total exports 1993)]/8	COMTRADE
PSCredit	Credit to the private sector relative to GDP	ratio	1995	none	World Development Indicators (World Bank)

*Patents were weighted as follows. For countries other than the US and EU, the measure is the number patents granted to nationals of the country in the U.S. divided by its exports to the U.S. plus the number of patents granted to nationals of the country in the E.U. divided by its exports to the E.U. For the U.S., its patents in its own market were divided by total domestic commerce (Non-services GDP minus total exports), and then added to E.U. patents divided by its exports to the E.U. For E.U. countries, patents in the U.S. were divided by exports to the U.S. and added to exports in the E.U. divided by total E.U. commerce (Non-services GDP minus total exports plus exports to other E.U. countries). The resulting measure weighs patenting activity by total commerce in the relevant market.

Correlations Among Explanatory Variables

	ln(GDP Per Capita)	ln(AvgEduc)	TertEnroll	ln(Journals)	ln(Patents)	ln(Bus Start)	ln(Initial Exports)	Export Growth	PSCredit
ln(GDP Per Capita)	1								
ln(AvgEduc)	0.7736	1							
TertEnroll	0.698	0.7033	1						
ln(Journals)	0.5572	0.5578	0.6634	1					
ln(Patents)	0.5367	0.5019	0.4598	0.5804	1				
ln(Bus Start)	-0.62	-0.6451	-0.6724	-0.4603	-0.2401	1			
ln(Initial Exports)	0.5862	0.4458	0.5207	0.878	0.4695	-0.3482	1		
Export Growth	0.082	-0.1283	-0.1456	-0.0011	-0.0977	0.231	0.3037	1	
PSCredit	0.559	0.475	0.2519	0.4285	0.4355	-0.3648	0.5467	-0.1472	1