EXPORT DIVERSIFICATION: WHAT'S BEHIND THE HUMP?

Olivier Cadot, Céline Carrère, and Vanessa Strauss-Kahn*

Abstract—The paper explores the evolution of export diversification patterns along the economic development path. Using a large database with 156 countries over 19 years at the HS6 level of disaggregation (4,991 product lines), we look for action at the intensive and extensive margins. We find a hump-shaped pattern of export diversification similar to what Imbs and Wacziarg (2003) found for production. Diversification and subsequent reconcentration take place mostly along the extensive margin. This hump-shaped pattern is consistent with the conjecture that countries travel across diversification cones, as discussed in Schott (2003, 2004) and Xiang (2007).

I. Introduction

WHY should export diversification be taken as a policy objective per se? There are two reasons that it should *not*. First, according to Ricardo, countries should specialize, not diversify. Second, the Heckscher-Ohlin model implies that export patterns are largely determined by endowments, so, if anything, we should worry about factor accumulation, not diversification. Yet export diversification is a constant preoccupation of policymakers in developing countries. As de Ferranti et al. (2002) note, "A recurrent preoccupation of [Latin American] policymakers is that their natural riches produce a highly concentrated structure of export revenues, which then leads to economic volatility and lower growth" (p. 38).

The notion that export patterns are fully determined by endowments is of course naive. The relationship of endowments, trade, and growth is a complex and imperfectly understood one. Intra industry trade models showed long ago that many factors other than endowments, including market failures and policies, can affect trade patterns. More recently, Hausmann, Hwang, and Rodrik (2007) argued that export patterns can display path dependence in the presence of externalities.

Policy concerns about a linkage between the concentration of exports on primary products and deteriorating terms of trade, income volatility and, ultimately, low growth go back to the work of Prebisch (1950) and Singer (1950). Subsequent work (for example, Neary & van Wijnbergen, 1986; Gelb, 1988; Auty, 1990; Sachs & Warner, 1999) showed a robustly negative correlation between dependence on primary products and future growth, a finding called the "natural-resource curse."[1] The negative correlation between natural resources and growth was, however, questioned by, among others, Brunnschweiler (2008) and Brunnschweiler and Bulte (2008), who argued that regressing growth on the share of primary products in exports or GDP suffered from fatal endogeneity problems.

While the relationship of endowments, trade, and growth has remained a controversial issue, how export patterns vary across time and countries has become a subject of intense descriptive analysis in recent years. Several papers (for example, Evenett & Venables, 2002; Hummels & Klenow, 2005; Kehoe & Ruhl, 2006; Brenton & Newfarmer, 2007) decompose cross-country export variations into intensive and extensive (new-products or new-markets) margins and study the contribution of these margins in export growth.[2] Digging deeper into the extensive margin, Hausmann and Klinger (2006) proposed a measure of "product proximity" based on the conditional probability that one product is exported given that the other is also exported.

In parallel with this literature, a widely cited paper by Imbs and Wacziarg (2003) uncovered a nonmonotone path of production and employment diversification as functions of per capita incomes, with diversification followed by reconcentration. Imbs and Wacziarg's work naturally raised the question of whether a similar pattern would hold for exports as well. Klinger and Lederman (2004, 2006) indeed found that exports diversify and then reconcentrate with income. While Imbs and Wacziarg's exercise was essentially an empirical one, Klinger and Lederman built on Hausmann and Rodrik (2003) to explore a causal link from market failures to insufficient diversification. The argument is that opening up new export lines is an entrepreneurial gamble; if it is successful, it is quickly imitated. The inability of "export entrepreneurs" to keep private the benefits of their activity thus leads to a classic public-good problem.

We revisit the issue using a different perspective, in which we derive and analyze a decomposition of Theil's concentration index that maps directly into the extensive and intensive margins of export diversification. In order to analyze how the two margins evolve as functions of GDP per capita, we construct a very large database covering 156 countries (including 141 developing ones) over all years available from the COMTRADE database at the highest disaggregation level (HS6). Using this database, we calculate for all countries and years three classes of variables of interest: export concentration indices (focusing on Theil's index and its decomposition), the number of active lines (lines with nonzero exports), and a measure of "new export

Received for publication December 7, 2007. Revision accepted for publication January 6, 2010.

* Cadot: World Bank, HEC Lausanne, CERDI, CEPR and CEPREMAP; Carrère: CERDI-CNRS, Université d'Auvergne; Strauss-Kahn: ESCP-Europe and CEPII.

Research on this paper was supported by a grant from the Interamerican Development Bank and by Switzerland's Fonds National pour la Recherche Scientifique. We thank the FERDI (Fondation pour les Etudes et Recherches sur le Développement International) for publication support. Special thanks go to Julien Gourdon for calling our attention to key data and estimation issues. We also thank Marius Brülhart, Antoni Estevadeordal, Christopher Grigoriou, Jaime de Melo, Marcelo Olarreaga, Christian Volpe, two anonymous referees, and the editor for useful comments.

[1] The Prebisch-Singer hypothesis implies that low growth is caused by dependence on primary products, not necessarily by concentration per se. However, preliminary findings by Dutt, Mihov and van Zandt (2008) suggest that diversification does accelerate future growth, especially when it is accompanied by convergence toward the U.S. pattern of exports.

[2] The intensive margin reflects variation in export values among existing exports, whereas the extensive margin reflects variation in the number of new products exported or in the number of new markets for existing exports.

The Review of Economics and Statistics, May 2011, 93(2): 590–605

products." We use these three variables to explore action along the intensive and extensive margins. In essence, we propose a decomposition of the Theil index in between-groups and within-groups components that can be easily mapped into the extensive and intensive margins, respectively.

We find a hump-shaped relationship between economic development and export diversification, like Imbs-Wacziarg and Klinger-Lederman, with a turning point around $25,000 per capita at purchasing power parity (PPP). The observed reconcentration might be spurious in a number of ways. For instance, it could be driven by small, rich, and concentrated oil producers. It could also be an artifact of the Harmonized system (see Appendix A). This would be the case if low- and middle-income countries were mainly exporting products from sectors with large numbers of export lines such as the textile sector. Alternatively, observed concentration pattern could be driven by unexplained heterogeneity between countries. We find that none of the obvious culprits stands scrutiny. In particular, the reconcentration holds strongly within country: all countries to the right of the turning point reconcentrate over time.

At income levels below the turning point, we find diversification at both the extensive and intensive margins, but mostly along the extensive margin until around PPP $22,000. The intensive margin briefly dominates around the turning point; thereafter, the extensive margin retakes the lead and explains the reconcentration, suggesting that rich countries close export lines. What are those products disappearing from rich-country export portfolios? We find that the factor intensities of those products are typically far away from the countries' endowments, as if they were left-overs from old export patterns kept alive only by hysteresis. That is, our evidence suggests that as countries travel across diversification cones, they fail to close a tail of export lines that no longer belong to their comparative advantage but artificially inflate their diversification, until finally comparative advantage catches up.

The paper is organized as follows. Section II reports econometric evidence on the stages of export diversification in the process of economic development. In order to better understand what is behind the hump-shaped diversification curve, section III analyzes action along the intensive and extensive margins by examining the evolution of the within and between component of the Theil concentration index. It also explores the specificities of the new export products that generate diversification. Section IV explores potential explanations behind the diversification curve. Section V concludes.

II. Stages of Diversification: Estimation

A. *Measures of Export Concentration/Diversification*

Our dataset comprises data on trade and income per capita. Export data are from UNCTAD's COMTRADE database at the HS6 level (4,991 lines).[3] The baseline sample covers 156 countries representing all regions and all levels of development between 1988 and 2006 (19 years), including 141 developing countries—non-high-income countries, defined by the World Bank as countries with 2006 per capita GDP under $16,000 in constant 2005 PPP international dollars. After we take out missing-year data, the usable sample has 2,797 observations (country-years).

In this section, we compute several measures of export concentration/diversification for each country and year: Herfindahl concentration indices, Theil and Gini indices of inequality in export shares, and the number of active export lines. The Herfindahl index, normalized to range between 0 and 1, is

$$H^* = \frac{\sum_k (s_k)^2 - 1/n}{1 - 1/n},$$

where $s_k = x_k / \sum_{k=1}^{n} x_k$ is the share of export line k (with amount exported x_k) in total exports and n is the number of export lines (omitting country and time subscripts). We use the following formula for the Gini index:

$$G = 1 - \sum_{k=1}^{n} (X_k - X_{k-1})/n,$$

where $X_k = \sum_{l=1}^{k} s_l$ represents the cumulative export shares. Theil's entropy index (Theil, 1972) is given by

$$T = \frac{1}{n}\sum_{k=1}^{n} \frac{x_k}{\mu} \ln\left(\frac{x_k}{\mu}\right) \quad \text{where} \quad \mu = \frac{1}{n}\sum_{k=1}^{n} x_k. \tag{1}$$

Table 1 shows descriptive statistics for these indices.

Observe that Gini indices are very high. The reason has to do with the level of disaggregation: we use a very disaggregated trade nomenclature. At that level, we have a large number of product lines with small trade values, while a relatively limited number of them account for the bulk of all countries' trade (especially for developing countries, but even for industrial ones). As for the average number of positive export lines—active lines with non zero trade values—it is relatively low at 2,062 per country per year—a little less than half the total, with a minimum of 8 for Kiribati in 1993 and a maximum of 4,988 for Germany in 1994 and the United States in 1995. This implies that there is room for a substantial extensive margin for developing countries, especially the poorest and least diversified ones.

Per capita GDPs are taken from the World Bank's World Development Indicators (WDI) and are expressed in 2005 purchasing power parity (PPP) dollars for comparability.

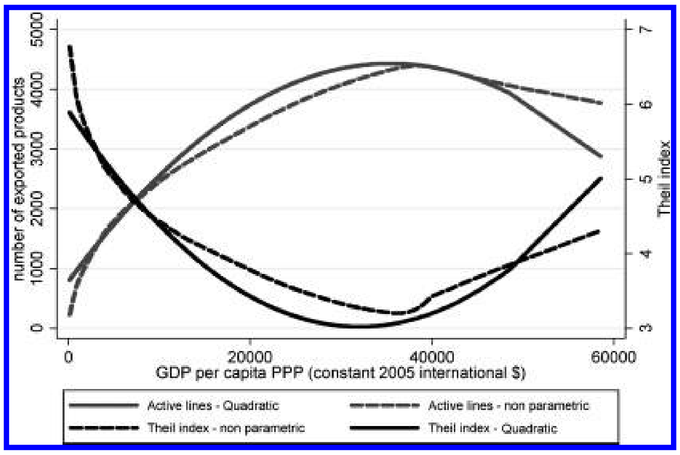
[3] The appendix provides further information on the COMTRADE HS6 level database.

Table 1.—Descriptive Statistics: 156 Countries over 1988–2006

Variable	Observation	Mean	S. D.	Minimum	Maximum
Gini	2,797	0.967	0.045	0.773	1.000
Herfindahl	2,797	0.189	0.235	0.002	0.989
Theil	2,797	4.865	1.797	1.478	8.465
Nber of active lines	2,797	2,061.8	1,669.6	8	4,988
GDPpc PPP, constant international 2005 dollars	2,695	9,442.1	11,130.9	136.5	73,276.9
Share of oil in exports	2,797	0.190	0.287	0	0.996

Author calculations using COMTRADE.

Figure 1.—Predicted Theil's Concentration Index & Number of Active Export Lines



Note:
Quadratic corresponds to the OLS estimation of $Y_{it} = \alpha_0 + \alpha_1 GDPpc_{it} + \alpha_2 GDPpc_{it}^2 + v_{it}$, with Y_{it} being alternatively the Theil index and the number of active export lines. *Nonparametric* corresponds to smoother nonparametric regressions of Y_{it} on the *GDPpc*. Author calculations using COMTRADE.

B. *Parametric Evidence*

Figure 1 depicts curves representing predicted values of Theil's index, as well as curves representing the predicted number of active export lines.[4] The latter, which are concave and increasing at the origin, are easy to distinguish from the former, which are convex and decreasing at the origin.

The Theil curve is fitted using quadratic polynomial regressions of the Theil concentration index on per capita GDP using pooled OLS with White-corrected standard errors. We find a turning point around $30,000 in PPP (2005 constant).[5] We also estimated smoother nonparametric regressions (dashed curves). This consists of reestimating the regression for overlapping samples centered on each observation. Smoother regressions impose no functional form and are therefore suited to the exploration of highly nonlinear relationships. The nonparametric estimates validate the use of a quadratic form to approximate the relation between the export concentration and per capita GDP (figure 1).

One issue is whether the turning point is driven by microstates and island economies, which are very heterogeneous in GDP per capita and at the same time very concentrated—say, in bananas or fish products. Because microstates are potential outliers, we omit them in the rest of the analysis (that is, we exclude fifteen countries with populations below 1 million).

A second issue is that of omitted variables. First, spurious correlation could be introduced by fluctuations in the world price of oil and other commodities, as higher commodity prices would raise both per capita incomes and export concentration for primary-product exporters. Columns 1 to 4 of table 2, which report pooled estimates with time effects, show a turning point around 25,000 PPP international (2005 constant) dollars. This turning point is quite similar to the one found by Imbs and Wacziarg for production and by Klinger and Lederman (2004) for exports on a panel of 130 countries from 1992 to 2003 ($22,500 in constant 2000 dollars).[6]

Second, given the panel structure of our data set, a natural question is the type of estimator—within, between, or pooled—we should use. Imbs and Wacziarg's estimation on production data relies on fixed effects (that is, within). Columns 5 to 12 of table 2 show our results using the within and between estimators. The turning point stays significant and at a similar level of GDP per capita. Apart from its level, what matters is which countries are on either side of the turning point. When Theil regressions are used the between and pooled estimators return the same list of 21 countries to the right of the turning point. The within estimator adds only 2 (Israel and New Zealand).[7]

Table 3 reports a number of robustness checks. First, we consider censoring, as Gini coefficients are bounded left and right, at 0 and 1, respectively, although neither is binding the strictest sense. We thus perform a logistics transfor-

[4] Fitted curves for Herfindahl and Gini indices have similar shapes.

[5] We also explore the turning point's stability across different definitions of GDP per capita (i.e., per capita GDP at PPP from the Penn World Tables and per capita GDP in constant US dollars from the WDI). Results, which are similar across definitions, are available on request.

[6] The value of our turning point is not directly comparable to that of Imbs and Wacziarg, as they used Summers-Heston per capita incomes in constant 1985 dollars. They note, however, that their turning point occurs roughly at the level of income reached by Ireland in 1992. Our turning point corresponds roughly to Ireland's income level in 1996.

[7] Measurement errors in explanatory variables, if they are correlated with the error term, create a downward bias in estimated coefficients that is especially severe with fixed effects (see Griliches and Hausman, 1986). If present, this would push the turning point to the left compared to pooled and between estimates.

TABLE 2.—POOLED, WITHIN, AND BETWEEN ESTIMATES

	Pooled				Within				Between			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependent	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber
GDPpc	−1.89E-05	−0.0002516	−5.98E-06	2.65E-01	−6.50E-06	−0.0000779	−2.63E-06	3.90E-01	−1.89E-05	−0.0002573	−5.84E-06	2.68E-01
	12.48***	23.40***	21.34***	36.53***	2.22***	4.91***	9.46***	23.96***	2.57**	4.85***	4,52***	7.68***
$GDPpc^2$	4.09E-10	4.99E-09	1.12E-10	−4.67E-06	1.38E-10	1.83E-09	5.87E-11	−6.98E-06	4.21E-10	5.20E-09	9.95E-11	−4.79E-06
	9.49***	15.40***	10.27***	21.00***	2.52***	6.18***	11.27***	18.67***	1.90*	3.11***	2.00**	4.26***
Turning point ($)	23,105	25,210	26,744	28,396	23,551	21,284	22,402	27,928	22,447	24,740	29,347	28,012
R^2	0.12	0.37	0.50	0.64	0.10	0.32	0.43	0.58	0.10	0.36	0.51	0.63
Observations	2,497	2,497	2,497	2,497	2,497	2,497	2,497	2,497	141	141	141	141
Number of countries	141	141	141	141	141	141	141	141	141	141	141	141
Period	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006
Countries on the Right of the Turning Point, 2006												
	Australia	Australia	Australia	Australia	Australia	Australia	Australia	Australia	Australia	Australia	Australia	Australia
	Austria	Austria	Austria	Austria	Austria	Austria	Austria	Austria	Austria	Austria	Austria	Austria
	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium	Belgium
	Canada	Canada	Canada	Canada	Canada	Canada	Canada	Canada	Canada	Canada	Canada	Canada
	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark	Denmark
	Finland	Finland	Finland	Finland	Finland	Finland	Finland	Finland	Finland	Finland	Finland	Finland
	France	France	France	France	France	France	France		France	France	France	France
	Greece	Greece	Greece	Greece	Greece	Greece	Greece		Greece	Greece	Greece	
	Germany	Germany	Germany	Germany	Germany	Germany	Germany	Germany	Germany	Germany	Germany	Germany
	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong	Hong Kong
	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland	Ireland
	Israel					Israel	Israel		Israel			
	Italy	Italy	Italy		Italy	Italy	Italy		Italy	Italy		Italy
	Japan	Japan	Japan	Japan	Japan	Japan	Japan		Japan	Japan	Japan	Japan
	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands	Netherlands
	New Zealand				New Zealand	New Zealand			New Zealand			
	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway	Norway
	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore	Singapore
	Spain	Spain	Spain		Spain	Spain	Spain		Spain	Spain		
	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden	Sweden
	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland	Switzerland
	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom	United Kingdom
	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States

Absolute value of robust *t*-statistics under coefficients. ***, **, *: Significant at, respectively, 1%, 5%, and 10% level. The full sample except microstates, GDP per capita PPP in constant 2005 international dollars from WDI. Author calculations using COMTRADE.

Table 3.—Robustness

	Logistic Transformation		Negative Binomial	System GMM			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent	Gini	Nber	Nber	HHI	Theil	Gini	Nber
GDPpc	−2.55E-04	3.18E-04	1.45E-04	−2.14E-05	−2.80E-04	−6.95E-06	2.97E-01
	25.57***	31.20***	29.32**	4.23***	7.10***	7.18***	11.42***
GDPpc^2	4.85E-09	−4.98E-09	−2.83E-09	4.46E-10	5.61E-09	1.20E-10	−5.41E-06
	16.21***	16.30***	20.42***	3.49***	5.35***	4.61***	7.46***
Turning point ($)	26,320	31,908	25,583	23,991	24,955	28,958	27,412
Observations	2497	2497	2497	2497	2497	2497	2497
Number of countries	141	141	141	141	141	141	141
Period	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006

Absolute value of robust *t*-statistics under coefficients. ***, **, *: Significant at respectively 1%, 5%, and 10% level. Full sample except microstates. GDP per capita PPP in constant 2005 international dollars from WDI. Author calculations using COMTRADE.

mation whose results are reported in columns 1 and 2. The turning point is at the usual level of about $26,000. Second, we correct for the potential endogeneity of GDP per capita to export concentration. As we have no valid outside instrument for GDP per capita for our large panel, we carry out a system GMM estimation. Results, presented in the columns 4 to 7, show a turning point varying between $24,000 (Herfindahl) and $29,000 (Gini), with the same countries to the right of the turning point.[8] Thus, by and large, both the existence of a turning point in export concentration and its location around a GDP per capita of about $22,000 to $27,000 at PPP in constant 2005 international dollars—a very late point in the development process—are fairly robust.

A glance at the columns entitled “Nber” in tables 2 and 3 shows a clear hump-shaped relation between the number of active export lines and GDP per capita. The turning point for the number of active export lines is always roughly at the same level of GDP per capita as that of the concentration indices (see also figure 1). As the number of lines is a count variable, we also run a negative binomial estimation. Results, reported in column 3 of table 3 are consistent with previous findings. The rising part of the curve corresponds to the introduction of new products as countries develop. Its decreasing part illustrates one of the striking findings of this paper: that high-income countries tend to close down export lines faster than they open up new ones, resulting in reconcentration at the extensive margin. We return to this point later.

Thus, our analysis, regressing concentration indices and the number of active lines on GDP per capita, shows a hump-shaped relationship between economic development and export diversification. Our next task is to understand what is behind the hump.

[8] A crucial issue with system GMM (Blundell & Bond, 1998) is the number of instruments to use. This number should not exceed the number of individuals in the panel (see Roodman, 2006). We make the standard choice of using two lags for the instruments of the differenced equation and one lag for the instruments of the level equation. Following Arellano and Bond (1991), we use the Sargan/Hansen test of overidentifying restrictions and a direct test for the absence of second-order serial correlation. Both fail to reject the null of no serial correlation.

III. Stages of Diversification: Extensive versus Intensive Margins

That export diversification would proceed in parallel with economic development is to be expected. Pretty much like human beings colonized new land to alleviate competitive pressure on existing pastures, entrepreneurs can be expected to look for new pastures and open up production and export lines at the extensive margin. As capital accumulates, this becomes easier. But the later reconcentration, although consistent with Imbs and Wacziarg’s findings for production and employment, is somewhat of a puzzle. In order to better understand what is behind the hump in the curve, we turn to a systematic analysis of the intensive and extensive margins using the decomposability property of Theil’s index.

The nonmonotone pattern of diversification revealed in section II (decreasing concentration up to $25,000 and increasing concentration thereafter) could be explained by change at the extensive margin, the intensive margin, or both. Diversification at the extensive margin occurs when the number of active lines rises. Diversification at the intensive margin occurs when the distribution of trade values across existing export lines becomes more even. That is, diversification at the intensive margin during a period t_0 to t_1 means convergence in export shares among goods that were exported at t_0. The evolution in the number of active lines identified in section II is suggestive of action at the extensive margin. In order to shed more light on the issue, we turn to a decomposition of Theil’s index, which can be usefully mapped into the intensive and extensive margins thus defined.

A. *Mapping the Theil Decomposition with the Extensive and Intensive Margins*

In this section, we combine the classic decomposition of Theil’s index into between- and within-groups components with a partition of export lines into active and inactive ones. The result is a perfect mapping of changes in the between-groups component of Theil's index into changes in the extensive margin of exports and of changes in its within-groups components into changes in the intensive margin of exports.

Theil's index has the property that it can be calculated for groups of individuals (export lines) and decomposed additively into within-groups and between-groups components (that is, the within- and between-groups components add up to the overall index). Specifically, let n be the total number of potential export lines (the 4,991 lines of the HS6 system) and μ their average dollar value. Consider some partition of that total number of potential exports (of a given country in a given year) into $J + 1$ groups denoted $G_j, j = 0,\ldots,J$. Let n_j be the number of export lines in group j and μ_j their average dollar value. Also let T_j stand for Theil's index for group j, calculated using equation (1) on the n_j lines making up group j. Finally, let x_k be the dollar value of export line k, regardless of which group it belongs to. The between-groups component of Theil's index is defined as

$$T^B = \sum_{j=0}^{J} \frac{n_j}{n}\frac{\mu_j}{\mu} \ln\left(\frac{\mu_j}{\mu}\right), \tag{2}$$

and its within-groups component is defined as

$$T^W = \sum_{j=0}^{J} \frac{n_j}{n}\frac{\mu_j}{\mu} T_j = \sum_{j=0}^{J} \frac{n_j}{n}\frac{\mu_j}{\mu} \left[\frac{1}{n_j}\sum_{k\in G_j}\frac{x_k}{\mu_j}\ln\left(\frac{x_k}{\mu_j}\right)\right]. \tag{3}$$

It is easily verified that $T^W + T^B = T$.

Suppose that for a given country and year, we partition the 4,991 lines making up the HS6 nomenclature into two groups: G_1 made of active export lines for that country and year, and G_0 made of inactive export lines. We want to use this partition to construct group Theil subindices, one for each group $j = 0,1$, and the within and between components of the Theil. The between-groups subindex is not defined since $x_k = 0$ for all k in G_0, so that $\mu_0 = 0$ and consequently the logarithm in expression (2) is not defined for $j = 0$. However, applying L'Hôpital's rule gives

$$\lim_{\mu_0\to 0}\left[\frac{\mu_0}{\mu}\ln\left(\frac{\mu_0}{\mu}\right)\right] = 0, \tag{4}$$

so given our partition,

$$\lim_{\mu_0\to 0} T^B = \frac{n_1}{n}\frac{\mu_1}{\mu}\ln\left(\frac{\mu_1}{\mu}\right). \tag{5}$$

As $\mu = \sum_{k=1}^{n} x_k/n$, $\mu_1 = \sum_{k\in G_1} x_k/n_1$, and $\lim_{\mu_0\to 0} \sum_{k\in G_1} x_k = \sum_{k=1}^{n} x_k$ (since lines outside G_1 must all tend to 0 for their mean to also tend to 0), it follows that $n_1\mu_1 \to n\mu$, so

$$\lim_{\mu_0\to 0} T^B = \ln\left(\frac{\mu_1}{\mu}\right) = \ln\left(\frac{n}{n_1}\right). \tag{6}$$

Letting Δ denote a period-to-period change and observing that n is time invariant, we have finally that

$$\lim_{\mu_0\to 0} \Delta T^B = -\Delta \ln n_1. \tag{7}$$

That is, given our partition, changes in the between-groups component of Theil's index measure changes at the extensive margin (proportional changes in the number of active lines).

As for the within-groups component, it is a weighted average of terms combining group-specific means (μ_j/μ) and group-specific Theil indices T_j (the terms in square brackets), the weights being n_j/n. In our case, T^W reduces to T_1, the group Theil index for active lines. To see this, write equation (3) in full as

$$\begin{aligned} T^W = {} & \frac{n_0}{n}\frac{\mu_0}{\mu}\left[\frac{1}{n_0}\sum_{k\in G_0}\frac{x_k}{\mu_0}\ln\left(\frac{x_k}{\mu_0}\right)\right] \\ & + \frac{n_1}{n}\frac{\mu_1}{\mu}\left[\frac{1}{n_1}\sum_{k\in G_1}\frac{x_k}{\mu_1}\ln\left(\frac{x_k}{\mu_1}\right)\right]. \end{aligned} \tag{8}$$

In group G_0, suppose that all lines have the same arbitrary, strictly positive value x_0, so $\mu_0 = x_0$. Then the first term in equation (8) is well defined and boils down to

$$\frac{n_0}{n}\frac{\mu_0}{\mu}\ln(1) = 0.$$

Moreover, this remains true as x_0 is made arbitrarily close to 0. Thus,

$$\lim_{x_0\to 0} T^W = \frac{n_1}{n}\frac{\mu_1}{\mu}\left[\frac{1}{n_1}\sum_{k\in G_1}\frac{x_k}{\mu_1}\ln\left(\frac{x_k}{\mu_1}\right)\right]. \tag{9}$$

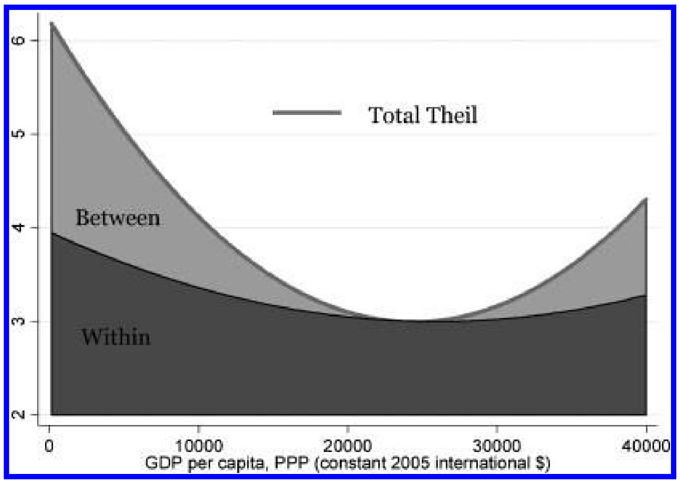
Now, as x_0 tends to 0, we noted already that $n_1\mu_1 \to n\mu$. It follows that

$$\lim_{x_0\to 0} T^W = \frac{1}{n_1}\sum_{k\in G_1}\frac{x_k}{\mu_1}\ln\left(\frac{x_k}{\mu_1}\right) = T_1. \tag{10}$$

Thus, given our partition, changes in the within-groups Theil index (ΔT^W) measure changes at the intensive margin (ΔT_1, that is, changes in concentration among active lines only).

In sum, the decomposition of Theil's index with our partition of export lines into active and inactive ones allows distinguishing changes in overall concentration into extensive- and intensive-margin changes. The evolution of the between component of the Theil corresponds to changes at the exten-

FIGURE 2.—WITHIN AND BETWEEN COMPONENTS OF THEIL'S INDEX



Author calculations using COMTRADE (quadratic estimates).

sive margin, whereas the evolution of the within component of the Theil reflects changes at the intensive margin.

We now put this decomposition to work. Figure 2 depicts the contribution of the between and within components to the overall Theil. We observe that in levels, the within component dominates the index, but in terms of evolution, most of the action is in the between component.[9]

Until about PPP$22,000, the between component shrinks faster than the within, so diversification occurs mostly at the extensive margin. Past that point and until the turnaround (at around PPP$25,000), it is the within component that decreases faster, so diversification occurs mostly at the intensive margin. That is, individual export values (and shares) converge among active lines.

Beyond the turning point, the index starts rising again, and its rise is driven almost exclusively by the between component. That is, reconcentration occurs at the extensive margin as countries close down active export lines. What are those lines?

Table B1 in appendix B shows the sectors and chapters mostly concerned with closure. The majority of chapters listed in table B1 are declining industries in high-income countries. Among the fifteen chapters that experienced the highest number of closed lines, three belong to the textiles sector, a fourth concerns raw hides and skins and leather, two belong to the vegetable products sector, two others to the live animal and animal products, two are from the mineral products sector, and one concerns iron and steel. Textiles (chapter 53) and leather (chapter 41) are among the most active "closers" (8.6 percent of the chapter's active lines for the former, 9.4 percent for the latter). The case of chemicals (chapters 29 and 28) is worth investigating. Although the chemicals sector does not necessarily come across as a declining sector for most developed countries (figure 4b confirms that high-income countries specialize in chemicals), chapters 29 and 28 rank high in their number of closed lines. The simultaneous occurrence of rising specialization and line closures in the chemical sector is, however, consistent with Schott's (2004) finding that specialization occurs within sectors, as high-tech exports replace low-tech ones when countries become more prosperous. The closure of export lines in the leather sector, by contrast, suggests between-product specialization, as leather or cotton works are labor-intensive activities in which countries lose comparative advantage when they grow. We explore this last point more intensively in section IV.

B. *What Are the New Export Products That Generate Trade Diversification?*

Although the most intriguing feature of the U-shape pattern is the exports' reconcentration of the richest countries, patterns of diversification at lower income levels are also of interest. As most of the diversification occurs at the extensive margin, one may indeed wonder what the characteristics of those new export products (new lines at the HS6 level) are.

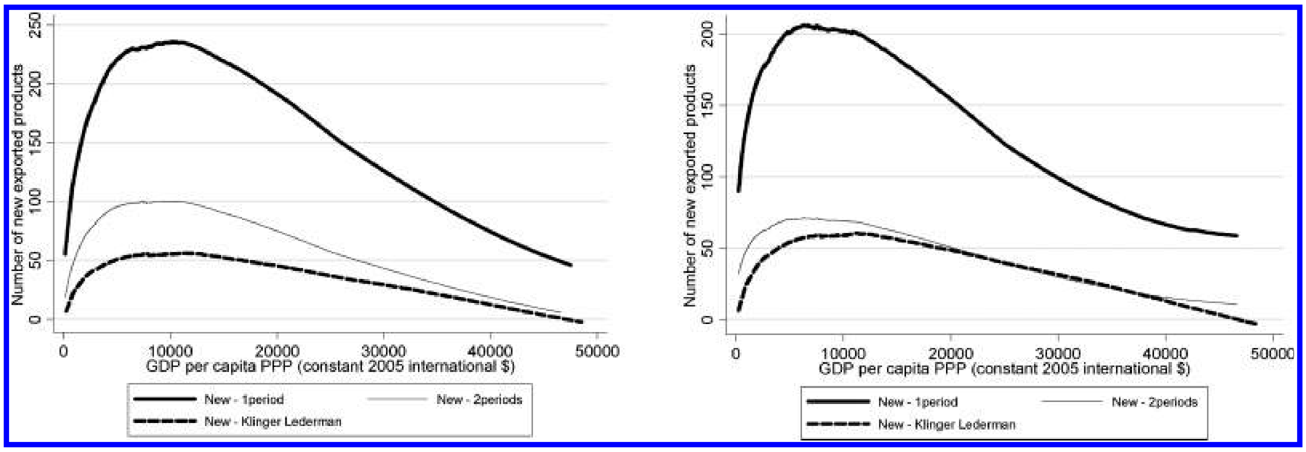
The number of new export products should be interpreted somewhat cautiously, as these products are not necessarily true entrepreneurial discoveries. In most cases, they correspond to the opening of new export lines that are already active in other countries. This is particularly true for developing countries that are copying existing products invented elsewhere and exporting those products as new export lines. In contrast, genuine innovations are incorporated within the HS6 classification in the course of periodic revisions and may not show up as new export lines.[10] Our new export products thus correspond to what Klinger and Lederman (2006) called "inside-the-frontier innovations." The focus of our paper is not innovation but export diversification within an existing (although arbitrarily limited) product nomenclature. Exporting a product for the first time (that is, opening a new export line), even if it were already produced or exported to other destinations, is an entrepreneurial risk worth investigating.

There is no conventional definition of new export products. In order to stay as close as possible to the definition of active lines and in the tradition of Besedes and Prusa (2006b), we first define new export products for a year and country as those lines that were not active in the country's export trade in the preceding year but were exported in the following year (one-year cutoff). This definition, based on a moving three-year window, reduces the sample period to 1989 to 2005, one year being taken out at both ends. As alternatives, we use (a) Klinger and Lederman's (2006) definition and (b) lines that were inactive in the country's

[9] When the slope of the overall Theil is at least twice that of its within component, the between component contributes for more than 50 percent to the overall index's decrease.

[10] At the HS6 level, reclassifications are limited, but we follow Besedes and Prusa (2006a) in treating them as censored, that is, a spell of, say, five years ending with a reclassification is treated as a spell of at least five years, like one at the end of the sample.

FIGURE 3.—PREDICTED NEW EXPORT LINES: NONPARAMETRIC ESTIMATES



(a) Using the available data for each definition of new products (1989–2005 for the one year cut-off, 1990–2004 for the two year cut-off, and 1997–2006 for Klinger and Lederman).
(b) All new products are computed over the 1997–2005 period.
Author calculations using COMTRADE.

export trade in the preceding two years but were exported in the following two years (two-years cutoff). This latter definition strikes a balance between the very conservative definition used by Klinger and Lederman (2006) and the very liberal one used by Besedes and Prusa (2006b).

Klinger and Lederman (2006) define "discoveries" as products not exported in the early part of their sample (1994–1996) but with over $10,000 of exports in the latter part (2002–2003). What is the difference between this definition and definitions that account for years of inactivity and activity around the first appearance of a product (one-year or two-year cutoff)? Conceptually these notions of new export products are essentially the same, being based on the idea that imperfectly informed entrepreneurs search for profitable export opportunities. Uncertainty can be about production costs, as in Hausmann and Rodrik (2003), or about foreign demand, as in Vettas (2000), but the point is that starting to export a product is an entrepreneurial gamble that may fail. Whereas Klinger and Lederman's definition singles out successful export line development (new lines that reach a threshold value), the other definitions include small-volume, short-spell lines in order to pick up the trial-and-error process at the extensive margin. The shorter the spell, the more discoveries or new products there should be, as new entrepreneurs try again a few months or years later. Besedes and Prusa (2006a) found that over half of all trade relationships were observed for a single year and 80 percent lasted less than five years. Our more aggregated HS6 data are likely to smooth some of those entries and exits, but Besedes and Prusa showed the high churning rate to be robust to aggregation.

Figure 3 shows the predicted number of new export products (per country-year, with several alternative definitions of new export products) against GDP per capita using the nonparametric (smoother) estimator. In all cases, the turning point comes very early: in the PPP $5,000 to 10,000 range. The rapid decrease in export entrepreneurship apparent in the figure could conceivably be due to equally rapid convergence toward the absolute barrier to diversification (the 5,000 lines of the HS system), but it is not, as few countries approach this barrier and certainly not those at GDP per capita levels around $5,000 to $10,000.[11]

The relationship between income and new export products is robust to the choice of definition of *new products*. The lower number of Klinger and Lederman's new export products in figure 3a could be expected from the more conservative aspect of their definition. It could also result from the shorter time frame on which new products are measured. As ten years are required to compute a new product according to Klinger and Lederman, we measured these new products for the 1997–2006 period against 1989–2005 for the Besedes and Prusa definition (one-year cutoff) and 1990–2004 for the two-year cutoff. Figure 3b depicts the nonparametric estimates of the predicted number of new export products against GDP per capita for the 1997–2005 period, which is common to all definitions. Once corrected for the number of years available, new export products per Klinger and Lederman are similar to new export products defined by the two-year cutoff. The one-year cutoff unsurprisingly counts more new products because it includes several of these new exports with extremely short spells, which can be assimilated to trial-and-error export products.

We finally ask whether new export products are any different from other traditional exports. Table 4 gives a characterization of export goods using Rauch's index of product differentiation. Rauch (1999) distinguished products traded on organized exchanges, products with reference prices, and differentiated ones. Table 4 shows the proportion of each of Rauch's categories in traditional and new export lines as measured according to Besedes and Prusa's

[11] Recall that on average, only half the HS6 lines are active for any country and year.

TABLE 4.—CHARACTERIZATION OF PRODUCTS BY DEGREE OF DIFFERENTIATION

	New Products (count number)	New Products	All Products	World Trade, 1990 (Rauch)[b]
Conservative classification[a]				
Homogeneous	7.6%	15.0%	31.9%	12.6%
Reference priced	28.1%	32.5%	27.4%	20.3%
Differentiated	64.4%	52.5%	41.0%	67.1%
Liberal classification[a]				
Homogeneous	12.0%	22.6%	39.1%	16.0%
Reference priced	26.8%	28.2%	19.7%	19.5%
Differentiated	61.2%	49.2%	41.3%	64.2%

In value of total trade unless otherwise indicated. Author calculations using COMTRADE.
[a] Because the classification of some products cannot be asserted unambiguously, Rauch's conservative classification assigns fewer products to the homogeneous and reference-priced categories than his liberal ones.
[b] From table 2 of Rauch (1999).

(2006b) definition. Using other definitions for new export products provides similar shares.

We find a lower share (in terms of export value) of homogeneous product exports among new than among traditional ones (15.0% versus 31.9% using Rauch's conservative classification and 22.6% versus 39.1% according to his liberal classification). The reverse is true for reference-priced and differentiated goods, suggesting that the bulk of diversification is made on these types of products. This feature is emphasized by the proportion of each of Rauch's categories in terms of the number of new lines. Differentiated goods account for 61.2% to 64.4% of new export lines in average over the 1989–2005 period.

Finally, like Besedes and Prusa (2006b) and Rauch and Watson (2003), we observe that initial trade in homogeneous products requires higher values than initial trade in differentiated products. The proportion of homogeneous goods in the total number of new export lines is smaller than its proportion in the total value of these new exports (7.6% versus 15% using Rauch's conservative classification and 12.0% versus 22.6% according to his liberal classification). The contrary is true for differentiated products (64.4% versus 52.5% using Rauch's conservative classification and 61.2% versus 49.2% according to his liberal classification).

Thus, new export products are essentially low-value-differentiated goods traded by low-income countries. These findings are consistent with the existing literature. Interestingly, they are independent of the definition chosen.

IV. Stages of Diversification: Alternative Explanations

Our decomposition of the Theil index highlights the importance of distinguishing the extensive from the intensive margins in the evolution of export diversification. It also suggested slow adjustment across diversification cones. We must, however, consider alternative explanations that could artificially create or reinforce a hump-shaped pattern. The diversification curve may result from spurious statistical effects, for example. Alternative explanations include (a) the potential role of primary resource exports as large exporters of mineral products (those for which mineral products represent over 50% of exports) are either low- or middle-income countries or very high-income ones in our database and (b) the structure of the HS6 COMTRADE classification, as textiles and clothing, essentially exported by low- to middle-income countries, have a large number of lines per dollar of export. We show in the next section that the hump-shaped relationship is robust to controls for these alternative explanations and then explore characteristics of closed lines that may help us understand what drives the hump shape.

A. *Primary Products*

We consider here the prevalence of primary resources in exports as an explanation for the U-shaped pattern of export concentration evidenced in section II. Where do we find large primary-resource exporters along the income axis? Figures 4 shows selected sectoral shares against GDP per capita.

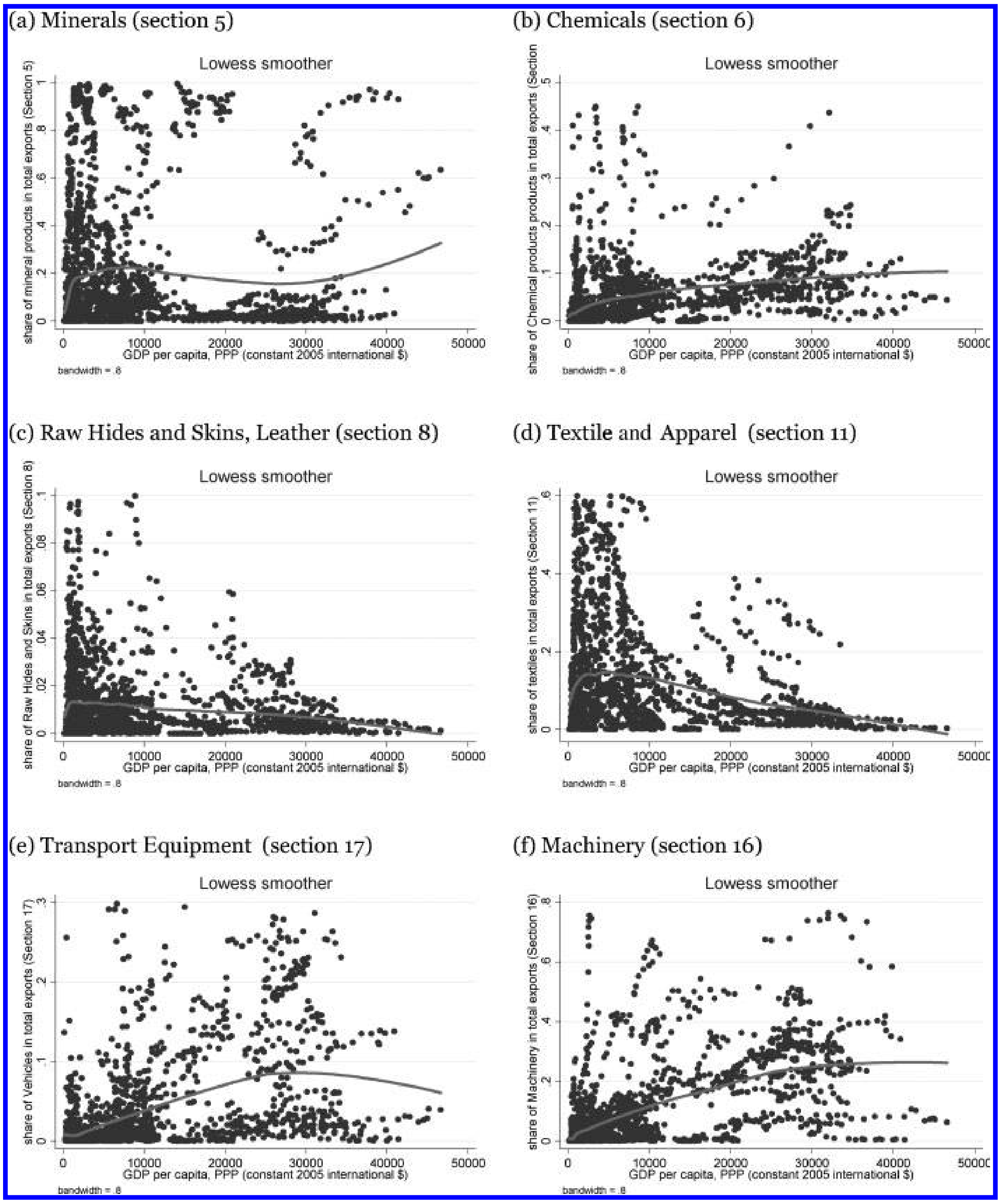
Figure 4a for minerals (HS section 5) shows a fairly distinct pattern whereby large exporters of mineral products (those for which mineral products represent over 50% of exports) are either low/middle income countries or very high-income ones. This pattern, which is confirmed by the nonparametric regression curve, is likely to contribute to the U-shaped pattern of export concentration.

As the large primary-product exporter status is a largely time-invariant country characteristic, the country fixed-effects estimator used in section II already suggests that the U-shaped pattern of export concentration is not a spurious one due to primary product exports. However, given the importance of primary product exports in the debate linking export concentration and development, we choose to go beyond the country-fixed-effects approach in two ways.

First, we exploit the time variation in the share of primary products in exports over the 1988–2006 period by including this variable (in an additive way) in our usual quadratic. We thus introduce in the model the share of HS chapters 26 (ores, slag, and ashes) and 27 (mineral fuels, mineral oils, and products of their distillation).[12] Results are shown in table 5.

[12] Chapters 26 and 27 belong to section 5.

FIGURES 4.—SELECTED SECTORAL SHARES AGAINST GDP PER CAPITA



Author calculations using COMTRADE.

Unsurprisingly, the share of raw materials comes out as a positive and significant contributor to export concentration (this is to be expected, as a large share of one narrow class of products is likely to be associated with high concentration) and as a negative one to the number of active lines (columns 1 to 4 of table 5). But the striking result is that coefficients on GDP per capita and its square are not affected by much, nor is the turning point.

Second, we want to know if the share of raw materials changes only the level of export concentration or if it also has an impact on the magnitude of the U-shape and the level of the turning point. We thus interact the share of raw materials in exports with GDP per capita (columns 5 to 8 of table 5). Figure 5 plots predicted Theil indices against GDP per capita for various levels of raw material export shares. Except for very high values of the share of raw materials

TABLE 5.—ESTIMATES WITH RAW MATERIAL EXPORT SHARES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber
GDPpc	−1.91E-05	−2.53E-04	−6.00E-06	2.66E-01	−2.80E-05	−3.23E-04	−6.73E-06	3.11E-01
	16.49***	33.46***	27.22***	39.98***	19.98***	33.96***	23.39***	44.87***
GDPpc^2	4.15E-10	5.03E-09	1.12E-10	−4.68E-06	6.76E-10	6.76E-09	9.15E-11	−5.76E-06
	12.6***	22.60***	13.78***	23.02***	16.88***	22.87***	9.70***	27.31***
Raw materials	0.5142	3.4746	0.0533	−1245.7	0.3425	1.5409	−0.0013	74.01
	36.39***	45.72***	24.96***	16.03***	12.57***	15.44***	0.88	0.78
GDPpc × Raw materials					4.01E-05	3.80E-04	7.82E-06	−2.53E-01
					10.84***	22.25***	18.4***	12.35***
GDPpc^2 × Raw materials					−1.05E-09	−8.49E-09	−1.04E-10	5.50E-06
					11.24***	16.5***	7.48***	9.17***
Turning point ($)	23,012	25,139	26,690	28,385	–	–	–	–
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,497	2,497	2,497	2,497	2,497	2,497	2,497	2,497
Number of countries	141	141	141	141	141	141	141	141
Period	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006	1988–2006

Absolute value of robust *t*-statistics under coefficients. ***, **, *: Significant at, respectively, 1%, 5%, and 10% level. Full sample except microstates. GDP per capita PPP in constant 2005 international dollars, from WDI. Author calculations using COMTRADE.

(over 70%), the U-shaped relationship is maintained with an almost unchanged turning point.

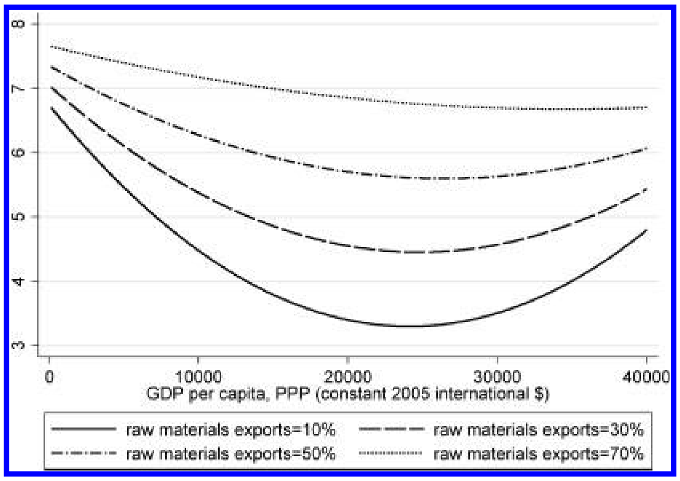
B. *The Harmonized System's Classification*

The harmonized system's classification used by COMTRADE could also potentially explain the hump-shaped relationship between economic development and export diversification. This classification is derived from nomenclatures originally designed for tariff collection purposes rather than to generate meaningful economics. Consequently, some sections have a large number of economically irrelevant categories (for example, the textile-clothing sector, section 11), whereas in other sections (for example, machinery, section 16), economically important categories are lumped together in a few lines. Now assume that products in section 11 are essentially exported by middle-income countries, whereas products in section 16 are essentially exported by high-income countries (assumptions confirmed by figures 4d and 4f, respectively). Then the observed diversification and reconcentration pattern could be an illusion caused by the structure of the HS6 classification.

Figure 6a, which plots, for each section of the HS6 classification, total export value versus number of lines provides evidence of this feature. Sections 6, 11, 15, and 16 have a much higher number of lines than others sectors of the HS6 classification. Section 16, however, differs from sections 6, 11, and 15 as it is well above the 45 degree line, reflecting a disproportionate high value per export line, whereas sections 6, 11, and 15 include a large number of small lines.

In order to control for the conjecture that the U-shape pattern of diversification may be a consequence of the structure of the HS6 classification, we went back to our raw database and reaggregated the lines in sections 6, 11, and 15 from HS6 (subheading) to HS4 (heading) level (because of its specificity, we treated section 16 separately, as we explain below). The number of lines in these sectors thus

FIGURE 5.—THEIL INDICES AGAINST GDP AND THE SHARE OF RAW MATERIALS



Author calculations using COMTRADE.

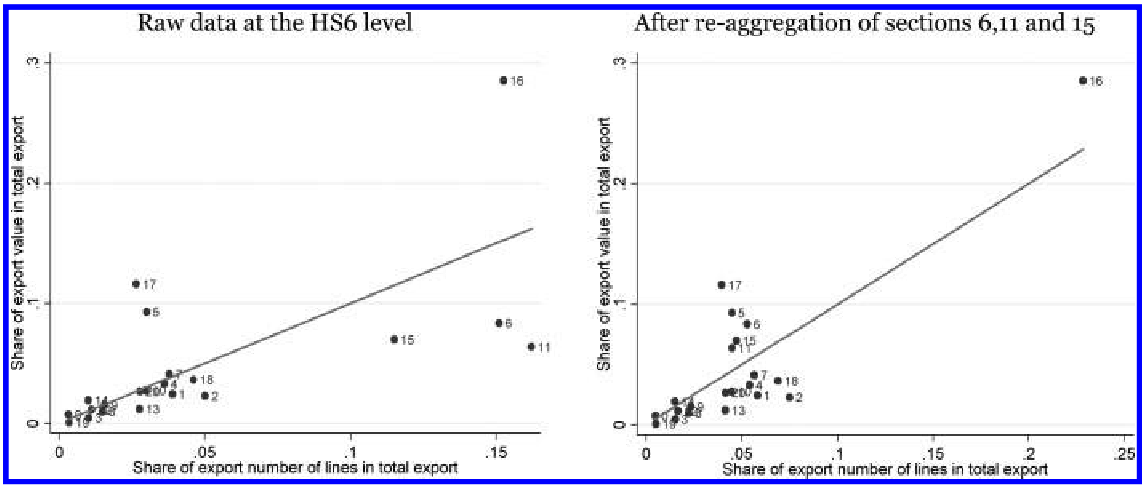
shrinks drastically, reducing the average value per line to a level comparable to that of other sections, as reported in figure 6b.

Our new classification (HS4 for sections 15, 6, or 11 and HS6 otherwise) contains 3,336 product lines instead of 4,991 for the benchmark classification. Results obtained with Theil indices calculated on the modified database are not significantly different from the ones obtained above: the turning point is consistent with previous findings under pooled or within estimation.[13]

Figure 6 reveals that section 16 has both a large number of lines and a disproportionately high value per export line (the section represents around 25 percent of the total value of exports). The high value per export lines suggests that the number of existing lines is not extended enough to represent production in this section in a similar way as other

[13] Results available on request.

FIGURE 6.—SECTION SHARES IN NUMBER OF LINES AND TRADE VALUE



Author calculations using COMTRADE.

sections of the HS6 classification. Mammoth lines may indeed include many more products than lines in other sections. This could artificially lead to the high concentration of high-income countries.

We thus need to control for the particular design of section 16. As we cannot further disaggregate section 16, we dropped this sector from the database. Our final classification thus contains 2,575 product lines. Results (not reported here but available on request) are similar to the one obtained with the benchmark classification: the turning point is robust to the aggregation of section 6, 11, or 15 and the elimination of section 16 in the pooled as well as in the within estimation.

The hump-shaped relationship between economic development and export diversification is thus not a consequence of spurious composition effects.[14]

C. *Traveling across Diversification Cones*

As Schott (2003, 2004) and Xiang (2007) discussed, countries travel across diversification cones when they accumulate capital. As they do, "old-cone" lines should become inactive while "new-cone" ones should become active. Suppose that old-cone lines are slow to die because of incumbency advantages, established ties with customers, or any other kind of support they may get. During the transition phase, new-cone lines become active, while old-cone ones do not want to die. As a result, exports diversify, and the total number of active lines rises. As time passes, however, comparative advantage catches up on old lines, and they slowly die, reducing diversification. Viewed in this way, high diversification at middle-income levels is essentially a transitory phenomenon between two steady states in terms of industrial specialization.

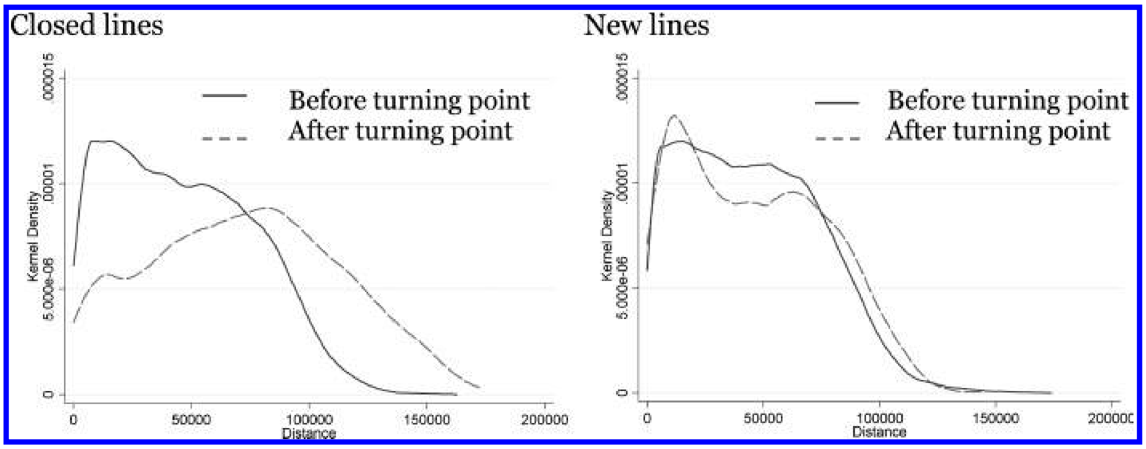
Besedes and Prusa's (2006b) finding that the hazard rate decreases rapidly in the first years of an export spell is indeed suggestive of a dual regime with high infant mortality, consistent with Hausmann and Rodrik's (2003) view of an entrepreneurial trial-and-error process, and persistence among "old" spells, consistent with the conjecture above. It is also consistent with Schott's (2003) finding that "estimated development paths deviate substantially from the theoretical archetypes of figures 4 [a systematic pattern of births for new-cone industries and deaths for old-cone ones]. Many sectors, including apparel and footwear, exhibit positive value-added per worker in more than two cones" (pp. 693–696). Apparel and footwear could indeed be slowly dying industries in many countries, not only on the import-competing side but also on the export side (the EU, for instance, is still a major exporter of textile and apparel products). If that were the case, the high diversification characterizing the middle part of the economic development process would not be a desirable outcome per se but simply an out-of-equilibrium one characterizing the transition from one steady state to another, each characterized by specialization according to comparative advantage.

A comparison of figures 4d and 4f, which show, respectively, the shares of textile and apparel products (section 11) and machinery (section 16) in exports as a function of GDP per capita, partly bears out this story, as the former follows a decreasing and only mildly convex trajectory (see the smoother fitted curves), while the latter follows a rising and concave one. The combination of the two generates a decrease in export concentration up to the $10,000 threshold, after which there is not much more action as both textiles and machinery stabilize at low (5 percent) and high (30 percent) shares, respectively.

Suppose that when a country reconcentrates, export lines that it closes are old-cone lines that were still in that coun-

[14] We also ran our baseline concentration regression with the share of service exports in GDP on the right-hand side. Results (available on request) were unchanged: the turning point was nearly the same. We thank Carsten Fink for giving us the service data.

FIGURE 7.—KERNEL DENSITY OF FACTOR-INTENSITY DISTANCES FOR CLOSED LINES



The horizontal axis measures the Euclidean distance between the factor intensity of closed lines (see the text for details on the calculation) and the factor endowment of the country closing it, all for the year in which the closure takes place. Author calculations using COMTRADE.

try's export portfolio essentially by inertia. In that case, lines closed by a country to the right of the diversification turning point would lie further from its comparative advantage than lines closed, in the process of normal churning, by countries to the left of the turning point.

This is a conjecture we can verify, albeit indirectly. To do this, we use a database compiled by Cadot, Shihotori, and Tumurchudur (2008). The databases contain national factor endowments (capital per worker and educational achievement) as well as revealed factor intensities calculated at the HS6 level as weighted averages of the factor endowments of countries exporting each good. The construction of these revealed factor intensities follows the logic of Hausmann et al.'s (2007) PRODY. That is, the revealed capital intensity of product k is

$$\hat{\kappa}_k = \sum_i \omega_{ik} \kappa_i, \tag{11}$$

where κ_i is country i's capital/labor endowment calculated according to Easterly-Levine (2001) and ω_{ik} is its (Balassa) index of revealed comparative advantage in good k. Human capital intensities (h_i for country i) are from Barro and Lee's (2000) national educational achievements database, and the revealed human capital intensity of product k is calculated in a way similar to equation (11). We compare the revealed factor intensity of closed line k, computed in this way, with the endowment of the country closing it, using a Euclidean distance formula:

$$d^e_{ik} = \left[\left(h_i - \hat{h}_k\right)^2 + \left(\kappa_i - \hat{\kappa}_k\right)^2\right]^{1/2}. \tag{12}$$

If our conjecture were correct, d^e_{ik} should be larger for lines closed by countries to the right of the turning point (declining industries) than for lines closed by countries to the left of it (normal churning). The left panel of figure 7 shows just that pattern. The density of d^e_{ik} for lines closed by countries to the left of the diversification turning point (solid line) peaks near the vertical axis, suggesting small distances between their factor intensities and the endowments of countries closing them ("accidental" closures).[15] By contrast, the density of d^e_{ik} for lines closed by countries to the right of the turning point (broken line) peaks far from the vertical axis, suggesting large distances (products far from the closing country's current diversification cone). To make the argument clear, the average intensity of lines closed by countries to the right of the turning point is between the factor endowments of Chile and Malaysia, whose income is about half the turning point.

The right panel of figure 7 provides a counterfactual. Densities estimated in a similar way for new export lines peak near 0, suggesting that the factor intensity of new export lines coincides roughly with the endowment of the countries introducing them. Moreover, there is no clear difference between the lines introduced by countries to the right of the turning point and those introduced by countries to the left.

In order to go beyond descriptive statistics, we regressed endowment-intensity distances (d^e_{ik}) on the status of countries (a dummy variable equal to 1 for countries to the right of the turning point and 0 otherwise) first on the subsample of closed lines, and then on the subsample of new lines for the counterfactual. Table 6 presents the results (see columns 1 and 2, respectively) and confirms the findings of figure 7. The coefficient on the status dummy is positive and significant for closed lines but insignificant for new lines.

Columns 3 and 4 of table 6 show that the factor intensities of lines closed to the right of the turning point are not

[15] In order to limit the number of one year trial-and-error cases in our estimation, we define closed lines (in a similar way as "new export lines") as lines that were open for two years and remained subsequently closed for two years. The kernel estimation is thus performed on lines closed between 1990 and 2003 (endowment-intensity distances are not available in the Cadot et al., 2008, database for 2004). Note that we also run the exercise defining closed lines as lines that had been open for one year and remained subsequently closed for one year. Although there are around five times more closed lines with this definition, we observe the same patterns as the ones described in this section.

TABLE 6.—REGRESSION RESULTS: INTENSITY/ENDOWMENT DISTANCES ON CLOSED LINE STATUS

Dependent Variable	Intensity/Endowment Distance		Intensity/Endowment Difference: Human Capital	Intensity/Endowment Difference: Capital
Sample:	Closed Lines (1)	New Lines (2)	Closed Lines (3)	Closed Lines (4)
Status (ATP=1)	25,171.40	465.22	4.96	108,333.00
	(39.5)***	(0.8)	(84.8)***	(129.8)***
Observations	31,372	98,390	31,372	31,372
R^3	0.06	0.01	0.19	0.36

Estimation is by OLS; year dummies are not reported in order to save space. Absolute value of robust t-statistics under coefficients. ***, **, *: Significant at respectively 1%, 5%, and 10% level. The dependent variable in columns 1 and 2 is the Euclidean distance between the factor intensity of closed lines (see text for details on the calculation) and the factor endowment of the country closing it, all for the year in which the closure takes place. Columns 3 and 4, report the algebraic difference between the factor endowment of the closing country and the factor intensity of the closed line (for human and physical capital respectively). The status regressor is a dummy variable equal to 1 when the line is closed by a country to the right of the turning point in year t. Thus, ignoring the year dummies, column 4 says that

$$\Delta K = -22,909 + 108,333 I_R,$$

where capital is measured in 2000 PPP dollars and the status dummy I_R is

$$I_R = \begin{cases} 1 & \text{if country is to the right of the turning point in } t \\ 0 & \text{otherwise.} \end{cases}$$

Thus, the negative intercept means that a closed line is on average \$22,909 more capital intensive than the endowment of the country closing it when it is left of the turning point, and \$108,333−27,909 = \$80,424 less intensive to the right of the turning point. By way of comparison, France's capital endowment (capital per worker at 2000 PPP dollars) was \$139,000 in 2003. Author calculations using COMTRADE.

just far from the endowments of the countries closing them, but also less intensive in human capital and physical capital. That is, in column 3, the dependent variable is $d_{ik}^h = h_i - h_k$, and in column 4, it is $d_{ik}^\kappa = \kappa_i - \kappa_k$. The status dummy is again positive and highly significant.

The evidence brought together in this section is only suggestive of a pattern whereby the closure of export lines in declining industries is delayed, but it certainly goes in that direction. It also confirms the prima facie evidence in Appendix B, where declining industries figure prominently among closed lines.

V. Conclusion

The results presented so far suggest two observations. First, there seems to be, across countries and time, a robust hump-shaped relationship between export diversification and the level of income (the mirror image of our U-shaped concentration indices). This nonmonotonicity holds both between and within countries. The reconcentration of exports above a threshold around PPP \$25,000 is especially striking. Diversification occurs mostly at the extensive margin, especially early in the development process, as new export items multiply and are marketed at increasingly large initial scales. This relationship does not appear to be spurious or driven only by variations in the share of primary products. From a policy perspective, it thus appears as a key element of the economic development process and is, if not necessarily an objective per se, at least an important policy indicator. From an econometric perspective, our findings justify treating export diversification as endogenous in growth regressions, as de Ferranti et al. (2002) do.

The second observation is that diversification at middle to high levels of income may simply reflect a slow adjustment process between two equilibria, with new export sectors being faster to appear than old ones are to die. We find evidence that countries to the right of the turning point close lines that are typically, in terms of factor intensities, far from their endowments—outliers in their export portfolios. The hump-shaped relationship between diversification and development may be explained by this slow adjustment as countries travel across diversification cones.

REFERENCES

Arellano, M., and S. Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies* 58 (1991), 277–297.

Auty, R., *Resource-Based Industrialization: Sowing the Oil in Eight Developing Countries* (New York: Oxford University Press, 1990).

Barro, R. J., and J.-W. Lee, "International Data on Educational Attainment: Updates and Implications," CID working paper no. 42 (2000).

Besedes, T., and T. Prusa, "Surviving the U.S. Import Market: The Role of Product Differentiation," *Journal of International Economics* 70 (2006a), 339–358.

——— "Ins, Outs, and the Duration of Trade," *Canadian Journal of Economics* 39 (2006b), 266–295.

Blundell, R., and S. Bond, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* 87 (1998), 115–143.

Brenton, P., and R. Newfarmer, "Watching More Than the Discovery Channel: Export Cycles and Diversification in Development," World Bank policy research working paper, no. 4302 (2007).

Brunnschweiler, C., "Cursing the Blessings? Natural Resource Abundance, Institutions, and Economic Growth," *World Development* 36 (2008), 399–419.

Brunnschweiler, C., and E. Bulte, "Linking Natural Resources to Slow Growth and More Conflict," *Science 320* (2008), 616–617.

Cadot, O., M. Shirotori, and B. Tumurchudur, "Revealed Factor Intensities at the Product Level," mimeograph, UNCTAD (2008).

de Ferranti, D., G. Perry, D. Lederman, and W. Maloney, *From Natural Resources to the Knowledge Economy: Trade and Job Quality* (Washington, DC: World Bank, 2002).

Dutt, P., I. Mihov, and T. van Zandt, "Trade Diversification and Economic Development," mimeograph, INSEAD (2008).

Easterly, W., and R. Levine, "It's Not Factor Accumulation: Stylized Facts and Growth Models," *World Bank Economic Review* 15 (2001), 177–219.
Evenett, S., and A. Venables, "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows," mimeograph (2002).
Feenstra, Robert C., "U.S. Exports, 1972–1994: With State Exports and Other U.S. Data," NBER working paper no. 5990 (1997).
Feenstra, Robert C., John Romalis, and Peter K. Schott, "U.S. Imports, Exports, and Tariff Data, 1989–2001," NBER working paper no. 9387 (2002).
Gelb, A., *Windfall Gains: Blessing or Curse?* (New York: Oxford University Press, 1988).
Griliches, Z., and J. A. Hausman, "Errors in Variables in Panel Data," *Journal of Econometrics* 31 (1986), 93–118.
Hausmann, R., J. Hwang, and D. Rodrik, "What You Export Matters," *Journal of Economic Growth* 12 (2007), 1–25.
Hausmann, R., B. Klinger, "Structural Transformation and Patterns of Comparative Advantage in the Product Space," mimeograph, Harvard University (2006).
Hausmann, R., and D. Rodrik, "Economic Development as Self-Discovery," *Journal of Development Economics* 72 (2003), 603–633.
Hummels, D., and P. Klenow, "The Variety and Quality of a Nation's Exports," *American Economic Review* 95 (2005), 704–723.
Imbs, J., and R. Wacziarg, "Stages of Diversification," *American Economic Review* 1993 (2003), 63–86.
Kehoe, T. J., and K. J. Ruhl, "How Important Is the New Goods Margin in International Trade?" Society for Economic Dynamics meeting papers no. 733 (2006).
Klinger, B., and D. Lederman, "Discovery and Development: An Empirical Exploration of 'New' Products," mimeograph (2004).
——— "Diversification, Innovation, and Imitation inside the Global Technology Frontier," World Bank Policy Research working paper no. 3872 (2006).
Neary, R., and S. van Wijnbergen, *Natural Resources and the Macroeconomy* (Oxford: Basil Blackwell, and Cambridge, MA: MIT Press, 1986).
Prebisch, R., *The Economic Development of Latin America and Its Principal Problems* (Lake Success, NY: United Nations Department of Economic Affairs, 1950). Reprinted in *Economic Bulletin for Latin America* 7 (1962), 11–22.
Rauch, J., "Networks versus Markets in International Trade," *Journal of International Economics* 48 (1999), 7–35.
Rauch, J., and J. Watson, "Starting Small in an Unfamiliar Environment," *International Journal of Industrial Organization* 21 (2003), 1021–1042.
Roodman, D., "How to Do xtabond2: An Introduction to 'Difference' and 'System' GMM in Stata," Center for Global Development working paper no. 103 (2006).
Sachs, J., and A. Warner, "The Big Rush, Natural Resource Booms and Growth," *Journal of Development Economics* 59 (1999), 43–76.
Schott, P., "One Size Fits All? Heckscher-Ohlin Specialization in Global Production," *American Economic Review* 93 (2003), 686–708.
——— "Across-Product Versus Within-Product Specialization in International Trade," *Quarterly Journal of Economics* 119 (2004), 647–678.
Singer, H., "US Foreign Investment in Underdeveloped Areas: The Distribution of Gains Between Investing and Borrowing Countries," *American Economic Review* 40 (1950), 473–485.
Theil, H., *Statistical Decomposition Analysis* (Amsterdam: North-Holland, 1972).
Vettas, N., "Investment Dynamics in Markets with Endogenous Demand," *Journal of Industrial Economics* 48 (2000), 189–203.
Xiang, C., "Diversification Cones, Trade Costs and Factor Market Linkages," *Journal of International Economics* 71 (2007), 448–466.

APPENDIX A

Data Description

The Harmonized System's classification of goods is defined by the number of digits used, which goes from 1 (sections, numbering 21) to 2 (chapters, numbering 99), 4 (headings, numbering 1,243), and 6 (subheadings, numbering around 5,000). Between 1988 and 2006, there were three classifications: HS0-1988/1992 (5,015 products), HS1-1996 (5,111 products), and HS2-2002 (5,222 products). We convert the HS1 and HS2 classifications into HS0 (using WITS conversion tables) and drop 24 HS0 lines that were no longer present in the HS1 and HS2 classification. This yields 4,991 lines.

Further degrees of disaggregation (HS 8, 10, and beyond) are not harmonized across members of the World Customs Organization and require caution in using. For instance, Eurostat, the European Union's statistical division, frequently reclassifies goods, shifting them back and forth between different HS8 codes from one year to another. This problem also affects U.S. trade data compiled by Feenstra in the NBERTD (see Feenstra, 1997, and Feenstra, Romalis, & Schott, 2002).

COMTRADE does not always report inactive export lines as zero lines, as national customs often omit those lines. In a first step, we have thus harmonized sample size for all countries and years by adding the missing lines and assigning them zero trade values. We thus work with 4,991 observation per country-years. However, we do not have a perfectly balanced country-year database. Actually, for our baseline country-year regressions, we use 2,497 observations, corresponding to 141 countries over 1988 to 2006 (with an average number of observations per country of 18, with a minimum of 7 and a maximum of 19).

Finally, in order to limit potential errors in reported trade flows, we use mirrored data. Such data are more accurate than direct export data, in particular for developing countries. Actually, it is well known that imports are better reported than exports. Moreover, remaining errors in reported trade flows, when using mirror data, are no more related to exporting countries' income levels, limiting measurement error issues in the estimation.

APPENDIX B

Closed Lines by Chapter

TABLE B1.—CUMULATED CLOSED LINES, 2003–2005, BY MAIN CHAPTERS—COUNTRIES WITH GDP PER CAPITA OVER PPP$25,000

	Chapter		Corresponding Section	Cumulated Closed Lines 2003-2005, Country Average: Number in % of Chapter Active Lines in 2000	Number in % of Total Closed Lines	Value in % of Total Export Value in 2002
72	Iron and Steel	15	Base Metals and Articles of Base Metal	2.4%	10.8%	0.0005%
28	Inorganic Chemicals; Organic or Inorganic Compounds of Precious Metals, Of Rare-earth Metals, of Radioactive Elements or of Isotopes	6	Products of the Chemical or Allied Industries	3.3%	7.4%	0.0002%
29	Organic Chemicals	6	Products of the Chemical or Allied Industries	2.8%	7.3%	0.0005%
41	Raw Hides and Skins (Other Than Furskins) and Leather	8	Raw Hides and Skins,Leather, Furskins and Articles Thereof; Saddlery and Harness; Travel Goods, Handbags, and Similar Containers	9.4%	6.1%	0.0006%
52	Cotton	11	Textiles and Textile Articles	3.0%	4.8%	0.0001%
25	Salt, Sulphur, Earths and Stone; Plastering Materials, Lime and Cement	5	Mineral Products	3.5%	4.0%	0.0002%
68	Articles of Stone, Plaster, Cement, Asbestos, Mica or Similar Materials	13	Articles of Stone, Plaster, Cement, Asbestos, Mica or Similar Materials	2.6%	3.3%	0.0000%
48	Paper and Paperboard; Articles of Paper Pulp, of Paper Or of Paperboard	10	Pulp of Wood or of other Fibrous Cellulosic Material; Waste and Scrap of Paper or Paperboard; Paper and Paperboard and Articles Thereof	1.4%	2.9%	0.0003%
53	Other Vegetable Textile Fibres; Paper Yarn and Woven Fabrics of Paper Yarn	11	Textiles and Textile Articles	8.6%	2.8%	0.0001%
26	Ores, Slag and Ash	5	Mineral Products	7.8%	2.7%	0.0004%
11	Products of the Milling Industry; Malt; Starches; Inulin; Wheat Gluten	2	Vegetable Products	3.9%	2.6%	0.0001%
3	Fish & Crustaceans, Molluscs & Other Aquatic Invertebrates	1	Live Animals; Animal Products	3.4%	2.5%	0.0000%
12	Oil Seeds and Oleaginous Fruits; Misc, Grains, Seeds & Fruit; Industrial or Medicinal Plants; Straw and Fodder	2	Vegetable Products	4.6%	2.1%	0.0000%
2	Meat and Edible Meat Offal	1	Live Animals; Animal Products	4.4%	2.1%	0.0001%
55	Man-made Staple Fibres	11	Textiles and Textile Articles	2.4%	2.0%	0.0001%

Closed lines at date t are defined as lines with positive exports at $t-2$ and $t-1$ and 0 exports at t, $t+1$ and $t+2$. The sample is restricted here to countries with populations above 1 million (no microstates) and GDP per capita above PPP$25,000 (at the right of the turning point). Data are cumulated over 2003–2005 for robustness. Author calculations using COMTRADE.

This article has been cited by:

1. Sandra Poncet, Felipe Starosta de Waldemar. 2013. Export Upgrading and Growth: The Prerequisite of Domestic Embeddedness. *World Development* **51**, 104-118. [CrossRef]
2. Olivier Cadot, Céline Carrère, Vanessa Strauss-Kahn. 2013. OECD imports: diversification of suppliers and quality search. *Review of World Economics* . [CrossRef]
3. Olivier Cadot, Céline Carrère, Vanessa Strauss-Kahn. 2013. TRADE DIVERSIFICATION, INCOME, AND GROWTH: WHAT DO WE KNOW?. *Journal of Economic Surveys* **27**:4, 790-812. [CrossRef]
4. Iza Lejárraga, Peter Walkenhorst. 2013. Economic policy, tourism trade and productive diversification. *International Economics* . [CrossRef]
5. Julie Regolo. 2013. Export Diversification: How Much Does the Choice of the Trading Partner Matter?. *Journal of International Economics* . [CrossRef]
6. Mario Cimoli, Sebastian Fleitas, Gabriel Porcile. 2013. Technological intensity of the export structure and the real exchange rate. *Economics of Innovation and New Technology* **22**:4, 353-372. [CrossRef]
7. Aleksandra Parteka, Massimo Tamberi. 2013. What Determines Export Diversification in the Development Process? Empirical Assessment. *The World Economy* **36**:6, 807-826. [CrossRef]
8. M. Jaud, O. Cadot, A. Suwa-Eisenmann. 2013. Do food scares explain supplier concentration? An analysis of EU agri-food imports. *European Review of Agricultural Economics* . [CrossRef]
9. Gordon H Hanson. 2012. The Rise of Middle Kingdoms: Emerging Economies in Global Trade. *Journal of Economic Perspectives* **26**:2, 41-64. [CrossRef]
10. Céline Carrère, Julien Gourdon, Marcelo Olarreaga. 2012. Regional integration and natural resources: Who benefits? Evidence from MENA. *International Economics* **131**, 23-41. [CrossRef]
11. Céline Carrère. 2012. UEMOA, CEMAC : quelle performance en matière de commerce ?. *Revue d'économie du développement* **27**:1, 33. [CrossRef]
12. Arne Wiig, Ivar Kolstad. 2011. If diversification is good, why don't countries diversify more? The political economy of diversification in resource-rich countries. *Energy Policy* . [CrossRef]