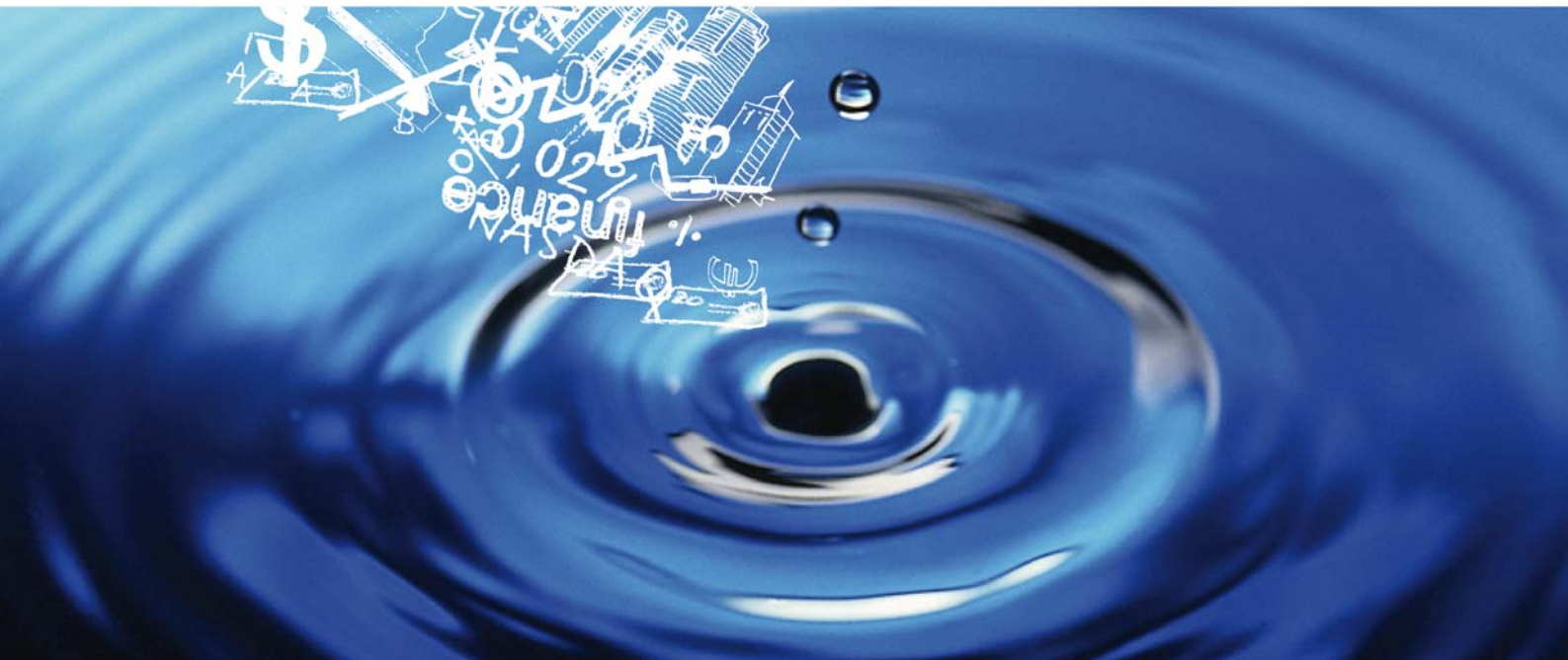
UNIL | Université de Lausanne
CREA – Institut de
macroéconomie appliquée

HEC Lausanne

Export Diversification: What's behind the Hump?

Céline Carrère, Vanessa Strauss-Kahn, Olivier Cadot

Novembre 2007



Export Diversification: What's behind the Hump?*

November 2007

Céline Carrère+
Vanessa Strauss-Kahn§
Olivier Cadot*

Abstract

The paper explores the evolution of export diversification patterns along the economic development path. Using a large database with 159 countries over 17 years at the HS6 level of disaggregation (4'998 product lines) we look for action at the "intensive" and "extensive" margins (diversification of export values among active product lines and by addition of new product lines respectively) using various export concentration indices and the number of active export lines. We also look at new product introduction as an indicator of "export-entrepreneurship". We find a hump-shaped pattern of export diversification similar to what Imbs and Wacziarg (2003) found for production and employment. Low and Middle income countries diversify mostly along the extensive margin whereas high income countries diversify along the intensive margin and ultimately re-concentrate their exports towards fewer products. Such hump-shaped pattern is consistent with the conjecture that countries travel across diversification cones as discussed in Schott (2003, 2004) and Xiang (2007).

Keywords: Export diversification, International trade, Latin America

JEL classification codes: F1, O11

---

* Research on this paper was supported by a grant from the Interamerican Development Bank and by Switzerland's Fonds National pour la Recherche Scientifique. Without implicating them, the authors would like to thank Antoni Estevadeordal and Christian Volpe for useful comments.
+ CERDI-CNRS, Université d'Auvergne.
§ INSEAD and CEPR.
* HEC Lausanne, CERDI, CEPR and CEPREMAP.

1. Introduction

Why should we care about export diversification? David Ricardo showed a century and a half ago that countries should *specialize*, not diversify. Monopolistic-competition models suggest that larger countries produce a broader range of products, but that hardly makes diversification a policy objective in itself. Yet as de Ferranti et al. (2002) note, "[a] recurrent preoccupation of [Latin American] policymakers is that their natural riches produces a highly concentrated structure of export revenues, which then leads to economic volatility and lower growth" (p. 38).

The view that concentration is associated with deteriorating terms of trade, income volatility and, ultimately, low growth, goes back to the work of Prebisch (1950) and Singer (1950). Though it has proved difficult to ascertain whether the terms of trade of primary-product exporters do have a deterministic downward trend or not (on this, see Cuddington et al. 2001), evidence in favor of the Prebisch-Singer hypothesis is fairly strong. Regressions on cross-sections of countries (see e.g. Sachs and Warner 1995, or more recently Gylfason 2004) and panels (de Ferranti et al. 2002) suggest that export concentration is indeed statistically associated with slow growth, in particular when export concentration reflects the predominance of primary products, as it usually does. Interestingly, Herzer (2004) also found a long-run statistical association between growth and export diversification on the basis of time-series data from Chile.[1] However evidence in favor of the Prebisch-Singer hypothesis only means that moving away from primary products is desirable; not that diversification is desirable *per se*. Assessing whether or not the quest for export diversification is a meaningful policy objective in itself requires a deeper understanding of how it relates with economic development.

[1] Herzer uses Perron's test for unit roots in the presence of structural breaks, which is of course particularly important given Chile's choppy growth history.

How export diversification evolves, empirically, along the path of economic development was, up to now, a relatively little-explored question. Imbs and Wacziarg (2003) were the first to uncover a non-monotone path of production and employment as functions of per-capita incomes, with diversification followed by re-concentration. Klinger and Lederman (2004, 2005) shortly followed suit with a similar result on export data. While Imbs and Wacziarg's exercise was a purely empirical one, Klinger and Lederman built on Hausmann and Rodrik (2003) to explore a causal link from market failures to insufficient diversification. Essentially, the story is that opening up new export markets is an entrepreneurial gamble which, if successful, is quickly imitated. The inability of export entrepreneurs to keep private the benefits of their activity leads to a classic public-good problem. Poor institutions, Klinger and Lederman show, appear empirically to compound the problem, lending support to the Hausmann-Rodrik view.

More recently, a number of paper have analyzed more closely the evolution and path-dependence of export patterns along the development process. Hausmann, Hwang and Rodrik (2005) argued that in the presence of externalities, specialization patterns are not fully determined by endowments; as a result, they can display path dependence. They proposed a measure of the technology content of exports based on the average income level of exporters of the same product and showed that it correlates with future growth (on this, see also Klinger 2007). Hausmann and Klinger argued that export patterns do not evolve smoothly across a continuous product space, but progress in leaps and bounds across a heterogeneous space. They proposed a measure of "product proximity" based on the conditional probability that one product is exported given that the other is also exported and used it to show graphically variations in the density of the product space. Hidalgo, Klinger, Barabasi and Hausmann (2007) and Hausmann and Klinger (2007) showed that the density of the product space as

measured through this metric is higher for high-technology products, making export redeployment easier for exporters of those products.

We revisit the issue of diversification using a different methodological perspective. We constructed a very large database covering 159 countries (including 121 developing countries) over all years available from the COMTRADE database at the highest disaggregation level (HS6). Using this database, we calculated for all countries and years three variables of interest: an export concentration index (we will use alternatively the Herfindahl, Theil and Gini indices), the number of active lines (lines with nonzero exports), and a measure of "new export products" identified, for each year and country in the sample, as export lines that are active and would remain so for two years but had been inactive during the previous two years.

Using Hummels and Klenow's (2005) terminology, we use these three variables to explore action along the "intensive" and "extensive" margins (diversification of export values among active product lines and by addition of new product lines respectively), as well as structural differences between traditional and new products.

We find a hump-shaped relationship between economic development and export diversification, like Imbs-Wacziarg and Klinger-Lederman, with a turning point around 20'000-22'000 dollars per capita at purchasing-power parity (PPP). At incomes levels below this turning point, there is diversification at both the extensive and intensive margins. Importantly, for the low and middle income countries (i.e., with GDP per capita below 14'000 dollars PPP) diversification occurs mostly along the extensive margin.[2] The intensive margin dominates thereafter. For incomes levels above the turning point, we observe a re-concentration of exports towards fewer products. Such hump-shaped diversification curve is consistent with the conjecture that countries travel

[2] PPP $ 14,000 is roughly the World Bank threshold for high income country.

across diversification cones as discussed in Schott (2003, 2004) and Xiang (2007).

We also find that if the share of raw materials is a significant contributor to export concentration, its inclusion in regressions does not affect the turning point or the significance of income levels, suggesting that the non-monotone path of diversification is an inherent feature of the economic development process (rather than a reflection of the predominance or not of primary-product exports). Moreover, we evidence that public infrastructure contributes to export diversification but only along the intensive margin.

The paper is organized as follows. Section 2 presents prima-facie (descriptive) evidence on traditional and new export products. Section 3 reports econometric evidence on the stages of export diversification along the economic development process. In order to better understand what is behind the hump-shaped diversification curve evidenced in preceding section, Section 4 analyses action along the "intensive" and "extensive" margins by (i) comparing changes in export concentration indices and number of active lines and (ii) analyzing the evolution of the "within" and "between" component of the Theil concentration index. We argue that results are consistent with the conjecture that countries travel across diversification cones. Section 5 explores other potential explanations of the diversification process curve. Section 6 concludes.

2. Prima-facie evidence

2.1 Measures of export concentration/diversification

Our dataset comprises data on trade, income per capita, and public capital. The export data is taken from UNCTAD's COMTRADE database at the HS6 level of

disaggregation (4'998 lines).[3] The baseline sample covers 159 countries representing all regions and all levels of development between 1988 and 2004 (17 years), including 121 developing countries, i.e. non high-income countries as defined by the World Bank (incomes per capita roughly under 2006 US$10'000). See appendix A.1 for a detailed sample composition. Taking out missing year data the usable sample has 1'574 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 zero and one, is

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

where $s_k = x_k / x$ is the share of export line k in total exports and n is the number of export lines (omitting country and time subscripts).[4]

Theil's entropy index (Theil 1972) is given by

[3] 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 (sub-headings, numbering 4'998 according to the HS 1989-92 nomenclature). Further degrees of disaggregation (HS 8, 10 and beyond) are not harmonized across members of the World Customs Organization and require extremely cautious handling. 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. As a result, an HS8 code may correspond to a good at time t, to another good at time t+1, and back to the same good at time t+2. This problem also affects US trade data compiled by Feenstra in the NBERTD (see Feenstra 1997 and Feenstra, Romalis and Schott 2002). Eurostat HS10 data is not publicly available.

[4] Note that 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. Thus, n = 4'998 (according to the HS 1989-92 nomenclature) for every country and year.

$$T = \frac{1}{n}\sum_{k=1}^{n}\frac{x_k}{\mu}\ln\left(\frac{x_k}{\mu}\right) \quad where \quad \mu = \frac{\sum_{k=1}^{n} x_k}{n} \qquad (2)$$

For Gini indices, we use Brown's formula; that is, for each country and year, we first sort export lines, indexed by *k*, by increasing order of trade value *x* so that $x_k < x_{k+1}$. Cumulative export shares are

$$X_k = \sum_{\ell=1}^{k} x_\ell \Big/ \sum_{\ell=1}^{n} x_\ell \qquad (3)$$

and cumulative shares in the number of export lines are simply *k/n*. Brown's formula for the Gini coefficient is then

$$G = \left|1 - \sum_{k=1}^{n}\left(X_k - X_{k-1}\right)(2k-1)/n\right|. \qquad (4)$$

Table 1 shows descriptive statistics for these indices.

Table 1
Descriptive statistics – 159 countries over 1988-2004

Observe that Gini coefficients are very high, corresponding to Lorenz curves that are almost right-angle ones. This contrasts with those calculated by Imbs and Wacziarg (2003) on production and employment (typically around 0.5, see their Table 1). The reason has to do with the level of disaggregation rather than with any conceptual difference between trade, production and employment shares. Whereas Imbs and Wacziarg calculated their indices at a relatively high degree of aggregation (ILO 1 digit, UNIDO 3 digits and OECD 2 digits) 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 so of course for developing

countries but even for industrial ones). The reason for this pattern is that the harmonized system used by COMTRADE is derived from nomenclatures originally designed for tariff-collection purposes rather than to generate meaningful economic statistics. Thus, it has a large number of economically irrelevant categories e.g. in the textile-clothing sector while economically important categories in machinery, vehicles, computer equipment etc. are lumped together in "mammoth" lines. High Gini indices are thus to be expected for all countries. Note that we are interested here in the evolution of the Gini index and not in its level.

As for the average number of "positive" export lines –active lines with non-zero trade values– it is relatively low at 2'492 per country per year, i.e. a little less than half the total, with a minimum of 13 for Kiribati in 1993 and a maximum of 4'957 for the United States in 1994. 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 2000 Purchasing Power Parity (PPP) dollars for comparability. The last line of Table 1 report descriptive statistics for an index of public infrastructure capital which we use in the regression analysis. It is a composite of fixed-line telephone density (number of lines per thousand inhabitant), railroad density (km per inhabitant), road density (km per inhabitant), and the proportion of paved roads. Data is from the WDI and individual components were combined into a single index using principal components analysis.[5]

[5] Missing data was completed by linear or geometric interpolation and limited extrapolation. As extrapolation often resulted in overshooting compared to trends, the choice between linear and geometric was based on minimization of extreme values. We are grateful to Claudio Sfreddo for making this data available to us.

2.2 Defining "new products"

"New products" (i.e. lines at the HS6 level) for a year and country are defined in our database as those that were not active in the country's export trade in the preceding two years but were exported in each of the following two years (the definition is thus based on a moving 5-year sub-sample). This reduces the sample of "new products" to 1990-2002, two years being taken out at both ends. Our definition differs from that used by Klinger and Lederman (2004) who define "discoveries" (the equivalent of our "new products") as products that represent more than US$1 million of exports per year in the latter part of their sample (1999-2002) and less than $10'000 in the beginning (1992-1993). By their definition, there were a total of 1'710 discoveries at the HS6 level over the whole sample period, whereas we have on average 57 new products per country per year (see table 1), i.e. a total of 51'626 "new products" (new for a country and a year, not in the absolute) for the entire sample period.

Why the difference? Conceptually, our notions of new 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, leading to short-lived export "spells". The shorter those spells, the more discoveries or new products there should be, as new entrepreneurs try again a few months or years later, incurring the sunk cost of reaching foreign markets anew.[6]

Detailed evidence on the length of export spells and on product turnover in international trade was recently analyzed by Besedes and Prusa (2006a) using

[6] On this, see Roberts and Tybout (1997), who found that the probability that a firm is active in export markets depends on its status the previous year but not further back, suggesting very rapid decay of incumbency advantages (information, networks etc.).

the Feenstra, Romalis and Schott (2002) database for the US.[7] Strikingly, they find that over half of all trade relationships (defined as nonzero export lines for a given exporting country, the importing country being always the US) are observed for a single year, while 80% are observed for less than five years. Survival analysis shows a rapidly decreasing hazard rate, suggestive of two regimes: rapid failure vs. long-term success. These numbers indicate very rapid turnover in international trade, a finding that is quite consistent with the entrepreneurial-search view of Vettas or Hausmann and Rodrik.

By aggregation, HS6 data are likely to smooth some of these entries and exits (though Besedes and Prusa's results seem robust to at least some aggregation), so one would want to err on the side of too many new products rather than too few. In addition, they find shorter median spells for Southern exporters (two years) than for Northern ones (six years), so our data is likely to be characterized by high unobserved turnover.[8] Finally, we treat two successive export spells in the same product line for the same country as two new products. The reason is that the product marketed by second-timers after an initial failure may not be –indeed, is unlikely to be– identical to what was tried by the first-timers, lest it would be likely to fail again; thus, two spells in the same HS6 line, treated as two new products in our definition, are indeed likely to be two new products, not one.

This said, the number of new products should be interpreted somewhat cautiously, as they do not necessarily represent true entrepreneurial "discoveries". First, as discussed in an earlier footnote, at very high levels of disaggregation such as HS8 or HS10, there is constant reclassification of

---

[7] This database is an extension to 2001 of Feenstra's database on US trade at the HS10 level of aggregation. See Feenstra, Romalis and Schott (2002).
[8] This however must be interpreted cautiously. If some of the apparent failures are simply measurement errors (unrecorded trade), smaller trade volumes are more likely to be censored in a way that cannot be detected. This is likely to affect developing countries whose export volumes are low (see below and the discussion in Besedes and Prusa 2006a).

products across HS codes, giving rise to artificial births and deaths.[9] Second, among the countries with the highest number of "new-product" lines thus defined, one finds transition countries whose trade statistics were gradually put in place during the 1990s, such as Romania (with 1'331 lines in 1991).[10] "New products" in those cases may well be discoveries of their country's statistical office only. Third, one also finds very poor countries whose trade statistics are particularly erratic and report zero trade as a result of mismeasurement, such as Zambia.[11] In that case what looks like two spells may be one with non-recorded trade in the middle. Thus our definition, which requires two zero-trade years instead of one to end a spell, strikes a balance between the very conservative one used by Klinger and Lederman (2004) and the very liberal one used by Besedes and Prusa (2006b).

2.3 Are 'new products' any different from others?

Table 2 gives a characterization of export goods using Rauch's index of product differentiation. Rauch (1999) distinguished between products traded on organized exchanges such as the LME, products with reference prices (listed in widely available publications like the *Knight-Ridder CRB Commodity Yearbook*), and differentiated products whose prices are determined by

[9] In their survival analysis, Besedes and Prusa (2006a) chose to treat reclassifications as censored observations; that is, a spell of, say, five years ending with a reclassification is treated as a spell of *at least* five years, like a spell at the end of the sample.
[10] Romania also figures prominently in Klinger and Lederman's (2004) discoveries (see their Table 2).
[11] Trade statistics are seriously error-prone in poor countries. The data is provided by UNCTAD's member states and is typically compiled by national statistical offices and reviewed by Trade Ministries on the basis of raw data provided by Customs administrations. Under automated systems such as ASYCUDA, data is increasingly entered in computer systems directly by employees of transit companies, sometimes resulting in input errors. Many Least Developed Countries have benefited in recent years from technical-assistance programs designed to raise the awareness of customs administrations to the need to provide government authorities with reliable data and improving their capacity to do so, but progress is slow.

branding.[12] Rauch's classification is likely to be of importance for our analysis as Besedes and Prusa (2006a) found that export spell lengths are significantly lower for homogenous goods than for reference-priced and differentiated ones. This suggests that, in accordance with intuition, search costs are higher for the latter than for the former, leading to more stable trading relationships. One would thus expect entrepreneurs in poor countries to establish the kind of trade networks needed to export differentiated goods only progressively, leading over time to more stable export patterns. This, in turn, would suggest a higher share of differentiated among new products than among traditional ones. Table 2 shows the proportion of each of Rauch's categories in traditional and new export lines.

Table 2
Characterization of products by degree of differentiation

We find a lower share (in terms of export value) of homogenous-product exports among new lines (according to our definition) than among traditional ones (11.4% vs. 22.3% using Rauch's "conservative" classification and 17% vs. 29.3% according to his "liberal" classification). The reverse is true of "reference-priced" goods, the third category (differentiated goods) having similar shares in new and traditional products. Thus, products that appeared in developing countries' exports in the 1990s were no more differentiated than products they had been exporting thus far, suggesting a general failure to establish efficient networks.

[12] Rauch argued that finding markets for differentiated goods involves a sequential search for trading partners that can be long and costly and will in all likelihood involve networks based on ethnic, linguistic or other factors of proximity. Exporting products listed on organized exchanges, by contrast, (or, to a lesser extent, reference-priced products) involves anonymous markets and hence lower search costs. Evidence from a gravity equation supported this view.

3. Stages of diversification: Estimation

3.1 Parametric evidence

Although Tables 3 to 6 report estimation results for both concentration indices and the number of active export lines, we will first limit the discussion to the former and then turn to the latter and their co-movement.

Table 3 explores the turning point's stability across different definitions of GDP per capita. The first bloc (columns (1)-(4)) uses per capita GDP at PPP from the WDI; the second (columns (5)-(8)) uses per capita GDP at PPP from the Penn World Tables; and the third (columns (9)-(12)) uses GDP per capita in constant US dollars from the WDI. Estimates are from pooled OLS, with White-corrected standard errors. Using WDI definition gives a turning point around $26'000 while other definition of GDP provide turning points between $20'000 and $28'000.[13]

Table 3
Quadratic regression results, pooled OLS

In Figure 1, fitted curves show predicted values of Theil index together with the predicted number of active export lines.[14] The latter, which is concave and increasing at the origin, are easy to distinguish from the former, which is convex and decreasing at the origin.

Figure 1
Predicted Theil's concentration index & number of active export lines

[13] The turning point is much higher when the Gini index is used instead of Theil or HHI indices. Using more sophisticated econometric techniques (i.e., logistic transformation, Generalized Moments Method), which correct for most biases, gives turning points roughly at the same level of GDP per capita for all measures of concentration (Table 6).
[14] Fitted curves using Herfindahl and Gini indices have similar shape. We present the fitted curve for the Theil index as this measure is further studied (i.e., decomposed) in section 4.2.

The curves shown are fitted using the quadratic polynomial regressions discussed in the previous paragraph.[15] One issue is whether the turning point is driven by microstates and island economies, which could have middle-range per capita GDPs and at the same time be very concentrated –say, in bananas or fish products. The first two columns of Table 4 (using the WDI definition, to which we will stick from now on) show that excluding 24 countries with populations below one million shifts the turning point forward to $23'000 at PPP. As microstates are potential outliers, we omit them in the rest of the analysis.

Table 4
Results without microstates

Our turning point is substantially higher than that found by Imbs and Wacziarg for production ($14'600 in 1996 dollars, or about 16'500 in constant 2000 dollars) but quite similar to what Klinger and Lederman (2005) found for exports on a panel of 130 countries over 1992-2003 ($22'500 in constant 2000 dollars).

An additional issue is whether our result is driven by 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. The first block of Table 5, which reports pooled estimates with time effects, shows that the turning point is unaffected.

Table 5
Pooled, within and between estimates

[15] We also estimated "smoother" regressions which consist of re-estimating the regression for overlapping samples centered on each observation. Smoother regressions impose no functional form and are therefore suited to the exploration of highly non-linear relationships. However, because the overlapping samples get smaller near the bounds of the estimation interval, the shape of smoother-regression curves near those bounds can be unduly affected by a small number of observations, resulting in either artificially-generated or undetected turning points. Hence, because our turning point comes at a high level of income, this non parametric method is not relevant here.

Second, given the panel structure of our data set, a natural question is which type of estimator –within, between, random-effect or pooled– should be used. Imbs and Wacziarg estimated their production turning point using fixed effects; however, their sample was long in the time dimension (1969-1997) whereas ours has only 11 years per country (with a minimum of 2 years and a maximum of 17–see appendix A.1) With such a short time dimension, estimating the turning point on the basis of the within-country dimension only would be of debatable value. Indeed, the second block of Table 5, which reports estimates with time and country fixed effects, shows no turning point at all. By contrast, the third block, which reports between estimates, has the usual turning point.[16]

Table 6 reports a number of robustness checks. One issue has to do with censoring: in order to take account of the fact that Gini coefficients are bounded left and right at zero and one respectively –although neither is binding *stricto sensu*– the first bloc reports estimates from a logistics transformation. The result is to make the turning point appear at the usual level of about $23'000. The second has to do with the potential endogeneity of GDP per capita to export concentration. As we have no valid outside instrument for GDP per capita, the table's last block –columns (4) to (7)– shows system Generalized Moments Method (GMM) estimation results, with a turning point varying between $19'524 (Herfindahl) and $24'500 (Gini).[17]

---

[16] A natural way of combining the within and the between dimensions of the data would be to use random effects, but a Hausman test rejects the null of no correlation between GDPs and country random effects. The natural fix for such a problem would be to use instrumental-variable techniques such as Hausman-Taylor's (1981), but with only one RHS variable (GDP and its square) instrumentation is not possible.

[17] Blundell and Bond's (1998) system GMM estimator uses lagged differences as instruments for current levels and lagged levels for current differences. As is well-known, a crucial issue when using GMM and especially system GMM is the number of instruments, which 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.

Table 6
Robustness

Thus, by and large both the existence of a turning point in export concentration and its location around a GDP per capita of about $20'000-23'000 at PPP –a very late point in the development process– are fairly robust.

3.2 Number of active lines and "new" products

A glance at the columns entitled "Nber" in Tables 3-6 shows that there is 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 being always roughly at the same level of GDP per capita as that of Herfindahl and Theil indices (see also figure 1). This applies as well to column (3) of Table 6, which reports negative binomial estimation to take into account the fact that the number of lines is a count variable. [18] The rising part of the curve corresponds to the introduction of new products as countries develop (see more evidence below). Its decreasing part illustrates one of the striking findings of this paper –namely, that high income countries tend to "close down" export lines faster than they open up new ones, resulting in re-concentration at the extensive margin. We will return to this point later on.

Figure 2 shows the predicted number of "new" export lines (per country-year, defined as per section 2.2 above) against GDP per capita. The curve is fitted with the usual quadratic polynomial.[19] Its turning point comes very early –at PPP$4'150– as in Klinger and Lederman, in spite of our different definitions of

[18] Note that, for active lines, we have one additional reason for using time effects –namely, that their number can be affected by commodity reclassifications (see footnote 1) as those happen even at the HS6 level (e.g. goods that are suddenly reported separately though they were previously lumped together in the 'not elsewhere specified' category). As reclassifications cannot be identified directly, we follow Klinger and Lederman (2005) and control for them simply by year dummies.
[19] This specification is justified by the non-parametric ("smoother") estimates, also reported in figure 3.

"new products". 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 five thousand 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 $4'150.[20]

Figure 2
Predicted New Exports: non-parametric & quadratic estimates

Figure 3 compares the number of new export lines and their average value (per line) against GDP per capita (non-parametric "smoother" estimates). It can be seen that, if the number of new products peaks at $4000 per capita, their value per line shoots up very late in the development process, largely after $25'000 per capita, so the increase in value concerns only a few countries.

Figure 3
New products and their average value: nonparametric curves

Thus our analysis, using regressions of concentration indices as well as number of active line on GDP per capita, evidences a hump-shaped relationship between economic development and export diversification. Our next task is to understand what is behind the hump.

4. Stages of diversification: "extensive" *vs.* "intensive" margins

That export diversification would proceed in parallel with economic development is something 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,

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

this becomes easier. But the later re-concentration, 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 now turn to a systematic analysis of the "intensive" and "extensive" margins. By these, we mean respectively variations in trade values for existing products and variations in the number of active lines. The analysis will be carried out using Theil's index because of its decomposability properties.

4.1 Number of active export lines *vs*. concentration indices

Using alternatively indices of concentration and the number of active product lines as the dependent variable (the explanatory variable being GDP per capita), four broad patterns are possible: (i) with both concentration indices and the number of active lines rising with GDP per capita, there is diversification along the extensive margin and concentration along the intensive one; (ii) with concentration indices decreasing and the number of active lines rising, there is diversification along the extensive or both margins; (iii) with concentration indices rising and the number of active lines decreasing, there is concentration along the extensive or both margins; (iv) with concentration indices and the number of active lines decreasing, there is concentration along the extensive margin (retrenchment), but diversification among existing products.

According to results in Tables 3-6, only scenarii (ii) and (iii) are relevant. At income levels below the turning point, concentration indices decrease and the number of active lines increases suggesting diversification along the extensive *or both* margins. Similarly, at income levels above the turning point, there is concentration along the extensive *or both* margins. In order to determine the type of diversification (extensive *vs*. intensive) that predominates during the

development process, we thus need to further analyse the concentration indices through decomposition.

4.2 Digging deeper: Theil decompositions

Let us now look at concentration measures within and between three groups of products indexed by *j* (each group being country-specific): traditional ones (exported by the country since the beginning of the sample), new ones (as defined in Section 2.2), and non-traded ones (whose exports are zero in the whole sample period for that country). Using these groups, we decompose Theil's index into a "within" component

$$\begin{aligned} T^W &= \sum_{j=1}^{J} \frac{n_j}{n} \frac{\mu_j}{\mu} T^j \\ &= \sum_{j=1}^{J} \frac{n_j}{n} \frac{\mu_j}{\mu} \left\{ \frac{1}{n_j} \sum_{k \in j} \frac{x_k}{\mu_j} \ln\left(\frac{x_k}{\mu_j} \right) \right\} \end{aligned} \qquad (5)$$

and a "between" component

$$T^B = \sum_{j=1}^{J} \frac{n_j}{n} \frac{\mu_j}{\mu} \ln\left(\frac{\mu_j}{\mu} \right) \qquad (6)$$

where $T^W + T^B = T$. In (5) and (6), n_j is the number of export lines in group *j* and μ_j is the group's average export value, in dollars.

The "within" component captures the concentration of exports within groups. Note however that it is *not* a size-weighted average of the group-specific Theil indices T^j, because size weights $\omega_j = n_j / n$ are multiplied by the ratios of group means (of export values) to the sample's overall mean (μ_j / μ). Thus, when the mean value of exports goes up for a country, the within component of Theil's index goes down mechanically even if all group indices T^j are unaffected. Second, observe that zero-export lines must be excluded from the within's

calculation because of the log term. This means that all action along the extensive margin will be reflected in the index's between component only. In our case, because traditional products account for 99% of all products and means don't differ that much between traditional and new products, the index's within component is largely dominated by the traditional products' group index, so we will use the latter as a (more intuitive) approximation of the within index.

Unlike the within, the "between" component does not involve individual values, being a function of group means and sizes only. It is zero whenever average export values are equal across all groups, irrespective of their distribution inside groups, and positive if and only if group means differ.

Figure 4
"Within" and "between" components of Theil's index

Figure 4 depicts the contribution of the between and within component to the Theil. Observe that in levels, the "within" component dominates the index. But in terms of evolution, most of the action is in the between component, which shrinks to almost zero at the index's turning point and starts rising again thereafter.

The rapid decrease in the between component is what drives diversification at low- to middle- income levels (below PPP$14'000, which roughly corresponds to the World Bank's high-income threshold), although the within also shrinks.[21] That is, diversification occurs mostly at the "extensive" margin, meaning convergence in average export values across groups. Two effects contribute to this. First, as more product lines become active in export, the size of the group made of inactive export lines shrinks; because this group has a very different mean from the other two, its shrinkage mechanically reduces the between component. Second, new products are launched at higher scales, approaching

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

that of traditional ones. If, following our earlier discussion, one thinks of new products as "entrepreneurial gambles", what we see here is export entrepreneurs taking increasingly large gambles, possibly reflecting the better information and lower risks of multinational companies when they set up relatively large, export-oriented factories in low-income countries (say in the textile sector).

After PPP$14'000 and until the index's turning point at around PPP$24,000, the between component still shrinks but slower than the within, so diversification is mostly along the intensive margin. That is, individual export values converge, mostly among traditional products since, as noted, those account for 99% of all active export lines. Put simply, all exports are at an industrial scale.

Finally, above PPP$24'000, re-concentration occurs at the extensive margin since the index's rise is driven almost entirely by the between component, the within remaining more or less flat. Recall that the between rises when group means diverge; here, this divergence can conceptually be due to two forces. First, the number of inactive product lines start rising again as rich countries "close down" export lines in sectors where they lose comparative advantage (textile and the like), as shown in Table 7; second, the average value of new products (computers and the like) shoots up, as we saw in Figure 3. As it turns out, most of the action is in the first force –the lines closed down.

Table 7
Cumulated "Closed" lines, 2001-2003, GDP per cap. above PPP$24'000

Appendix Table A.2 lists the chapters with the highest number of closed lines. Chemicals (Chapter 29) and Leather (Chapter 41) are among the most active "closers" (5.4% of the lines for the former, 29.5% for the latter). In both cases, the closed lines' value is low (about 0.03% of the total). But note that, at the same time, Figure 5b shows that high-income countries specialize in chemicals.

The simultaneous occurrence of rising specialization and line closures in the chemical sector is consistent with Schott (2004)'s finding that specialization occurs *within* sectors, as high-tech exports replace low-tech ones when countries grow. The closure of export lines in the leather sector, by contrast, suggests between-product specialization, as leatherworks are a labor-intensive industry in which countries lose comparative advantage when they grow.

4.3 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 kind of support they may get. During the transition phase, then, new-cone lines become active while old-cone ones don't 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 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 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 Hausman and Rodrik's 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 Figure 4 [i.e. 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. 693-6). Apparel and footwear could indeed be slow-dying industries in many

countries, not only on the import-competing side but also on the export side (the EU for instance is still today 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 5c and Figure 5d, 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 isn't much action any more as both textiles and machinery stabilize at low (5%) and high (30%) shares respectively.

5. Stages of diversification: alternative explanations

As made clear by the Theil index's decomposition, explanations of trade diversification should allow for the extensive/intensive margin interpretation and for the conjecture of a slow adjustment across diversification cones.[22] We must however consider alternative explanations which could *artificially* create or reinforce a hump-shaped pattern. The diversification curve may e.g. result from spurious statistical effects rather than reflecting meaningful economic

[22] Recall: low to middle income countries diversify mostly along the extensive margin whereas high income countries diversify along the intensive margin and eventually re-concentrate their exports towards fewer products.

effects linked to the development process. Alternative explanations include (i) the discrepancy in primary-resource exports emphasized by the between-country aspect of the database, (ii) the structure of the HS6 COMTRADE classification, and (iii) the uneven levels of public infrastructure across countries.

5.1 Primary products

Given that a substantial chunk of the U-shaped pattern of export concentration evidenced in section 3 is generated by the between-country dimension of the data, a likely candidate for the underlying cause is the prevalence of primary resources in poor-country exports. Figure 5 shows selected sectoral shares against GDP per capita.

Figures 5a-5e
Selected sectoral shares against GDP per capita

It can be seen that for many sectors the data shows substantial heterogeneity with large outliers. For minerals (HS section 5) there is a fairly distinct pattern whereby large exporters of mineral products (those for which mineral products represent over 20% of exports) are either low/middle income countries (below $12'000) or very high-income ones (above $25'000). This pattern, which is confirmed by the non-parametric regression curve, is of course likely to contribute to the U-shaped pattern of export concentration.

In order to verify the conjecture that primary products contribute to the U-shape of export concentration on our dataset, we ran our usual quadratic-polynomial regressions controlling for the share of raw materials in exports, as proxied by the share of HS chapters 26 (ores, slag and ashes) and 27 (mineral

fuels, mineral oils and products of their distillation).[23] Results are shown in Table 8.

Table 8
Estimates with raw-material export shares

Unsurprisingly, the share of raw materials comes out as a positive and significant contributor to export concentration and as a negative one to the number of active lines (columns (4) and (8)). There is thus evidence of concentration and of some degree of Dutch disease. But the striking result is that coefficients on GDP per capita and its square are not affected by much; nor is the turning point.

As a further exercise, we interact the share of raw materials in exports with GDP per capita (second block of Table 7). We plot in Figure 6 predicted Theil indices against GDP per capita for various levels of raw-material export shares.

Figure 6
Predicted Theil indices against GDP per capita and the share of raw materials in export

Except for very high values of the share of raw materials (over 70%), the U-shaped relationship is maintained with an almost unchanged turning point.

5.2 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

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

number of economically irrelevant categories (e.g. the textile-clothing sector –section 11), whereas in other sections (e.g. machinery –section 16, or transport equipments –section 17) economically important categories are lumped together in a few lines. Figure 7, which plots, for each section of the HS6 classification, total export value *versus* number of lines provides evidence of such feature. Sections 16 and 17 are well above the 45° line, reflecting a disproportionate high value per export line, while section 11 includes a large number of small lines.

Figure 7
Shares Value/number of lines by section weighted average

Now, assume that products in section 11 are essentially exported by middle income countries whereas products in sections 16 and 17 are essentially exported by high income countries (assumptions confirmed by Figures 5a, 5d and 5e respectively). Then, the observed diversification/re-concentration pattern could be an *illusion* caused by the structure of the HS6 classification.

In order to verify this conjecture, we include controls for the export share of sections 5, 11, 16 and 17 (per country-year) in the regressions. The curve's shape and its turning point are robust to the introduction of these controls. Thus, the hump-shaped relationship between economic development and export diversification is thus not due to a spurious "composition" effect.[24]

5.3 Public infrastructure capital

Active infrastructure policies could also influence the diversification process. For instance, better infrastructure can affect a country's ability to export manufactured products more than its ability to export raw materials (say because mining companies know how to set up their own, private infrastructure). Table 9 shows regressions results including our measure of

[24] Results are available upon request.

public infrastructure capital described in section 2. It can be seen that infrastructure capital contributes to reduce export concentration along the intensive margin (concentration coefficients, columns (5)-(7)) but not the extensive one (one active lines, column (8)). This is somewhat disappointing, suggesting only limited scope for supply-side policies as vehicles to encourage export entrepreneurship. The turning point, however, remains, again, unchanged.

Table 9
Effect of public infrastructure capital

It should be kept in mind, however, that public-infrastructure measures are very rough proxies based on fragmentary data. The role of supply-side policies in export diversification clearly deserves more scrutiny.

6. Concluding remarks

The results presented so far suggest two observations and one caveat. 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). The re-concentration of exports above a threshold of PPP$24'000 is especially striking. Diversification occurs mostly at the extensive margin for low- to middle-income countries, 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. do.

The second observation is that diversification at high to very 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. The hump-shaped relationship between diversification and development may be explained by this slow adjustment and thus corresponds to traveling across diversification cones.

The caveat is that diversification can be the by-product of two policy distortions. First, support for declining industries may be the reason for over-diversification during the transition phase. Second, and perhaps more importantly, trade diversion fostered by preferential agreements like Mercosur or NAFTA can result in welfare-reducing export diversification. Sanguinetti et al.'s (2001) findings are definitely suggestive of this, and more work is needed in this direction.

References

Arellano, M., and S. Bond (1991), “Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations”, Review of Economic Studies 58, 277–97.

Besedes, T. and T. Prusa (2006a), “Surviving the U.S. Import Market: The Role of Product Differentiation”, Journal of International Economics, 70(2), 339-358.

– and – (2006b), “Ins, Outs, and the Duration of Trade”, Canadian Journal of Economics 39, 266-95.

Blundell, R., and S. Bond (1998), “Initial conditions and moment restrictions in dynamic panel data models”, Journal of Econometrics 87, 11–143.

Cleveland, W. (1979), “Robust Locally Weighted Regression and Smoothing Scatterplots”; Journal of the American Statistical Association 74, 829-836.

Cuddington, John; R. Ludema, and S. Jayasuriya (2001), “Prebisch-Singer Redux”; mimeo, The World Bank.

Feenstra, R. (1997), “NBER Trade Database, Disk 1:US Imports, 1972-1994”, NBER Working Paper 5975.

–, J. Romalis and P. Schott (2002), “US imports, Exports and Tariff Data, 1989-2001”, NBER working paper 9387.

de Ferranti, David; G. Perry, D. Lederman and W. Maloney (2002), From Natural Resources to the Knowledge Economy; The World Bank.

Gylfason, T. (2004) "Natural Resources and Economic Growth: From Dependence to Diversification," CEPR Discussion Papers 4804

Hausman A. and E. Taylor (1981), “Panel data and unobservable individual effects”, Econometrica, 49 (6), November, 1377-1398.

Hausmann, R. and D. Rodrik (2003), “Economic Development as Self-Discovery”; Journal of Development Economics 72, 603-633.

–, J. Hwant and D. Rodrik (2005), “What You Export Matters”; NBER working paper 11905.

– and B. Klinger (2006), “Structural Transformation and Patterns of Comparative Advantage in the Product Space”; mimeo, Harvard University.

– and – (2007), “The Structure of the Product Space and the Evolution of Comparative Advantage”; CID working paper 146, Harvard University.

Herzer, Dierk (2004), “Export Diversification, Externalities and Growth”; University of Göttingen Discussion Paper #99.

Hidalgo, Cesar; B. Klinger, A.-L. Barabasi and R. Hausmann (2007), “The Product Space Conditions the Development of Nations”; mimeo.

Hummels, David, and P. Klenow (2005), “The Variety and Quality of a Nation’s Exports”; AER 95, 704-723.

Imbs, J. and R. Wacziarg (2003), “Stages of Diversification”; American Economic Review 1993, 63-86.

Klinger, Bailey (2007), “Development and the Topology of the Product Space”; paper prepared for the UNCTAD Conference on Trade and Development, Geneva, November 2007.

– and D. Lederman (2004), “Discovery and Development: An Empirical Exploration of ‘New’ Products; mimeo.

– and – (2005), “Diversification, Innovation, and Imitation off the Global Technology Frontier”; mimeo.

Prebisch, Raúl (1950), “The Economic Development of Latin America and its Principal Problems”; reprinted in Economic Bulletin for Latin America 7, 1962, 11-22.

Rauch, J. (1999), “Networks versus Markets in International Trade”; Journal of International Economics 48, 7-35.

Roberts, M. and J. Tybout (1997), “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs”, American Economic Review 87, 545-64.

Roodman, D. (2006), “How to Do xtabond2: An Introduction to “Difference” and “System” GMM in Stata”, Center for Global Development, Working Paper Number 103, Decembre.

Sachs, Jeffrey, and A. Warner (1995), “Natural Resource Abundance and Economic Growth”; NBER working paper #5398.

Sanguinetti, Pablo; J. Pantano and J. Posadas (2001), “Trade Liberalization and the Dynamics of the Trade Structure in Argentina and Uruguay”; mimeo, The World Bank.

Schott, P. (2003), “One Size Fits All? Heckscher-Ohlin Specialization in Global Production”; AER 93, 686-708.

– (2004), “Across-product versus Within-product Specialization in International Trade”; QJE 119, 647-678.

Singer, H. W. (1950), “US Foreign Investment in Underdeveloped Areas: The Distribution of Gains Between Investing and Borrowing Countries”, AER P&P 40, 473-485.

Theil, Henri (1972), Statistical Decomposition Analysis; North Holland.

Vettas, N. (2000), “Investment Dynamics in Markets with Endogenous Demand”, Journal of Industrial Economics 48, 189-203.

Xiang, C. (2007), “Diversification cones, trade costs and factor market linkages”, Journal of International Economics 71, 448-466.

Tables and figures

Tables

Table 1
Descriptive statistics – 159 countries over 1988-2004

Variable	Obs	Mean	Std. Dev.	Min	Max
Export concentration indices:					
Gini	1'574	.959	.045	.793	.999
Herfindahl	1'574	.131	.183	.003	.987
Theil	1'574	4.392	1.669	1.589	8.461
Nber of active lines	1'574	2'492	1'630.6	13	4'957
Nber of new export lines a/	912	56.61	67.60	0	1'151
GDPpc, const. 2000 US$	1'574	7'324.8	9'501.5	106.09	48'419.3
GDPpc, PPP	1'545	10'247.8	9'488.8	486.47	64'298.64
Share of oil in exports	1'574	.129	.230	0	.996
Public infrastructure capital	790	1.018	.622	.019	2.348

a/ according to the "new" export lines definition (see section 2.2), the sample is reduced to (i) 1990-2002, two years being taken out at both ends and (ii) the 125 countries with available data for at least 5 consecutive years.

Table 2
Characterization of products by degree of differentiation

	All products	New products b/	World trade, 1990 (Rauch) c/
Conservative classification a/			
Homogenous	22.35	11.39	12.60
Reference priced	26.43	38.91	20.30
Differentiated	51.21	49.70	67.10
Liberal classification a/			
Homogenous	29.29	16.96	16.00
Reference priced	21.92	36.60	19.50
Differentiated	48.80	46.44	64.20

Notes
a/ Because the classification of some products cannot be asserted unambiguously, Rauch's conservative classification assigns fewer products to the "homogenous" and "reference-priced" categories than his liberal ones.
b/ According to the definition in the text
c/ From Table 2 of Rauch (1999)

Table 3
Income levels: pooled OLS, 1988-2004

	GDPpc, PPP in constant 2000 intern. $, WDI				GDPpc, PPP in constant 2000 intern. $, PWT				GDPpc, in constant 2000 US$, from WDI			
	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12
Dependant	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber
GDPpc	-1.14E-05	-0.000199	-4.40E-06	2.33E-01	-1.55E-05	-0.000228	-4.84E-06	2.58E-01	-9.80E-06	-0.000181	-5.54E-06	2.13E-01
	5.68***	12.86***	9.99***	19.41***	7.27***	13.56***	10.53***	19.07***	6.75***	14.51***	13.99***	20.65***
GDPpc²	2.11E-10	3.71E-09	4.43E-11	-4.27E-06	3.81E-10	5.10E-09	7.04E-11	-5.36E-06	1.71E-10	3.47E-09	9.60E-11	-4.03E-06
	3.35***	7.10***	2.72***	11.09***	5.44***	8.40***	4.04***	11.00***	3.72***	8.33***	7.31***	12.65***
Turn. Point ($)	27014	26873	49661	27280	20341	22353	34375	24029	28655	26066	28854	26384
R2	0.17	0.29	0.39	0.42	0.09	0.25	0.32	0.39	0.07	0.24	0.35	0.35
obs.	1545	1545	1545	1545	1540	1540	1540	1540	1574	1574	1574	1574
Nber countries	155	155	155	155	154	154	154	154	159	159	159	159
period	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004
	Countries on the right of the turning point in 2004											
	Australia	Australia	None	Australia	Australia	Australia	Norway	Australia	Denmark	Denmark	Denmark	Denmark
	Austria	Austria		Austria	Austria	Austria	US	Austria	Iceland	Hong Kong	Iceland	Hong Kong
	Belgium	Belgium		Belgium	Belgium	Belgium		Belgium	Japan	Iceland	Japan	Iceland
	Canada	Canada		Canada	Canada	Canada		Canada	Norway	Ireland	Norway	Ireland
	Denmark	Denmark		Denmark	Cyprus	Cyprus		Denmark	Sweden	Japan	Sweden	Japan
	Finland	Finland		Finland	Denmark	Denmark		Finland	Switzerland	Norway	Switzerland	Norway
	Hong Kong	France		Hong Kong	Finland	Finland		France	US	Sweden	US	Sweden
	Iceland	Hong Kong		Iceland	France	France		Germany		Switzerland		Switzerland
	Ireland	Iceland		Ireland	Germany	Germany		Hong Kong		UK		US
	Netherlands	Ireland		Netherlands	Hong Kong	Hong Kong		Iceland		US		
	Norway	Japan		Norway	Iceland	Iceland		Ireland				
	Sweden	Netherlands		Switzerland	Ireland	Ireland		Japan				
	Switzerland	Norway		UK	Israel	Italy		Netherlands				
	UK	Sweden		US	Italy	Japan		Norway				
	US	Switzerland			Japan	Netherlands		Singapore				
		UK			Netherlands	NZ		Sweden				
		US			NZ	Norway		Switzerland				
					Norway	Singapore		UK				
					Singapore	Sweden		US				
					Slovenia	Switzerland						
					Spain	UK						
					Sweden	US						
					Switzerland							
					UK							
					US							

Absolute value of robust t statistics under coefficients (White's correction for heteroskedasticity used)
***, **, * significant at respectively 1%, 5% and 10% level.

Table 4
Results without microstates

	(1)	(2)	(3)	(4)
Dependant	HHI	Theil	Gini	Nber
Method	Pooled	Pooled	Pooled	Pooled
GDPpc	-1.59E-05	-0.000262	-6.16E-06	3.09E-01
	7.72***	17.21***	13.75***	30.11***
GDPpc²	3.43E-10	5.49E-09	9.15E-11	-6.48E-06
	5.29***	10.65***	5.53***	19.85***
Turning Point ($)	23178	23825	33661	23878
R2	0.12	0.39	0.50	0.59
obs.	1359	1359	1359	1359
Nber of countries	131	131	131	131
period	1988-2004	1988-2004	1988-2004	1988-2004
Countries on the right of the turning point in 2004				
	Australia	Australia	Ireland	Australia
	Austria	Austria	Norway	Austria
	Belgium	Belgium	US	Belgium
	Canada	Canada		Canada
	Denmark	Denmark		Denmark
	Finland	Finland		Finland
	France	France		France
	Germany	Germany		Germany
	Hong Kong	Hong Kong		Hong Kong
	Ireland	Ireland		Ireland
	Italy	Italy		Italy
	Japan	Japan		Japan
	Netherlands	Netherlands		Netherlands
	Norway	Norway		Norway
	Singapore	Singapore		Singapore
	Sweden	Sweden		Sweden
	Switzerland	Switzerland		Switzerland
	UK	UK		UK
	US	US		US

Absolute value of robust t statistics under coefficients (White's correction for heteroskedasticity used)
***, **, * significant at respectively 1%, 5% and 10% level.

Table 5
Pooled, within and between estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dependant	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber
Method	Pooled	Pooled	Pooled	Pooled	Within	Within	Within	Within	Between	Between	Between	Between
GDPpc	-1.55E-05	-0.000257	-5.96E-06	3.09E-01	6.13E-06	0.000108	1.38E-06	5.29E-02	-1,31E-05	-0,000228	-6,96E-06	2,71E-01
	7.55***	16.69***	13.37***	28.92***	1.6	5.75***	3.17***	3.53***	2.16**	4.32***	4,51***	6.91***
GDPpc²	3.33E-10	5.35E-09	8.48E-11	-6.47E-06	-4.07E-11	-6.83E-10	1.89E-11	-2.02E-06	2,81E-10	4,97E-09	1,36E-10	-5,91E-06
	5.13***	10.20***	5.15***	19.14***	0.52	1.79*	2.15**	6.68***	1.66*	2.81***	2.60***	4,80***
Turning Point ($)	23273	24019	35142	23868	-	-	-	13085	23310	22948	25588	22926
Year effects	yes	yes	yes	yes	yes	yes	yes	yes	-	-	-	-
Country effects	no	no	no	no	yes	yes	yes	yes	-	-	-	-
R2	0.12	0.40	0.50	0.59	0.02	0.14	0.21	0.28	0.06	0.27	0.45	0.44
obs.	1359	1359	1359	1359	1359	1359	1359	1359	131	131	131	131
Nber of countries	131	131	131	131	131	131	131	131	131	131	131	131
period	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004

Absolute value of robust t statistics under coefficients.
***, **, * significant at respectively 1%, 5% and 10% level.
Note: all sample except microstates, GDP per capita PPP in constant 2000 international $, from WDI

Table 6
Robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependant	Gini	Nber	Nber	HHI	Theil	Gini	Nber
Method	Logistic transformation		Negative binomial	System GMM			
GDPpc	-2.70E-04	3.34E-04	1.51E-04	-2.87E-05	-3.62 E-04	-7.84E-06	3.75E-01
	20.72***	23.66***	22.97***	5.29***	7.55***	5.78***	10.50***
GDPpc²	5.73E-09	-6.54E-09	-3.57E-09	7.35E-10	8.81E-09	1.60E-10	-8.57E-06
	13.67***	14.19***	18.40***	4.49***	5.65***	3.35***	7.31***
Turning Point ($)	23543	25566	21176	19524	20556	24500	21865
Year effects	yes	yes	yes	yes	yes	yes	yes
obs.	1359	1359	1359	1359	1359	1359	1359
Nber of countries	131	131	131	131	131	131	131
period	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004	1988-2004

Absolute value of robust t statistics under coefficients.
***, **, * significant at respectively 1%, 5% and 10% level.
Note: all sample except microstates, GDP per capita PPP in constant 2000 international $, from WDI

Table 7
Cumulated "Closed" lines over 2001-2003 on average for countries with a GDP per capita > 24 000$

Cumulated Closed lines 2001-2003	Mean	Std. Dev.	Min	Max
Cumulated number of closed lines	140	71.8	73	295
Cumulated number of closed lines in % of total actives lines in 2000	3.16%	1.84%	1.49%	7.81%
Cumulated value of closed lines in % of total exports in 2000	0.60%	0.78%	0.06%	3.29%

"Closed" lines in date t are defined as a line with positive exports in t-2 and t-1 et zero exports in t, t+1 and t+2 for countries with a population higher than 1 million (not a microstate) and a GDP per capita higher than 24 000$ (hence for countries at the right of the turning point). To have a robust picture, we compute the "cumulated" closed lines over 2001-2003.

Table 8
Estimates with raw-material export shares

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber
GDPpc	-1.52E-05	-2.54E-04	-5.92E-06	0.308042	-2.00E-05	-3.02E-04	-6.54E-06	0.3332
	10.32***	22.69***	15.12***	29.85***	11.85***	22.46***	14.16***	28.59***
GDPpc²	3.84E-10	5.71E-09	8.99E-11	-6.58E-06	5.54E-10	7.06E-09	9.4E-11	-7.42E-06
	8.67***	16.34***	6.35***	20.41***	10.99***	16.18***	5.64***	19.68***
Raw materials	0.5039	3.5376	0.0507	-1118.25	0.3890	1.8221	0.0024	-416.99
	22.36***	34.85***	18.11***	9.11***	3.23***	9.35***	9.61***	2.20**
GDPpc*Raw mat.					3.06E-05	3.63E-04	7.54E-06	-0.171182
					4.19***	10.78***	9.61***	3.57***
GDPpc²*Raw mat.					-1.06E-09	-9.39E-09	-7.85E-11	5.41E-06
					6.86***	9.64***	2.67***	4.04***
Turning Point ($)	19792	22277	32925	23407	-	-	-	-
Year effects	yes				yes			
obs.	1359				1359			
Nber of countries	131				131			
period	1988-2004				1988-2004			

Absolute value of robust t statistics under coefficients.
***, **, * significant at respectively 1%, 5% and 10% level.
Note: all sample except microstates, GDP per capita PPP in constant 2000 international $, from WDI

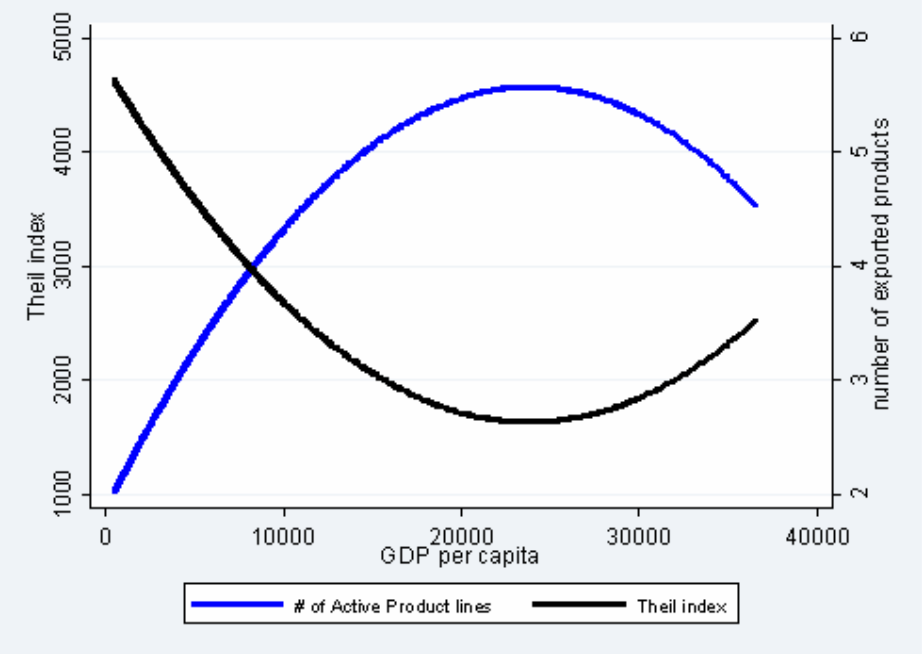
Table 9
Effect of public infrastructure capital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependant	HHI	Theil	Gini	Nber	HHI	Theil	Gini	Nber
GDPpc	-1.48E-05	-2.27E-04	-5.75E-06	0.2543	-1.02E-05	-1.98E-04	-5.25E-06	0.2578
	5.96***	11.67***	9.53***	17.87***	4.02**	8.74***	7.81***	15.59***
$GDPpc^2$	3.15E-10	4.64E-09	8.77E-11	-5.06E-06	2.09E-10	3.97E-09	7.61E-11	-5.14E-06
	4.23***	7.29***	4.16***	11.89***	2.79***	5.65***	3.37***	10.91***
Public capital					-0.0296	-0.1865	-0.0032	22.2698
					2.71***	2.71***	2.48**	0.47
Turning Point ($)	23492	24494	32782	25129	24402	24987	34494	25074
Year effects	yes				yes			
obs.	727				727			
Nber of countries	92				92			
period	1996-2004				1996-2004			

Absolute value of robust t statistics under coefficients.
***, **, * significant at respectively 1%, 5% and 10% level.
Note: all sample except microstates, GDP per capita PPP in constant 2000 international $, from WDI

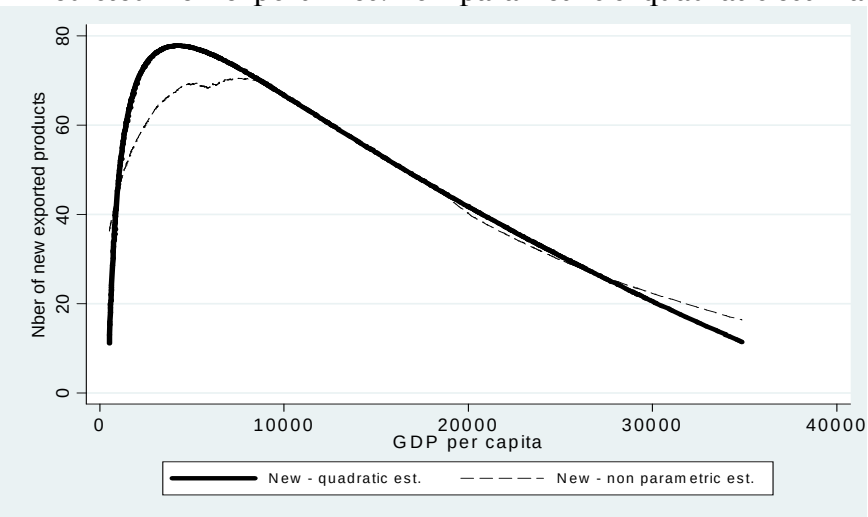
Figures

Figure 1
Predicted Theil's concentration index & number of active export lines



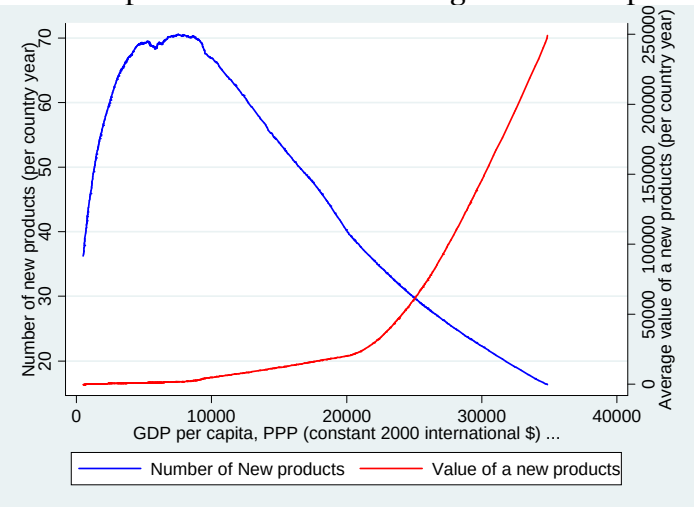
Source: author calculations using COMTRADE

Figure 2
Predicted New export lines: non-parametric & quadratic estimates



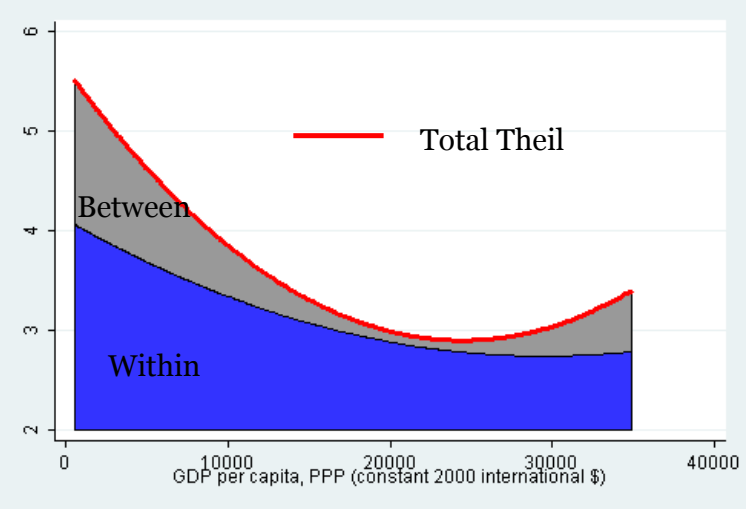
Source: author calculations using COMTRADE

Figure 3
New products and their average value: nonparametric curves



Source: author calculations using COMTRADE

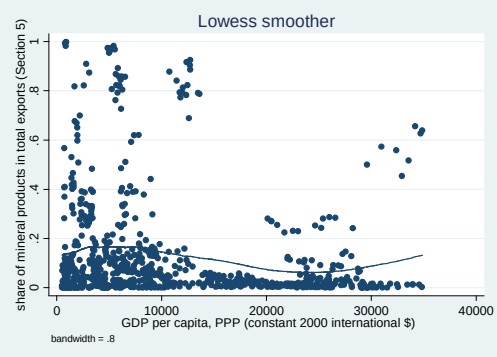
Figure 4
Within and between components of Theil's index



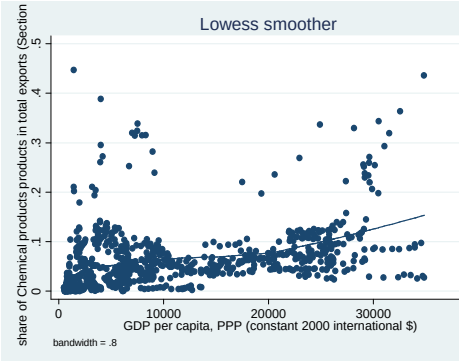
Source: author calculations using COMTRADE (quadratic estimates)

Figure 5
Selected sectoral shares against GDP per capita

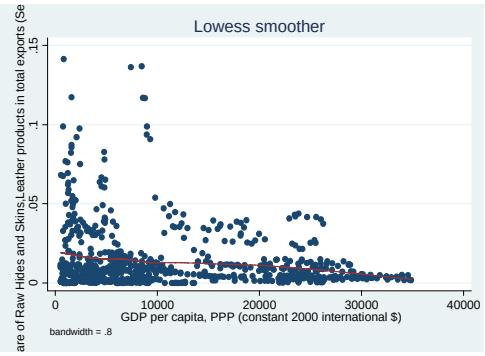
(a) Minerals (section 5)



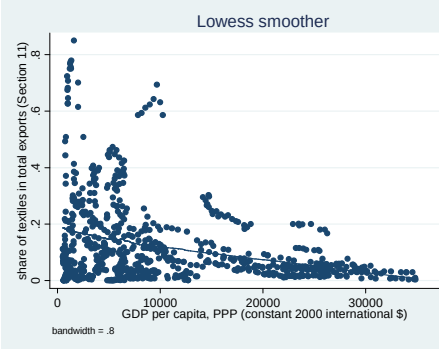
(b) Chemicals (section 6)



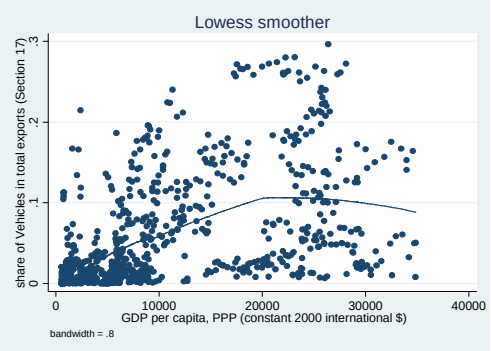
(c) Raw Hides and Skins, Leather (section 8)



(c) Textile & Apparel (section 11)



(e) Transport equipments (section 17)



(d) Machinery (section 16)

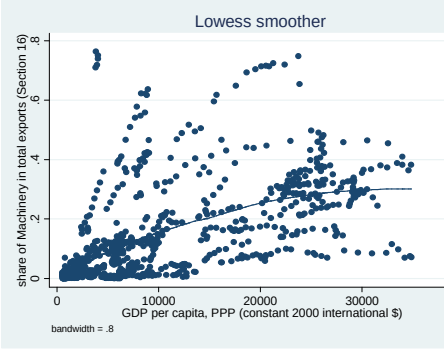


Figure 6
Theil indices against GDP and the share of raw materials

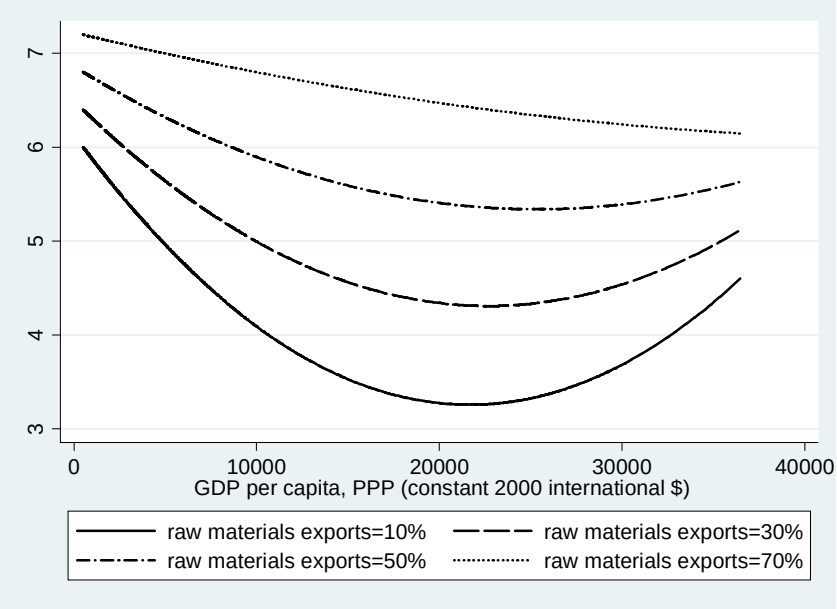


Figure 7 Shares Value/number of lines by sections
Weighted average

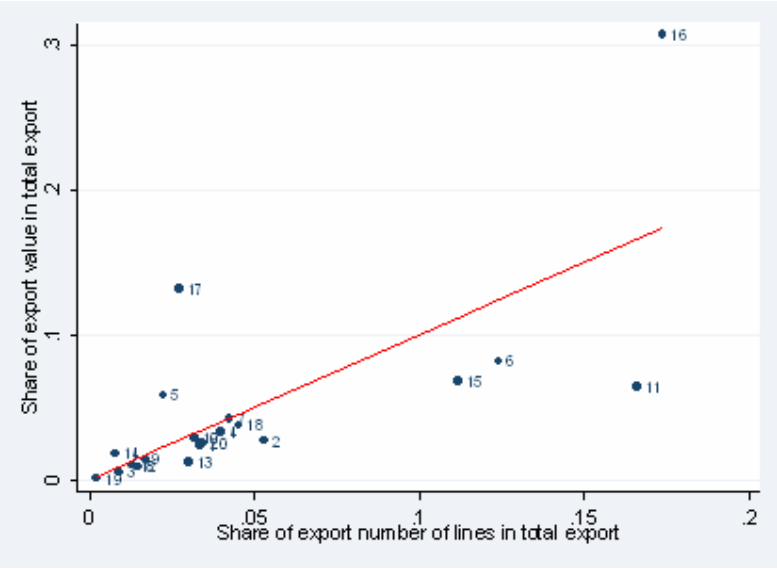


Table A.1
Countries in the sample (available time period in brackets)

High income		Low income		Lower middle income		Upper middle income	
Australia	[1988 - 2004]	Bangladesh	[1989 - 2004]	Albania	[2004 - 1996]	Ant. and Bar	[1999 - 2000]
Austria	[1994 - 2004]	Benin	[1998 - 2002]	Algeria	[2004 - 1992]	Argentina	[1993 - 2004]
Bahamas, The	[1997 - 2001]	Bhutan	[1993 - 1999]	Armenia	[2004 - 1997]	Belize	[1992 - 2003]
Bahrain	[2000 - 2003]	Burkina Faso	[1995 - 2004]	Azerbaijan	[2004 - 1996]	Botswana	[2000 - 2001]
Belgium	[1999 - 2004]	Burundi	[1993 - 2004]	Belarus	[2004 - 1998]	Chile	[1990 - 2004]
Canada	[1989 - 2004]	Cambodia	[2000 - 2004]	Bolivia	[2004 - 1992]	Costa Rica	[1994 - 2004]
Cyprus	[1989 - 2004]	CAR	[1993 - 2003]	Brazil	[2004 - 1989]	Croatia	[1992 - 2004]
Denmark	[1989 - 2004]	Comoros	[1995 - 2000]	Bulgaria	[2004 - 1996]	Czech Repub	[1993 - 2004]
Finland	[1988 - 2004]	Congo, Rep.	[1993 - 1995]	Cameroon	[2004 - 1995]	Dominica	[1993 - 2004]
France	[1994 - 2004]	Côte d'Ivoire	[1995 - 2003]	Cape Verde	[2004 - 1997]	Estonia	[1995 - 2004]
Fr. Polynesia	[1996 - 2004]	Eritrea	[- 2003]	China	[2004 - 1992]	Gabon	[1993 - 2004]
Germany	[1988 - 2004]	Ethiopia	[1995 - 2003]	Colombia	[2004 - 1991]	Grenada	[1993 - 2003]
Greece	[1988 - 2004]	Gambia, The	[1995 - 2003]	Dom. Rep.	[2001 - 1997]	Hungary	[1992 - 2004]
Hong Kong, Chin	[1993 - 2004]	Ghana	[1996 - 2004]	Ecuador	[2004 - 1991]	Latvia	[1994 - 2004]
Iceland	[1988 - 2004]	Guinea	[1995 - 2002]	Egypt	[2004 - 1994]	Lebanon	[1997 - 2003]
Ireland	[1992 - 2004]	Haiti	[1988 - 1997]	El Salvador	[2004 - 1994]	Lithuania	[1994 - 2004]
Israel	[1995 - 2004]	India	[1988 - 2004]	Fiji	[2004 - 2000]	Malaysia	[1989 - 2004]
Italy	[1994 - 2004]	Kenya	[1992 - 2004]	Georgia	[2004 - 1996]	Mauritius	[1993 - 2004]
Japan	[1988 - 2004]	Kyrgyz Rep.	[1995 - 2004]	Guatemala	[2004 - 1993]	Mexico	[1990 - 2004]
Korea, Rep.	[1988 - 2004]	Lesotho	[2000 - 2002]	Guyana	[2004 - 1997]	Oman	[1989 - 2004]
Kuwait	[2000 - 2001]	Madagascar	[1990 - 2004]	Honduras	[2003 - 1994]	Panama	[1995 - 2004]
Luxembourg	[1999 - 2004]	Malawi	[1990 - 2004]	Indonesia	[2004 - 1989]	Poland	[1994 - 2004]
Macao, China	[1991 - 2004]	Mali	[1996 - 2001]	Iran	[2003 - 1997]	Romania	[1989 - 2004]
Malta	[1994 - 2004]	Moldavia	[1994 - 2004]	Jamaica	[2002 - 1991]	Russia	[1996 - 2004]
Netherlands	[1992 - 2004]	Mongolia	[1996 - 2003]	Jordan	[2004 - 1994]	Seychelles	[1994 - 2004]

Table A1 (cont'd)

New Caledonia	[1999 -	2004]	Mozambique	[1994 -	2002]	Kazakhstan	[2004 -	1995]	Slovak Repul	[1994 -	2004]
NZ	[1989 -	2004]	Nepal	[1994 -	2003]	Kiribati	[1999 -	1995]	South Africa	[1992 -	2004]
Norway	[1993 -	2004]	Nicaragua	[1993 -	2004]	Macedonia	[2004 -	1994]	St. Kitts and	[1993 -	2003]
Portugal	[1988 -	2004]	Niger	[1995 -	2003]	Maldives	[2004 -	1995]	St. Lucia	[1992 -	2004]
Saudi Arabia	[1991 -	2002]	Nigeria	[1996 -	2003]	Morocco	[2004 -	1993]	St. V. and th	[1993 -	2004]
Singapore	[1989 -	2004]	Pakistan	[2003 -	2004]	Namibia	[2003 -	2000]	Trinidad and	[1991 -	2003]
Slovenia	[1994 -	2004]	Papua NG	[1998 -	2003]	Paraguay	[2004 -	1989]	Turkey	[1989 -	2004]
Spain	[1989 -	2004]	Rwanda	[1996 -	2003]	Peru	[2004 -	1992]	Uruguay	[1994 -	2004]
Sweden	[1992 -	2004]	São Tomé & Pr.	[1999 -	2003]	Philippines	[2004 -	1996]	Venezuela, R	[1994 -	2004]
Switzerland	[1988 -	2004]	Senegal	[1996 -	2004]	Samoa	[2004 -	2001]			
UAE	[1991 -	2001]	Sierra Leone	[-	2002]	Sri Lanka	[2004 -	1990]			
UK	[1993 -	2004]	Sudan	[1995 -	2003]	Suriname	[2001 -	1994]			
US	[1991 -	2004]	Tanzania	[1997 -	2004]	Swaziland	[2002 -	2000]			
			Togo	[1994 -	2004]	Syria	[2004 -	2001]			
			Uganda	[1994 -	2004]	Thailand	[2004 -	1989]			
			Yemen, Rep.	[-	2004]	Tunisia	[2004 -	1991]			
			Zambia	[1992 -	2004]	Turkmenistan	[2000 -	1997]			
			Zimbabwe	[1995 -	2004]	Ukraine	[2004 -	1996]			
						Vanuatu	[2000 -	1993]			

Table A.2
Cumulated "closed" lines, 2001-2003, by main chapters,
countries with GDP per capita over PPP$24'000

Chap.		Corresp. Section		Nber of lines	In % of chapter's active lines in 2000	In % of total closed lines	Value, in % of total export value in 2000
29	Organic Chemicals	6	Products of the Chemical or Allied Industries	11.9	5.40%	8.20%	2.79E-02
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;Articles of Animal Gut	8.8	29.50%	7.90%	2.69E-02
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	8.4	5.90%	6.50%	1.81E-02
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	7.2	6.80%	6.20%	2.22E-01
11	Products of the Milling Industry; Malt; Starches; Inulin; Wheat Gluten	2	Vegetable Products	4.6	16.60%	4.00%	1.28E-03
62	Articles of Apparel and Clothing Accessories, Not Knitted Or Crocheted	11	Textiles and Textile Articles	5.6	1.20%	3.70%	8.99E-03

Table A.2 (cont'd)

Chap.		Corresp. Section		Closed lines: Nber of lines	Closed lines: In % of chapter's active lines in 2000	Closed lines: In % of total closed lines	Closed lines: Value, in % of total export value in 2000
25	Salt, Sulphur, Earths and Stone; Plastering Materials, Lime and Cement	5	Mineral Products	4.6	7.50%	3.20%	4.73E-04
68	Articles of Stone, Plaster, Cement, Asbestos, Mica or Similar Materials	12	Footwear, Headgear, Umbrellas, Sun Umbrellas, Walking-Sticks, Seat-Sticks, Whips, Riding-Crops and Parts Thereof; Prepared Feathers and Articles Made Therewith; Artificial Flowers; Articles of Human Hair	3.3	6.70%	2.80%	2.57E-04
52	Cotton	11	Textiles and Textile Articles	4.1	4.50%	2.60%	7.12E-05
43	Furskins and Artificial Fur; Manufactures Thereof	8	Raw Hides and Skins,Leather, Furskins and Articles Thereof; Saddlery and Harness; Travel Goods, Handbags, and Similar Containers;Articles of Animal Gut	2.7	17.50%	2.60%	5.95E-04
12	Oil Seeds and Oleaginous Fruits; Misc, Grains, Seeds & Fruit; Industrial or Medicinal Plants; Straw and Fodder	2	Vegetable Products	3.2	10.10%	2.40%	4.48E-04
15	Animal or Vegetable Fats and Oils and their Cleavage Products; Prepared Edible Fats; Animal or Vegetable Waxes	3	Animal or Vegetable Fats and Oils and Their Cleavage Products; Prepared Edible Fats;	3	6.80%	2.30%	1.85E-03
3	Fish & Crustaceans, Molluscs & Other Aquatic Invertebrates	1	Live Animals; Animal Products	3.3	6.50%	2.10%	2.17E-04
53	Other Vegetable Textile Fibres; Paper Yarn and Woven Fabrics of Paper Yarn	11	Textiles and Textile Articles	2.8	13.80%	2.10%	2.65E-04
26	Ores, Slag and Ash	5	Mineral Products	3	15.40%	2.10%	6.49E-04
72	Iron and Steel	15	Base Metals and Articles of Base Metal	3.7	2.80%	2.00%	3.14E-02

Note: "Closed" lines at date *t* are defined as lines with positive exports at *t-2* and *t-1* and zero exports at *t*, *t+1* and *t+2*. The sample is restricted here to countries with populations above one million (no microstates) and GDP per capita above PPP$24'000 (at the right of the turning point). Data is cumulated over 2001-2003 for robustness.