Stages of Diversification*

Jean Imbs
London Business School
and CEPR

Romain Wacziarg
Stanford University

September 2000

Abstract

This paper studies the evolution of sectoral labor concentration in relation to the level of per capita income. We show that various measures of sectoral concentration follow a U-shaped pattern across a wide variety of data sources: countries first diversify, in the sense that labor is spread more equally across sectors, but there exists, relatively late in the development process, a point at which they start to specialize again. We introduce a model with endogenous costs of trading internationally that provides an explanation for this new empirical fact. The model highlights a trade-off between the benefits of diversification in the context of high trading costs, and the benefits of specialization in a Ricardian sense.

Keywords: Specialization, International Trade, Comparative Advantage.

JEL Classification: F43, F15, O40

*We thank Alberto Alesina, Lanier Benkard, John McMillan, Gian Maria Milesi-Ferretti, Morten Ravn, Paul Romer, Andrei Shleifer, seminar participants at the Universite de Lausanne, ECARES, Trinity College Dublin, ICRIER, CERGE-EI, the University of Chicago, the University of Manchester, the NBER Summer Institute 2000, the University of Maryland, Boston College, the IMF and Stanford University for useful comments. Zeynep Emre and Luca Bindelli provided diligent research assistance. The first author gratefully acknowldges financial assistance from the Fonds National de la Recherche Scientifique (Switzerland), Grant # 12-53584.98. The usual disclaimer applies. Contact information: Imbs: London Business School, Dept. of Economics, Regent's Park, London NW1 4SA, jimbs@london.edu. Wacziarg (corresponding author): Stanford Graduate School of Business, 518 Memorial Way, Stanford CA 94305-5015, wacziarg@gsb.stanford.edu

1 Introduction

This paper characterizes the pattern of sectoral labor diversification along the development path. Using data from a variety of sources covering a wide cross-section of countries, we provide new and robust evidence that economies grow through two stages of diversification. At first, labor is allocated increasingly equally between sectors, but there exists a level of per capita income beyond which the sectoral distribution of labor inputs starts concentrating again. In other words, the sectoral concentration of labor follows a U-shaped pattern in relation to per capita income.

At early stages of development, countries are usually specialized in exploiting their endowments of natural resources, and several theories suggest that their priority should be to diversify out of extreme specialization. For example, simple arguments based on the law of large numbers would stress that diversification helps dampening the aggregate effects of sector-specific shocks.[1] Similarly, in the context of high trading costs a country where consumers display a taste for diversity might find it efficient to open new sectors domestically rather than import desired goods at a high cost.[2] As productivity rises, the range of goods for which it is efficient to produce domestically increases and the economy diversifies.

On the other hand, there is now a substantial body of literature emphasizing important reasons for sectoral specialization or agglomeration to occur. When external to the firm, increasing returns to scale in aggregate production can make it efficient for factors of production to concentrate in one or a few sectors and generate faster long-run growth. In the economic geography literature, the presence of pecuniary externalities ("demand linkages") and costly trade make it optimal for monopolistic competitors to cluster since their profits increase in local expenditures, which are themselves a positive function of the number of local firms. The idea that transport costs fall with better technologies tends to reinforce the mechanism, as the de-

[1] This incentive has recently been related to the fact that aggregate volatility tends to fall with income per capita, in a model where opportunities to diversify are limited by the presence of indivisibilities in financing imperfectly correlated projects (see Acemoglu and Zilibotti (1997)). Successful investment in a risky project reduces aggregate volatility and increases the pool of available capital. Murphy, Shleifer and Vishny (1989) resort to a more specific pecuniary externality whereby profits in each sector depends on the number of existing sectors. There, diversification (industrialization) can only occur as a coordinated Big Push.

[2] This remains true even with access to perfect capital markets: the goods must be consumed domestically, and thus the cost of importing them acts to reduce trade.

mand linkages effect is decreasingly mitigated by the incentive to be located close to demand. Thus, activity in integrated economies tends to be agglomerated.[3] Such geographic clustering or agglomeration would naturally translate into high observed degrees of sectoral concentration. Similarly, Ricardian trade theory predicts that open economies ought to specialize in producing a specific range of goods - so that open economies ought to be relatively more specialized.[4]

Thus, theory is ambiguous as to how measures of sectoral diversification should evolve as countries grow, although the force of diversification is probably more at play amongst low-income countries, and the force of concentration amongst richer economies. The findings in this paper are consistent with this view. Indeed, both sets of theories seem to be at play, at different points in the development process: sectoral labor diversification goes through two stages, at first one of increasing diversification, and later one of increasing concentration.

Our result is an extremely robust feature of the data. The non-linearity holds above and beyond the well-known shift of factors of production from agriculture to manufacturing and on to services. It holds within countries as well as across countries, for a variety of levels of disaggregation and data sources. The turnaround point is estimated to occur quite late in the development process. Thus, increased sectoral specialization in rich countries, although a significant development, is no more than the tip of the iceberg: countries diversify over most of their development path.

We propose a model where our stylized fact results from the interaction of exogenous productivity increases and endogenous trading costs. Our (small) open economy benefits from "catch-up" growth, so that (relative) productivity rises at an exogenous but declining rate. Concurrently, the economy uses a proportion of its resources to invest in infrastructures that make trade less costly. In early stages, the former effect dominates, so that it is optimal to open new sectors rather than to import the goods at a high

[3]See for instance the seminal book by Krugman (1991), where employment in the US is shown to be more specialized than in Europe. Predictions of this literature do not stop here: when decreasing returns to specialization are introduced, in the form of endogenous congestion costs as in Krugman and Venables (1995), where labor is immobile and wages increase with specialization, there comes a point at which de-agglomeration becomes optimal. This has been documented by Kim (1993) for US regions. We find some weak evidence of this in our data as well, although there are too few observations at sufficiently high levels of income in our sample for the de-agglomeration pattern to show up significantly.

[4]See for instance Dornbusch, Fischer and Samuelson (1977).

cost. This translates into observed diversification, as standard measures of concentration fall for a given level of trading costs. However, as the country grows the second effect becomes increasingly important, until it actually dominates and the number of sectors falls again by virtue of a standard Ricardian argument. As a result, observed measures of concentration rise.

The paper is structured as follows: Section 2 contains empirical evidence on the evolution of sectoral concentration for a broad panel of countries, using a variety of data sources and levels of disaggregation. Section 3 presents a general equilibrium model consistent with the new empirical regularities documented in the first section. Section 4 presents a simple empirical test of the model, providing support to our modelling strategy. Finally, the last section concludes.

2 The Evolution of Sectoral Concentration

This section documents the main empirical result of the paper: the relationship between various measures of sectoral employment concentration and the level of per capita income displays a U-shaped pattern. It also identifies the level of income at which minimum concentration level is attained, and performs a series of robustness checks for the stylized fact.

2.1 Overview of the Data

We employ sectoral employment data from the ILO (1997) and UNIDO (1997) to examine the evolution of several measures of sectoral concentration through time and in relation to the level of development.[5] The data pertain to employment shares across sectors, at the 1-digit (ILO) and 3-digit levels (UNIDO). While the ILO data cover all economic activities, the UNIDO 3-digit data covers only manufacturing. Both data sets span a wide range of industrial and developing countries.[6] In addition to ILO and UNIDO data, we also use data from the OECD (1998) for a set of 14 industrial countries. For the OECD, the data exist at the 2-digit level of disaggregation and

[5] Appendix 1 contains details concerning data coverage.

[6] The dataset was constructed so that the number of sectors available *through time* for each country was constant. This required abandoning observations on some sectors, when observations for a given country were not available for all years. On the other hand, the number of sectors available to compute the indices of concentration varied *across countries*. We retained countries for which all 27 or more sectors were available (UNIDO), or where all 6 or more sectors were available (ILO). Such variation is reflected purely in the country-specific effects under fixed-effects regression.

cover all of national activity, not just manufacturing (although some non-manufacturing sectors appear at the 1-digit level of disaggregation only). The OECD data therefore provides a check on robustness with respect to the level of disaggregation and allows a greater focus on the upper end of the per capita income distribution.[7]

The choice of employment shares as a measure of sector size is motivated by the absence of output volume data for the country coverage we seek to obtain, and is usual in the empirical literature concerned with sectoral specialization.[8] Our measure of sector size therefore refers to the (labor) resources employed in the production of goods rather the share of the value of these goods in overall income. The ILO data cover the 1969-1997 time period, the UNIDO 3-digit data extend from 1963 to 1996, and the OECD data covers the period 1960-1997.[9] All four data sets contain annual observations, so that both within and between country variation can be employed.

Tables 1 and 2 contain summary statistics for the sectoral concentration data. We employ two different measures for the concentration of employment across sectors: a Gini coefficient for the inequality of sector shares (Gini) and a Herfindahl index for the concentration of employment across sectors (Herfindahl).[10] Our measures of sectoral concentration are highly correlated amongst themselves, as shown in Table 2. However, the correlations are not sufficiently high to warrant using only one of these variables, so we present results using both measures throughout.

2.2 Within-Country Results

2.2.1 Non-parametric estimates

In order to investigate the shape of the relationship between our indices and income levels, we attempt to impose as little structure on the functional form as possible. This motivates the use of a non-parametric procedure to

[7] We also use data from the UNIDO at the 4-digit level of disaggregation (Section 2.4). These data cover the 1977-1997 period. While they are of relatively poor quality and their coverage is limited, they will provide additional robustness checks on our results.

[8] See for instance Krugman (1991).

[9] Since our per capita income data from Summers and Heston only extends to 1992, in most cases the last years of available data for both the ILO and the UNIDO datasets had to be discarded, a fact that explains in part why the upper tail of our data is relatively thin.

[10] We also used the share of the largest sector in employment, and the results were very similar to those obtained using the Herfindahl index. These results are available upon request.

identify the shape of the relationship. We focus on within-country estimates, i.e. a curve obtained by employing the within-country variation in the data only. We also focus at first on the UNIDO and ILO data sets, as they cover the widest set of countries, including developing economies.

Specifically, we partition the data into N sub-samples according to (overlapping) income intervals of size S, with an overlap of size $S-\Delta$.[11] For each sub-sample, we run a simple fixed effects linear regression of the measures of concentration on income. For each regression we can compute the fitted value of the regression, evaluated at the midpoint of the income interval used to determine the sample, and plot these fitted values against the income midpoint of each estimation interval. In other words, these plots are obtained by a series of linear approximations. We report the results of this procedure for the Gini coefficient in Figures 1 and 2.[12] Both the UNIDO and ILO data show that the relationship between specialization and income per capita is highly non-monotonic: it appears to be U-shaped, and the U-shape seems particularly pronounced for the Gini coefficient. This is the central stylized fact evidenced in this paper, and the rest of this section will be spent examining its robustness.

Two observations are in order. Firstly, the U-shaped patterns are not symmetric: given currently observed maximal levels of income, the upward bending portion of the curve does not swing up back to the level at which the downward bending portion of the curve starts. Secondly, the shift towards reconcentration seems to occur late in the development process: the non-parametric estimates can be used to compute the level of income at which the point of minimum concentration occurs. Specifically, we can take the value of income at which the slope of the curve is zero as our estimated minimum point. Table 3 reports these minima.

For example, for the ILO data, using the Gini measure of sectoral labor concentration, we find that the minimum point occurs when per capita income equals approximately \$9,575 per year. The Summers-Heston data employed for the calculations is in constant 1985 US dollars, so this point

[11] For the ILO dataset, $N = 529$, for the UNIDO dataset $N = 489$ and for the OECD dataset $N = 409$. We use an income interval $S = \$5,000$. Each interval, given the range of income in the data, has a midpoint which is $\Delta = \$25$ away from the following interval. The results are not sensitive to the choice of the income interval nor the number of subsamples. This estimator is essentially equivalent to a Kernel estimator with a "flat" or rectangular Kernel.

[12] The corresponding figures for the Herfindahl index look similar and are available upon request.

occurs roughly at the level of income reached by Ireland in 1992.[13] In other words, according to these estimates the minimum point occurs quite late in the development process. It occurs somewhat earlier for the UNIDO data set, suggesting that reconcentration occurs earlier within manufacturing than across a broader and wider set of sectors. However, the curve is much flatter for the UNIDO data set, making the identification of the point of minimum concentration less precise.

2.2.2 Significance of the upward-bending portion of the U-curve

Since the U-curve is asymmetric and the minimum point occurs late in the development process, the question of the statistical significance of its upward bending portion arises. A systematic way to ascertain the presence of an upward bending portion of the curve is to examine the significance of the coefficient for each of the sub-samples used in the non-parametric estimation of the relationship. This limits the incidence of outliers. Figures 4 to 6 display the range of coefficients for which the slope coefficient estimates are significantly different from zero at the 95% confidence level. The figures show, for the various data sets and measures of concentration, that there exists a point strictly greater than the minimum point, at which the slope estimates become generally statistically significant and positive (Table 3 displays the interval midpoint income levels above which the estimated coefficients on income become statistically significant at the 5% level).

We also use sectoral employment data from the OECD to assess the significance of the upward bending portion of the curve. Indeed, focusing on data pertaining to a set of 14 industrial countries at a higher level of disaggregation than the ILO data and applying our non-parametric procedure demonstrates that the upward bending portion of the curve is much more pronounced for the OECD sample (Figure 3) than for the samples which included developing countries. By focusing on the higher end of the income distribution and extending the whole-economy data coverage to the 2-digit level of disaggregation, these OECD results provide further evidence that the upward bending portion of the U-curve is indeed present in the data. It is remarkable that the estimated OECD minimum points (Table 3) are similar to those obtained using ILO and UNIDO data.

[13] $9,637 in Summers-Heston dollars.

2.2.3 Quadratic Specification

The above results suggest that the relationship between measures of sectoral labor concentration and income may be well represented by a quadratic specification.[14] In order to examine the robustness of our result, we turn to within-country results based on a quadratic specification for sectoral concentration. An added advantage of such a specification is that, contrary to the non-parametric methodology described above, it allows out-of-sample predictions on the stages of diversification, and that it provides precise parameter estimates for the shape of the curve, above and beyond the shape uncovered in Figures 1-3.

Table 4 displays the coefficients from fixed-effect regressions of sectoral concentration on per capita income and the square of per capita income. As before, within-country results demonstrate the existence of a statistically significant U-shaped relationship between sectoral concentration of labor and the level of per capita income, irrespective of the measure used. Figures 7-9 confirm graphically the existence of the non-linearity, by displaying the estimated within-country relationship (solid line) against the backdrop of the pooled-data scatter plot.[15] The economic significance of this relationship appears clearly for the ILO and OECD data sets, that is, for a broader level of disaggregation covering a wider range of sectors. The relationship is flatter for the UNIDO data.[16] However, the estimated fixed-effects coefficients on income (negative) and on income squared (positive) for all data sets are highly significant statistically (Table 4).

The minimum point, after which countries begin to reconcentrate, can be calculated easily for each set of estimates, by setting the derivative of sectoral concentration with respect to income to zero. This leads to estimates that are comparable to the minimum point estimates obtained using our non-

[14] Above, we employed non-parametric estimation because fitting a quadratic relationship is not only forcing a very special case on the data: it may also lead to erroneous conclusions. For instance, a log specification on income levels may fit the data equally well, however it would not imply an upward bending portion for the curve, and hence we could not infer reconcentration at higher levels of income.

[15] In other words, the scatter plot does not display deviations from the means. This is to allow comparability with pooled least squares plots, which reveals the source of the variation - cross-sectional (between) or time series (within).

[16] Throughout the paper, parametric results implied by the UNIDO dataset are somewhat weaker than for other data, without however creating contradictions. We think this is mostly related to the relative poorer quality of those data. Furthermore, the U-shaped relationship for the UNIDO dataset appears clearly in our non-parametric within-country estimates.

parametric procedure.

In this parametric specification the fact that few data points lie above the minimum point might imply that the non-linearity results from a few outliers.[17] To further assess the significance of the upward portion of the curve, we can restrict the sample to observations with income greater (or lower) than the estimated point of minimum concentration and examine the evolution of sectoral concentration for each subsample. We can then run a simple fixed-effects regression of concentration measures on the level of income (without the quadratic term) for high income levels, as shown in Table 5. For all specifications, the coefficient on the income variable remains highly significant statistically, providing suggestive evidence that the reconcentration phase is indeed present in a statistically meaningful way. Turning to regressions for the subset of the data with income lying below the minimum level (Table 6), the coefficient is negative and statistically significant, as expected. Moreover, the absolute value of the coefficient on income is always greater, and sometimes as much as twice larger than the corresponding coefficients obtained in Table 5, providing further support for the fact that the U curve is asymmetric.

2.3 Between-country Estimates

The evidence presented above pertains to within-country evidence, which is perhaps the most relevant evidence as it describes what happens to measures of sectoral concentration in a typical country along the growth process. However, our result holds in a standard cross-country approach as well. To show this, we turn to between-country evidence.[18] Table 8 displays results based on a between estimator (OLS on country means) applied to the ILO and UNIDO data sets.[19] The U-shaped pattern is more pronounced than with within-country estimates in magnitude, and it remains highly significant statistically despite the much reduced number of data points. Figures 10 and 11 provide a graphical representations of the U-curves obtained from

[17] Using fixed effects, we can note that approximately 33.6% of the country-year observations lie above the inflexion point for ILO data, while this proportion is 24% for the UNIDO data.

[18] The set of countries is reduced somewhat because the number of sectors used for each country has to be equal (countries with missing sector observations had to be deleted). We no longer have fixed-effects to account for the incidence of cross-country, time invariant differences in the number of available sectors. We retained countries for which all 28 sectors were available (UNIDO), or where all 9 sectors were available (ILO).

[19] We did not compute results for the OECD data set as those would be based on 14 observations only.

between estimates, against the scatter plot of country-means. The cross-sectional U-shape appears clearly in these figures. Again, Table 7 reports the implied minimum points, which remain relatively high. For instance, using the Gini coefficient as a measure of concentration in the ILO data set, the minimum point is estimated to occur when income per capita reaches $9,645 per year, again roughly the level of Ireland in 1992 ($9,637).

Within and between-country estimates are very similar. However, the magnitude of the coefficients on income and income squared (that is, the curvature of the U) is generally larger using the between estimator than using the within estimator. This is especially true for the UNIDO data set. This probably stems from measurement error in these data, leading to more errors-in-variables bias under fixed-effects relative to OLS on means.[20].

These results are confirmed when both the within and between variations are pooled. Table 8 presents random effects estimates which optimally weigh the within and between variations in the data.[21] As expected the significance of the estimates is raised considerably, and their magnitude is larger than under fixed-effects.

2.4 Robustness

In this subsection we discuss a number of robustness checks that we conducted on the evidence presented above.[22] In most cases, these robustness checks failed to affect the main stylized fact described above.

4-Digit UNIDO data First, we examined the possibility that our pattern would hold at finer levels of disaggregation within the manufacturing sector. The UNIDO published annual data for manufacturing sectors at the 4-digit level of disaggregation (corresponding to a maximum of 81 sectors per country-year), from 1977 to 1997. One important caveat is that these data are quite spotty and likely to be of lesser quality than 1-digit and 3-digit employment data. Replicating the results for the quadratic specification presented above at the 4-digit level, however, led to interesting results: the

[20] Indeed, under classical (white-noise) measurement error, the bias is exacerbated since we are differencing a highly autocorrelated right hand side variable (income and income squared). Hence, the error-to-truth ratio is raised. In any event, the U-shaped pattern remained statistically significant even when it was estimated to be relatively flat under fixed effects.

[21] Again, we do not present OECD results using any between country variation, as too few countries share the same coverage in terms of the number of sectors.

[22] The estimates corresponding to these robustness checks are available upon request.

U-curve pattern was preserved for all measures of sectoral labor concentration. Our estimates on income and income squared were close to statistically significant for the Herfindahl index of concentration, and consistently significant when using the between-country variation in the data. They were not, however, statistically significant (although they were of the expected signs) when using a within estimator applied to the Gini coefficient of sectoral concentration. We ascribe the lower precision of our estimates at the 4-digit level to the lower quality of the data. In fact, given the poor quality of the 4-digit data it is remarkable that the stylized fact documented in this paper would show up at all.

Boundedness of the concentration measures Each measure of concentration used in this study is bounded above by 1 and below by 0. This may drive part of the non-linearity in the relationship between sectoral concentration and income levels. As the data stand, however, there are very few data points that lie near the bounds, where the boundedness of the measures could generate artificial non-linearities. However, in order to examine this issue, we considered logistic transforms of each of the concentration measures. These provide a way of transforming the variables so they are not bounded above or below anymore. We then used these transformed measures as dependent variables in fixed effects regressions on income and income squared. The results provided further evidence concerning the robustness of the U-curve, as the coefficients were of the expected signs and their statistical significance levels remained high.

Controlling for country size Large countries should be expected to be both less open to trade and more diversified. Hence, country size might be an important variable to include in our regressions to check for the robustness of our U-shaped pattern. Indeed, we may observe less of a tendency for large countries to re-specialize once they reach their minimum point - since they are less open to trade. To address these issues, we controlled for the log of population in the basic regression. As expected, larger countries were found to be more diversified than smaller ones, but the inclusion of country size did not modify the U-shaped pattern.

Removing the agricultural sector Finally, insofar as the results pertaining to the ILO dataset are driven by the secular structural shift away from agriculture, into manufacturing and services, and later from manufacturing to services, we might expect that the U-shaped pattern uncovered in the ILO sample would vanish when agriculture is removed from the sectoral coverage of this data set. Indeed, in this case the curve would appear relatively flat at first, and increasing at higher levels of income, rather than U-shaped. However, this was not the case: the quadratic functional form appears supported by the data for the ILO dataset even when the agricultural sector (as well as mining and quarrying) is excluded from the data. Furthermore, we are confident that our stylized fact is not due only to movements of resources away from agriculture and into manufacturing and services, since the U-shaped pattern holds within the manufacturing sector alone (UNIDO data set).

3 Trading Costs, Diversification and Economic Integration

We next turn to a model accounting for the U-shaped relationship between indices of specialization and the level of income per capita.[23] There is a continuum of goods, indexed by z, over $[0, 1]$. Without loss of generality, varieties of goods are ordered so that the (small) domestic economy is increasingly productive in z. Agents derive utility from consumption of all varieties, and domestic and foreign produced goods of the same variety are perfect substitutes. As a result, the economy produces only a subset of the continuum of goods, over which it is most productive. Furthermore, that range overlaps with that of goods produced in the rest of the world in the presence of transport costs (see Figure 12).

As the country catches up with the technological frontier, domestic aggregate productivity rises, and so does the number of goods produced domestically since the range where the domestic economy holds a comparative advantage expands. As a result, the domestic price schedule shifts to the Soutwest, and this is equivalent to a decrease in model-implied indices of specialization. On the other hand, the accumulation of infrastructures leads to a fall in transport costs. As this occurs the number of goods produced domestically tends to fall as the price schedule shifts to the Northeast, and concentration increases again. Put differently, the presence of transport costs forces diversification beyond comparative advantage. In this section, we aim at combining those dynamics to account for the empirical evidence documented previously.

[23] The starting point for the static model is Dornbusch, Fischer and Samuelson (1977).

3.1 Static Model

3.1.1 Demand

For simplicity, we choose to minimize the effects of demand-driven phenomena, and opt for Leontief preferences over all varieties. Choosing more general preferences would only obscure our point without modifying the basic economic intuition. Each period, agents in country 1 (the domestic economy) choose $X_1(z)$ to solve:

$$\min(X_1(0), ..., X_1(1)) s.t. \int_0^{z_1} p_2^1(z)\ X_1(z)\ dz + \int_{z_1}^1 p_1^1(z)\ X_1(z)\ dz = E_1$$

where $p_i^j(z)$ denotes the price of variety z produced in i and consumed in j, E_1 is total domestic (nominal) expenditure, and z_1 (z_2) denote the threshold goods that verify $p_2^1(z_1) = p_1^1(z_1)$ ($p_1^2(z_2) = p_2^2(z_2)$). In other words, it is cheaper to produce abroad those varieties between 0 and z_2, whereas the domestic economy produces all goods between z_1 and 1; of course, in the presence of trading costs, there is a range of non-traded goods going from z_2 to z_1. Since for all z, z', $X_1(z) = X_1(z') \equiv X_1$, the budget constraint becomes: $P\ X_1 = E_1$, where $P = \int_0^{z_1} p_2^1(z)dz + \int_{z_1}^1 p_1^1(z)dz$ is the domestic consumer price index. The foreign equivalent is straightforward, with $X_2 \left[\int_0^{z_2} p_2^2(z)dz + \int_{z_2}^1 p_1^2(z)dz\right] = E_2$.

3.1.2 Supply

Varieties are indexed by z, and without loss of generality, ordered so that sector-specific productivity $a(z)$ increases in z. Furthermore, A denotes general productivity level relative to the rest of the world, and is assumed to affect indiscriminately all sectors.[24] Perfectly competitive firms produce output by combining capital $K(z)$ and labor $L(z)$ using a constant returns to scale technology, so that producers of variety z choose $K(z)$ and $L(z)$ to maximize their profits:

$$\Pi(z) = p_1^1(z)\ a(z)\ A\ K(z)^\alpha\ L(z)^{1-\alpha} - wL(z) - rK(z) \tag{1}$$

where $\alpha < 1$, w is the wage paid in the domestic economy, and r is the (world) rate of interest. Each of the factor prices are equal across sectors,

[24] One way to interpret the inclusion of these two productivity indices in the production function is to view a(z) as purely domestic, sector-specific productivity, whereas A reflects the country's average distance from the "world frontier".

so that the less productive sectors do not disappear in equilibrium. Thus, for all z:

$$\frac{L(z)}{K(z)} = \frac{r}{w}\ \frac{1-\alpha}{\alpha} \tag{2}$$

The capital-labor ratio must be constant across sectors. Optimal choice of capital then implies $p_1^1(z) = \frac{1}{a(z)\ A} \frac{r}{\alpha} \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha}\right)^{\alpha-1}$. We have $p_1^2(z) = \psi\ p_1^1(z)$, where $\psi > 1$ denotes the cost associated with importing goods from abroad. We assume foreign prices are given exogenously to the small economy, with $p_2^1(z) = \psi\ p_2^2(z) \equiv \psi\,\Gamma$.

3.1.3 Equilibrium

Production in each sector must verify the following resource constraint

$$Y(z) = a(z)\ A\ K(z)^{\alpha}\ L(z)^{1-\alpha} = X_1(z) + X_2(z) = X_1 + X_2 \tag{3}$$

where the last equality comes from Leontief preferences. Then, given our assumptions on preferences and (2):

$$L(z) = \frac{1}{a(z)\ A}\ \left[X_1 + X_2\right] \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha}\right)^{\alpha} \tag{4}$$

Solving for the Distribution of Labor Our evidence, as well as our theoretical argument, is of a long-run nature. Although the sectoral concentration of activity presumably varies at high frequencies as well, our purpose is to analyze the determinants of long-run trends in that reallocation. We therefore assume trade is balanced, which requires:

$$\int_{z_2}^{1} p_1^2(z)\ X_2(z)dz = \int_0^{z_1} p_2^1(z)\ X_1(z)dz \tag{5}$$

Given firms' profit maximization, and the exogeneity of foreign prices, this can be rearranged to yield:

$$X_2\ = \frac{z_1\ \Gamma}{\int_{z_2}^{1} \frac{dz}{a(z)\ A} \frac{r}{\alpha} \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha}\right)^{\alpha-1}}\ X_1 \equiv \xi\ X_1 \tag{6}$$

We are now equipped to solve for the equilibrium of this economy. Substituting (6) in (4),

$$L(z) = \frac{X_1}{a(z)\ A}\ \left[1+\xi\right] \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha}\right)^{\alpha} \tag{7}$$

Aggregating over sectors, defining $L = \int_{z_1}^{1} L(z)\ dz$, and substituting the resulting expression for X_1 back into (7) yields:

$$L(z) = \frac{L}{a(z)} \ \left(\int_{z_1}^{1} \frac{dz}{a(z)} \right)^{-1} \tag{8}$$

This is the key relation in the paper. Since consumers demand an equal quantity of each variety, labor input in each sector is inversely related to productivity, and labor is concentrated in those sectors that are least productive. More interestingly, the threshold variety z_1 enters positively: as z_1 increases, the number of sectors effectively produced domestically falls, so that labor input in each variety can be higher. As we will show later, this implies that model-induced measures of sectoral labor concentration will rise.

Solving for Aggregate Output Define domestic nominal aggregate output

$$P\ Y = \int_{z_1}^{1} p_1^1(z)\ Y(z)\ dz = \frac{r}{\alpha} \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha} \right)^{\alpha-1} \int_{z_1}^{1} A\ K(z)\ \left(\frac{L(z)}{K(z)} \right)^{1-\alpha} dz \tag{9}$$

Using the fact that sectoral capital-labor ratios are equal to the constant given in (2), and letting $K = \int_{z_1}^{1} K(z)\ dz$, the expression simplifies to:

$$P\ Y = \frac{r}{\alpha} A\ K \tag{10}$$

Furthermore, the domestic price index P is given by

$$P = \psi\ \Gamma\ z_1 + \frac{1}{A} \frac{r}{\alpha} \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha} \right)^{\alpha-1} \int_{z_1}^{1} \frac{dz}{a(z)} \tag{11}$$

The impact of transport costs on the aggregate price index is ambiguous because of the presence of two offsetting effects: a direct positive one since the imported goods between 0 and z_1 become more expensive with ψ, and an indirect negative effect since z_1 falls in ψ, that is the range of imported goods shrinks with transport costs, and more is produced at home. For simplicity, we assume $a(z) = z^{\gamma}$ with $\gamma > 1$, so that:[25]

$$P = \psi \Gamma z_1 + \frac{1}{A} \frac{1}{\gamma - 1} \left[z_1^{1-\gamma} - 1 \right] \tag{12}$$

[25] We also choose units so that $\frac{r}{\alpha} \left(\frac{r}{w}\ \frac{1-\alpha}{\alpha} \right)^{\alpha-1} = 1$.

Finally, since $p_1^1(z_1) = p_2^1(z_1)$ we have:

$$z_1 = [\Gamma \psi A]^{-\frac{1}{\gamma}} \tag{13}$$

Substituting (13) into (12), it is easy to see that the price index increases in ψ, with $\frac{\partial P}{\partial \psi} = \Gamma^{1-\frac{1}{\gamma}} \ A^{-\frac{1}{\gamma}} \ \psi^{-\frac{1}{\gamma}}$. Thus from (10), real aggregate output Y increases in capital, falls in the cost of exchanging goods with the rest of the world and is given by:

$$Y = \frac{r}{\alpha} \ (\gamma - 1) \ \frac{A^2}{\gamma \ [\Gamma \ \psi \ A]^{1-\frac{1}{\gamma}} - 1} \ K \tag{14}$$

3.1.4 Artificial Measures of Concentration

This subsection will show that model-implied Gini and Herfindahl indice are increasing in z_1: there is a natural and monotonic correspondence between sectoral diversification (measured by $1 - z_1$) in our model, and the specific indicators we apply to our data.[26] However, the data used to compute the Gini and Herfindahl indices are based on imperfect observations, since our our actual sectoral coverage is both truncated and discrete, as opposed to complete and continuous.

Insofar as we chose to maximize its coverage to include non OECD members, our data makes it impossible to differentiate between those sectors with missing observations and those where labor was effectively close to zero. Moreover, the sectoral coverage of the data for any given country varied through time. As a result, we chose to restrict our empirical analysis to an arbitrary number of sectors for each country.[27] This is an important source of truncation in the range of sectors under consideration.[28]

However, this turns out not to be a problem. We can show that the artificial indices implied by the model, even thus arbitrarily truncated, are monotonous and increasing in z_1 (i.e. decreasing in the number of sectors).

[26] We define the model induced Gini and Herfindahl indices as follows:
$HERF = \int_{z_1}^{1} \left[\frac{L(z)}{L}\right]^2 dz$ and $GINI = \frac{2\int_{z_1}^{1} L(z)dz}{1-z_1}$

[27] There are two levels to this arbitrariness: first, we obviously had to follow the ISIC classification of sectors, itself somewhat arbitrary in the way new or disappearing sectors are accounted for. Second, the official data were truncated, by choosing a number of sectors constant over time for each country. This was necessary to allow comparability of the shuffling measures through time.

[28] An additional important source of truncation is that the UNIDO 3 and 4 digit data only cover the manufacturing sector.

A rise (fall) in the number of sectors results in a lower (higher) value for the indices of concentration. To see this, consider (8) and compute an observed Herfindahl index $ObsHer$ assuming truncation over an arbitrary interval I:

$$ObsHer = \int_I \left[\frac{L(z)}{L}\right]^2 dz = \frac{\beta\,(\gamma-1)^2}{\left[(\psi\ \Gamma\ A)^{1-\frac{1}{\gamma}}-1\right]^2} \tag{15}$$

with $\beta = \int_I \frac{dz}{a(z)^2}$. The numerator in $ObsHer$ fails to take into account any change in the number of sectors, that are therefore merely reflected in the denominator. Thus, the observed Herfindahl index increases in z_1 and from its definition, decreases in both the transport cost and aggregate productivity. As A grows, measures of concentration will tend to fall, whereas lower transport costs result in more concentrated activity. The intuition is straightforward and apparent from Figure 12: the range of domestically produced goods increases with aggregate productivity as the domestic prices schedule shifts to the South-West, z_1 falls and the integral at the denominator in $ObsHer$ rises. In other words, the labor resources freed up by higher productivity make it possible to open up new sectors. The exact opposite reasoning holds for decreases in trading costs, as the wedge between international prices shrinks, and z_1 rises as a result. Very similarly, one can easily compute $ObsGini$, an arbitrarily truncated Gini coefficient, as:

$$ObsGini = \frac{2\int_I\ L(z)dz}{1-z_1} \tag{16}$$

which, as is the case with the Herfindahl index, rises in z_1. Therefore, a fall in the number of sectors -a rise in z_1- is associated with more apparent concentration, as measured by both artificial indices.[29]

3.2 Dynamics

In this subsection, we aim at characterizing the dynamics of diversification. We assume the government can levy taxes and invest the proceeds in a stock of "infrastructure" capital S_t, which reduces trading costs. Thus, output in (14) grows through three distinct influences: (i) standard capital accumulation on the part of a representative consumer, (ii) growth in A (i.e.

[29] Using the formulas in footnote 26, it is easy to show that the monotonically increasing relation between Gini, Herfindahl and z_1 does not hinge on the arbitrary truncation over an interval I. In other words, the truncation is not crucial to our result, although it mimics realistically the treatment we impose on the data.

technological catch-up relative to the rest of the world), (iii) decrease in trading costs ψ. As evident from equation (13), only the second and third influences have effects on the range of goods produced domestically: technological catch-up tends to widen the range of goods produced domestically (which translates into diversification), whereas lower trading costs make specialization desirable. Thus, our rendition of the U-shaped pattern is based on technological conditions related to the dynamic evolution of A and ψ. As we discuss below, these conditions are quite general and consistent with a wide body of literature on technological catch-up and the evolution of transport costs.

The consumer's problem Formally, we assume consumers are infinitely-lived with subjective discount rate ρ and unit elasticity of intertemporal substitution, so that they maximize:

$$\int_0^\infty \ln C_t \ e^{-\rho t} dt \text{ subject to } C_t = \Delta(\ \psi,\ A)\ K_t - I_t - G_t \tag{17}$$

with:

$$\begin{aligned} \Delta(\ \psi,\ A) &= \frac{(\gamma-1)\ \frac{r}{\alpha}\ A^2}{\gamma\ \left(\Gamma\ \psi\ A\right)^{1-\frac{1}{\gamma}} - 1} \\ \dot{K}_t &= I_t - \delta K_t \\ G_t &= \tau\ I_t \end{aligned}$$

where δ is capital depreciation and τ is the (exogenous) investment tax rate, whose proceeds are used to finance infrastructure S_t. We further assume that infrastructure depreciates at the same rate as productive capital, so that $\dot{S}_t = G_t - \delta\ S_t$ and $S_t = \tau K_t$. We now need to impose some structure on the variables affecting the sectoral allocation of employment. We assume transport costs ψ fall with infrastructure capital, with $\frac{\dot{\psi}}{\psi} = -f\left(S_t\right) = -f\left(\tau K_t\right)$ where $f(.) > 0$. Furthermore, we assume $\frac{\dot{A}}{A} = g\left(K_t\right)$, with $g(.) > 0$ and $g'(.) \leqslant 0$. This is consistent with the abundantly documented idea that the ability to import world technology growth decreases with the distance to the technological leader. We discuss these assumptions in more details below.

Consumers are assumed atomistic, so that the effect of K_t on the dynamic behavior of A and ψ is external to the optimizing consumer. As the components of $\Delta(\psi,\ A)$ are beyond the control of the consumer, the economy behaves in effect as if aggregate technology were linear in the capital stock, which grows at $\frac{1}{1-\tau}\,\Delta(\ \psi(K_t),\ A(K_t)\ ) - \delta - \rho$. The implied dynamics of output are however non-trivial and depend on a choice of functional forms for $f(.)$ and $g(.)$. They are of no direct relevance for our purpose.

Stages of Diversification Consider the expression for $z_1 = [\psi\,\Gamma\ A]^{-\frac{1}{\gamma}}$: as long as $\frac{\dot{A}}{A} + \frac{\dot{\psi}}{\psi} > 0$, z_1 falls and the country diversifies. Specialization on the other hand occurs when $\frac{\dot{A}}{A} + \frac{\dot{\psi}}{\psi} < 0$, i.e. when transport costs fall above and beyond exogenous productivity increases. The crucial assumption needed to reproduce the observed U-shape pattern is therefore that transport costs initially fall at a rate lower than technological catch-up. Put differently, we need to assume productivity growth does not translate one for one into lower trading costs. More formally:

$$\begin{aligned} \text{Diversification} &\Leftrightarrow g\,(K_t) > f\,(K_t) \\ \text{Specialization} &\Leftrightarrow g\,(K_t) < f\,(K_t) \end{aligned}$$

Thus, the sufficient conditions for a U-shaped pattern of diversification are the following:

$$\begin{aligned} 1)\ g\,(K_0) &> f\,(K_0) \\ 2)\ g'\,(K_t) &< f'\,(K_t) \end{aligned}$$

There is a variety of arguments making these conditions empirically plausible. Consider first the case where the impact of infrastructure accumulation on transport costs is linear, so that $f'(.) = 0$. The U-shaped pattern exists whenever we maintain the hypothesis that closing the technological gap becomes increasingly difficult as countries grow (i.e. $g'(.) < 0$). This is a recurrent assumption in the literature about the so-called "technological gap".[30]

Suppose in contrast that technological progress is itself linear in K, so that closing the technological gap involves a constant cost irrespective of the proximity to the technological leader. In this case, the non-monotonicity will obtain as long as $f'(.) > 0$. The actual shape of $f(.)$ is a matter of debate. In recent work, Hummels (1999) has argued forcefully that transport costs have decreased very little since World War II, and some have actually risen. He concludes that the bulk of the decrease seems to have occurred very recently, thus providing support in favor of a monotonically increasing functional

[30] See for instance, Nelson and Phelps (1966), Fagerberg (1997) or Barro and Sala-i-Martin (1997).

form for $f(.)$, perhaps even convex. The evidence in Hummels (1999) and the recurrent case for infrastructure building that pervades the literature on economic growth provide further support for this assumption.[31]

4 A Simple Test of the Model

If indeed the pattern of diversification-specialization is governed by Ricardian arguments, a country's propensity to be open to trade ought to be a determinant of the stages of diversification. In particular, ceteris paribus, closed economies will tend to find it optimal to diversify later than more open countries, as international exchanges remain more costly longer given a rate of relative productivity growth. A natural test of the model would therefore consist in comparing the stages of diversification between open and closed economies. To explore this avenue, we can compute the expression for openness as implied by our model, which is given by:

$$\frac{X+M}{Y} = \frac{[1+\xi]\ X_1\ P}{\frac{r}{\alpha}\ K} \tag{18}$$

Using the expression for X_1 in (7), simplifying and solving for z_1, we get

$$z_1 = \left[1 - \frac{\gamma - 1}{A\ \left(\frac{X+M}{Y}\right) - 1}\right]^{\frac{1}{\gamma-1}} \tag{19}$$

This expression has at least three implications. First, since z_1 increases in $\frac{X+M}{Y}$, open economies tend to display higher degrees of specialization at all income levels. This prediction is however unsurprising and probably shared with a wide class of Ricardian models. Second, the non-linearity is not explained only by the degree of openness, even within countries. It results from the confluence of two forces, relative productivity growth A and transport costs ψ. Simply including the degree of openness in the set of independent variables in the estimations described in Section 2 will not constitute a conclusive test because, in equation (19), the ratio of imports plus exports to GDP is an endogenous variable. As a result, the non-linearity linking income and sectoral concentration could still prevail when controlling for the level of openness.[32] Finally, equation (19) shows how diversification

[31] See for instance Kelly (1997), Murphy, Shleifer and Vishny (1989) or Bougheas, Demetriades and Morgenroth (1999), among many others.

[32] We actually carried out this test, and inclusion of the openness variable had little effect on the U-shaped pattern.

responds endogenously to both relative productivity levels and openness. Investigating the dynamic impact of openness on the stages of diversification would require satisfactorily accounting for differences in productivity levels across countries, which is beyond the scope of this paper. Instead, we need an exogenous measure of the extent of transport costs, reflecting a country's "natural" degree of openness, prevailing at all levels of aggregate productivity.

To address this issue, we constructed the gravity component of a country's trade to GDP ratio by regressing the latter on a set of gravity variables, such as country size measured by population and area, whether the country is landlocked, an island, and so on. Table 9 presents such regressions. The results are in line with priors: larger countries tend to be more closed to trade, as are landlocked countries, while small countries and islands tend to be more open to trade. Using the fitted values from the regressions in Table 9, we construct a measure of a country's exogenous degree of openness.[33] We then split the sample between observations with predicted gravity openness greater and smaller than the sample median, providing us with a set of open and closed economies. *Ceteris paribus,* closed economies will tend to have lower $\Delta(\psi, A)$ and thus a lower marginal product of capital. This results in slower capital growth, i.e. slower productivity growth and build-up of infrastructures. Hence, these economies will diversify over a longer period.[34]

We then reproduce the non-parametric estimation described in Section 2 for these two subsamples. Figures 13-16 present results for the ILO and UNIDO three-digit data. First, as expected open countries generally tend to have more concentrated sectoral structures than closed countries, with higher degrees of specialization at all income levels in three of the four figures. Second and more importantly, open countries start to reconcentrate earlier than closed countries: this is particularly clear in Figures 13 and 14, where the minimum point is around \$8,800 for closed economies against roughly \$5,000 for open countries. The ILO results are somewhat less clear-cut: Figure 15 is surprising, in that it points to more specialization in closed economies, while Figure 16 remains supportive of the fact that countries with a high degree of "natural" openness tend to reconcentrate earlier. As a corollary, the slope of the U-curve in the reconcentration stage is more

[33] Such a measure is almost entirely time invariant since only the log of population, among the variables entered as regressors in the gravity equation, varies through time.

[34] This happens because the (negative) growth rate of transport costs depends negatively on their level.

pronounced in open countries. This is even more apparent when using the Herfindahl index. While these results do not constitute a complete test, they provide suggestive evidence that the degree of exogenous trade openness is related to the stages of diversification in a way that is consistent with our theory.

5 Concluding Comments

Using a wide panel of countries, we inform the debate about the stages of sectoral labor diversification across time and across countries. The empirical verdict is clear. Diversification is at first a priority for poor countries, where new sectors are developed. It is not until they have grown to relatively high levels of per capita income that incentives to specialize take over as the dominant economic force, and labor appears to re-concentrate into fewer sectors. This stylized fact is the first contribution of this paper.

We present a model where the number of active sectors arises endogenously as a result of ongoing (relative) productivity growth and prevalent trading costs: productivity increases translate into more sectors and less concentration, whereas falls in transport costs result in fewer sectors and higher concentration. The theory expands on the Ricardian model by including continuous trading costs, endogenizing these costs and introducing growth dynamics. Under general conditions on the nature of technological progress and on the dynamic evolution of transport costs, the theory replicates the stages of diversification found in the data, and their relationship to trade openness. This is the paper's second main contribution.

Appendix 1. Data Coverage

A. Sectoral Coverage

1. ILO 1-Digit Classification (9 sectors)

1. Agriculture, Hunting, Forestry and Fishing
2. Mining and Quarrying
3. Manufacturing
4. Electricity, Gas and Water
5. Construction
6. Wholesale and Retail Trade and Restaurants and Hotels
7. Transport, Storage and Communication
8. Financing, Insurance, Real Estate and Business Services
9. Community, Social and Personal Services

2. UNIDO 3-Digit Classification (28 sectors)

300 Total manufacturing
311 Food products
313 Beverages
314 Tobacco
321 Textiles
322 Wearing apparel, except footwear
323 Leather products
324 Footwear, except rubber or plastic
331 Wood products, except furniture
332 Furniture, except metal
341 Paper and products
342 Printing and publishing
351 Industrial chemicals
352 Other chemicals
353 Petroleum refineries
354 Miscellaneous petroleum and coal products
355 Rubber products
356 Plastic products
361 Pottery, china, earthenware
362 Glass and products
369 Other non-metallic mineral products
371 Iron and steel
372 Non-ferrous metals

381 Fabricated metal products
382 Machinery, except electrical
383 Machinery, electric
384 Transport equipment
385 Professional and scientific equipment
390 Other manufactured products

3. OECD 2-Digit Classification (20 sectors)

100. Agriculture, hunting, forestry and fishing
200. Mining and quarrying
310. Food, beverages and tobacco
320. Textiles, wearing apparel and leather industries
330. Wood and wood products, including furniture
340. Paper and paper products, printing and publishing
350. Chemicals, chemical petroleum, coal, rubber, plastic products
360. Non-metallic mineral products excl. products of petroleum & coal
370. Basic metal industries
380. Fabricated metal products, machinery and equipment
390. Other manufacturing industries
400. Electricity, gas and water
500. Construction
610+620. Wholesale trade and retail trade
630. Restaurants and hotels
700. Transport, storage and communication
810+820 . Financial institutions and insurance
830. Real estate and business services
910. Community and Social services
920. Personal services

B. Geographic Coverage

Argentina [a]
Australia [c]
Austria
Bangladesh
Barbados [b]
Belgium [b,c]
Benin [a]
Bolivia [b]
Brazil
Burkina Faso [a]
Burundi [a]
Canada [c]
Chile
China
Colombia
Costa Rica [b]
Cyprus
Czechoslovakia [a]
Denmark [c]
Dominican Rep.
Ecuador
Egypt
El Salvador
Ethiopia [a]
Finland [c]
France [b,c]
Gabon [a]
German Dem. Rep.[a]
Germany, Fed. Rep. [c]
Ghana a
Greece
Guatemala [a]
Haiti [a]
Honduras
Hong Kong [b]
Hungary
Iceland
India [a]
Indonesia [b]
Iran [a]
Ireland
Israel
Italy [c]
Jamaica [b]
Japan [c]
Jordan [a]
Korea
Kuwait [a]
Luxembourg [b]
Malaysia
Mexico [b]
Morocco [b]
Myanmar [b]
Netherlands [b,c]
New Zealand
Nicaragua [a]
Nigeria [a]
Norway [b,c]
Pakistan
Panama [b]
Paraguay
Peru [a]
Philippines
Poland
Portugal [b]
Puerto Rico [b]
Romania [b]
Saudi Arabia h
Senegal [a]
Seychelles [b]
Singapore [b]
South Africa [a]
Spain
Sri Lanka
Sweden [c]
Switzerland [b]
Syria [b]
Taiwan [a]
Thailand [b]
The Gambia [a]
Trinidad / Tobago [b]
Tunisia
Turkey
U.S.A [c]
U.S.S.R. [a]
United Kingdom [c]
Uruguay
Venezuela
Yugoslavia [a]
Zaire [a]
Zambia [a]

a: not in ILO dataset
b: not in UNIDO dataset
c: in OECD dataset

References

Acemoglu, Daron and Fabrizio Zilibotti (1997), *Was Prometheus Unbound by Chance ? Risk, Diversification, and Growth*, Journal of Political Economy, vol. 105, no. 4, pp. 709-751.

Barro, Robert J. and Xavier Sala-i-Martin (1995), *Economic Growth*, New York: McGraw Hill.

Barro, Robert J. and Xavier Sala-i-Martin (1997), *Technological Diffusion, Convergence, and Growth,* Journal of Economic Growth, Vol. 2, No. 1.

Bourgheas, Spiros, Panicos Demetriades and Edgar Morgenroth (1999), *Infrastructure, Transport Costs and Trade*, Journal of International Economics, no. 47, pp. 169-189.

Dornbusch, Rudiger, Stanley Fischer and Paul A. Samuelson (1977), *Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods*, American Economic Review, Vol. 67, no. 5, December, pp. 823-839.

Fagerberg, Jan (1997), A Technological Gap Approach to Why Growth Rates Differ, in Edward N. Wolff, ed., *The Economics of Productivity*, Elgar, New York.

Frankel, Jeffrey and David Romer (1999), *Trade and Growth: An Empirical Investigation*, American Economic Review.

Grossman, Gene M. and Elhanan Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.

Helliwell, John (1996), *How Much Do National Borders Matter?*, Brookings Institution Press, Washington, DC.

Hummels, David (1999), *Have International Transportation Costs Declined?*, Mimeo. University of Chicago.

Imbs, Jean (1999), *Co-Fluctuations*, CEPR Discussion Paper 2267.

International Labour Office (1997), Yearbook of Labor Statistics

Kelly, Morgan (1997), *The Dynamics of Smithian Growth,* Quarterly Journal of Economics, Vol. 112, p939-965.

Kim, Sukkoo (1995), *Expansion of Markets and the Geographic Distribution of Economic Activities: The Trends in U.S. Regional Manufacturing Structure, 1860-1987,* Quarterly Journal of Economics,vol. 110, p.881-908.

Kraay, Aart and Jaume Ventura (1998), *Comparative Advantage and the Cross-Section of Business Cycles*, mimeo, MIT.

Krugman, Paul (1991), *Geography and Trade*, MIT Press.

Krugman, Paul and Anthony Venables (1995), *Globalization and the Inequality of Nations*, Quarterly Journal of Economics, Vol. 110, 4, p. 857-880.

Lucas, Robert E. (1988), *On the Mechanics of Economic Development*, Journal of Monetary Economics, 22, July, 8-22.

Lucas, Robert E. (2000), *Some Macroeconomics for the 21st Century*, Journal of Economic Perspectives, Winter, 159-168.

McCallum, J. (1995), *National Borders Matter: Canada-US Regional Trade Patterns*, American Economic Review, June, 615-23.

Murphy, Kevin M., Andrei Shleifer and Robert W. Vishny (1989), *Industrialization and the Big Push*, Journal of Political Economy, vol. 87, no. 5, p. 1003-1026.

Nelson, Richard R. and Edmund Phelps (1966), *Technological Diffusion and Economic Growth*, American Economic Review, vol. 56, no. 1/2, March, pp. 69-75.

Organization for Economic Cooperation and Development (1998), International Sectoral Database (ISDB)

Romer, Paul (1986), *Increasing Returns and Long-Run Growth*, Journal of Political Economy, 94, 1002-37.

Sachs, Jeffrey and Andrew Warner (1995), *Economic Reform and the Process of Global Integration*, Brookings Papers on Economic Activity, 1, pp 1-95.

Solow, Robert (1970), *Growth Theory: An Exposition*, The Radcliffe Lectures, Oxford: Clarendon Press.

United Nations Industrial Development Organization (1997), Industrial Statistics Database, 3-Digit Level of ISIC Code.

Table 1 - Summary Statistics for the Sectoral Concentration Indices (pooled data)

ILO	Obs	# of countries	# of sectors[1]	Mean	Std. Dev.
Gini	885	64	9	0.479	0.098
Herfindahl	885	64	9	0.229	0.071
UNIDO					
Gini	1556	67	28	0.573	0.102
Herfindahl	1556	67	28	0.118	0.098
OECD					
Gini	356	14	20	0.531	0.042
Herfindahl	356	14	20	0.127	0.018

Table 2 - Correlation Matrix for the Sectoral Concentration Indices (pooled data)

	Gini
ILO (885 obs.)	
Herfindahl	0.759
UNIDO (1556 obs.)	
Herfindahl	0.828
OECD (356 obs.)	
Herfindahl	0.708

[1] Maximum number of available sectors used to compute the concentration indices.

Table 3 – Non-Parametric Estimates of the Inflexion Point ($)

ILO	Gini	Herfindahl
Inflexion Point	9,575	10,450
Y high*	10,950	10,825
Y low**	5,875	5,725
UNIDO		
Inflexion Point	8,675	8,825
Y high*	6,550	5,075
Y low**	5,025	8,650
OECD		
Inflexion Point	9,250	9,175
Y high*	8,950	8,725
Y low**	8,375	8,325

* Interval midpoint income level above which all estimated coefficients on income are statistically significant at the 5% level, and positive.
** Interval midpoint income level below which all estimated coefficients on income are statistically significant at the 5% level, and negative.

Table 4 - Fixed-Effects Regressions of Sectoral Concentration on Income and Income Squared (unbalanced panel)[2]

ILO	**Gini**	**Herfindahl**
Income	-0.027	-0.021
	(-12.11)	(-11.17)
Income^2	0.001	0.001
	(11.68)	(10.61)
Intercept	0.588	0.314
	(63.23)	(40.48)
# of obs.	885	885
# countries	64	64
R-sq	0.275	0.443
Inflexion Point ($)	11,548	11,722
UNIDO		
Income	-0.0104	-0.0095
	(-7.44)	(-5.76)
Income^2	0.0006	0.0005
	(9.14)	(5.83)
Intercept	0.5975	0.1455
	(134.97)	(27.63)
# of obs.	1556	1556
# countries	67	67
R-squared	0.287	0.166
Inflexion Point ($)	7,980	9,686
OECD		
Income	-0.030	-0.014
	(-11.67)	(-8.60)
Income^2	0.002	0.0007
	(13.44)	(10.00)
Intercept	0.659	0.184
	(47.64)	(21.49)
# of Obs	356	356
# of countries	14	14
R-squared	0.137	0.121
Inflexion point ($)	9,639	9,542

(t-statistics in parentheses)

[2] Throughout these tables, the data on per capita PPP income is entered in thousands of 1985 constant US dollars to facilitate readability of the numbers.

Table 5 - Fixed effects regressions on income for Income > estimated inflexion point

ILO	Gini	Herfindahl
Income	0.0041	0.0012
	(7.23)	(4.25)
Intercept	0.402	0.180
	(51.90)	(44.69)
# of Obs	193	185
# of countries	22	22
R-squared	0.071	0.017
UNIDO		
Income	0.0050	0.0016
	(15.75)	(7.57)
Intercept	0.4682	0.0645
	(128.36)	(24.74)
# of Obs	378	276
# of countries	22	17
R-squared	0.007	0.0006

(t-statistics in parentheses)

Table 6 - Fixed effects regressions on income for Income < estimated inflexion point

ILO	Gini	Herfindahl
Income	-0.0088	-0.0066
	(-7.06)	(-6.35)
Intercept	0.530	0.272
	(80.99)	(48.83)
# of Obs	692	700
# of countries	59	59
R-squared	0.251	0.338
UNIDO		
Income	-0.0056	-0.0042
	(-5.19)	(-3.89)
Intercept	0.6055	0.1403
	(174.40)	(35.52)
# of Obs	1178	1280
# of countries	62	65
R-squared	0.346	0.135

(t-statistics in parentheses)

Table 7 - Between Regressions of Sectoral Concentration on Income and Income Squared

ILO	Gini	Herfindahl
Income	-0.046	-0.042
	(-5.61)	(-5.85)
$Income^2$	0.0024	0.0022
	(4.65)	(4.86)
Intercept	0.651	0.361
	(27.45)	(17.48)
# of countries	51	51
R-squared	0.435	0.474
Inflexion Point ($)	9,645	9,608
UNIDO		
Income	-0.0627	-0.0441
	(-4.99)	(-2.94)
$Income^2$	0.0034	0.0023
	(3.89)	(2.20)
Intercept	0.7547	0.2532
	(25.09)	(7.05)
# of countries	51	51
R-squared	0.393	0.206
Inflexion Point ($)	9,220	9,616

(t-statistics in parentheses)

Table 8 – Random-Effects Regressions of Sectoral Concentration on Income and Income Squared (unbalanced panel)

ILO	Gini	Herfindahl
Income	-0.030	-0.022
	(-13.67)	(-11.99)
$Income^2$	0.0013	0.0009
	(12.74)	(11.02)
Intercept	0.619	0.314
	(52.54)	(31.14)
# of obs.	672	672
# countries	51	51
R-squared	0.417	0.451
Inflexion Point ($)	11,789	12,002
UNIDO		
Income	-0.0137	-0.0126
	(-9.32)	(-6.58)
$Income^2$	0.0007	0.0006
	(10.07)	(6.08)
Intercept	0.6269	0.1711
	(47.88)	(10.89)
# of obs.	1234	1234
# countries	51	51
R-squared	0.392	0.208
Inflexion Point ($)	9,184	10,718

(t-statistics in parentheses)

Table 9 – Regressions of Trade to GDP Ratios on Gravity Variables (between-country estimator)

	ILO	UNIDO
Intercept	1.819 (10.17)	1.439 (8.70)
Log Population	-0.076 (-3.43)	-0.080 (-4.10)
Log Area	-0.089 (-3.70)	-0.031 (-1.38)
Small country Dummy	0.081 (0.73)	0.003 (0.03)
Small Island Dummy	-0.202 (-1.38)	-0.092 (-0.66)
Island Dummy	0.034 (0.34)	0.109 (1.47)
Landlocked Country dummy	-0.080 (-0.99)	-0.052 (-0.59)
R-Squared	0.442	0.427
# of countries	137	64
# of Observations	2984	1482

(t-statistics in parentheses)

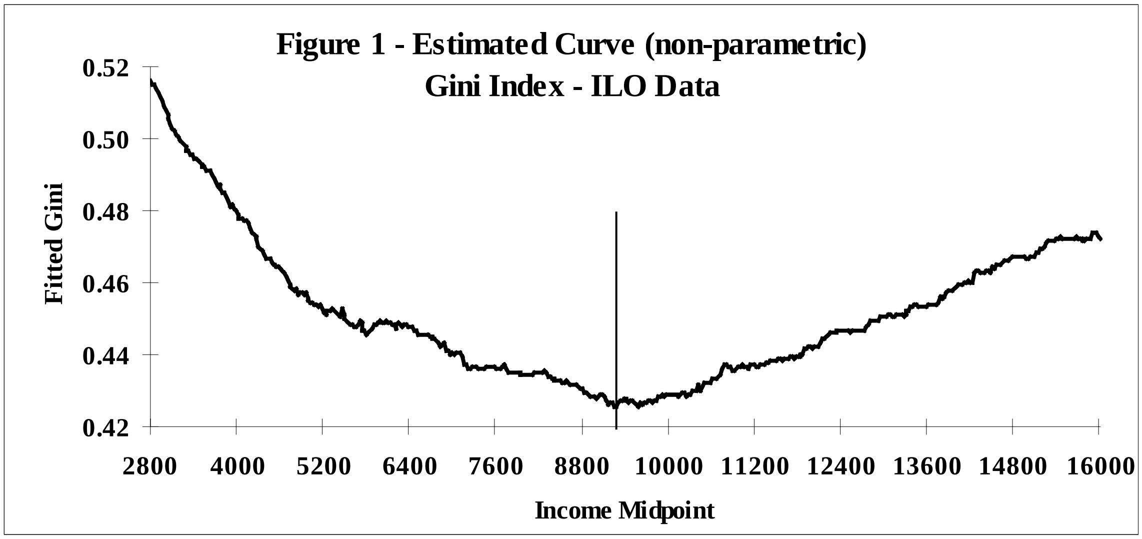


Figure 1 - Estimated Curve (non-parametric)
Gini Index - ILO Data
Fitted Gini
0.52
0.50
0.48
0.46
0.44
0.42
2800
4000
5200
6400
7600
8800
10000
11200
12400
13600
14800
16000
Income Midpoint

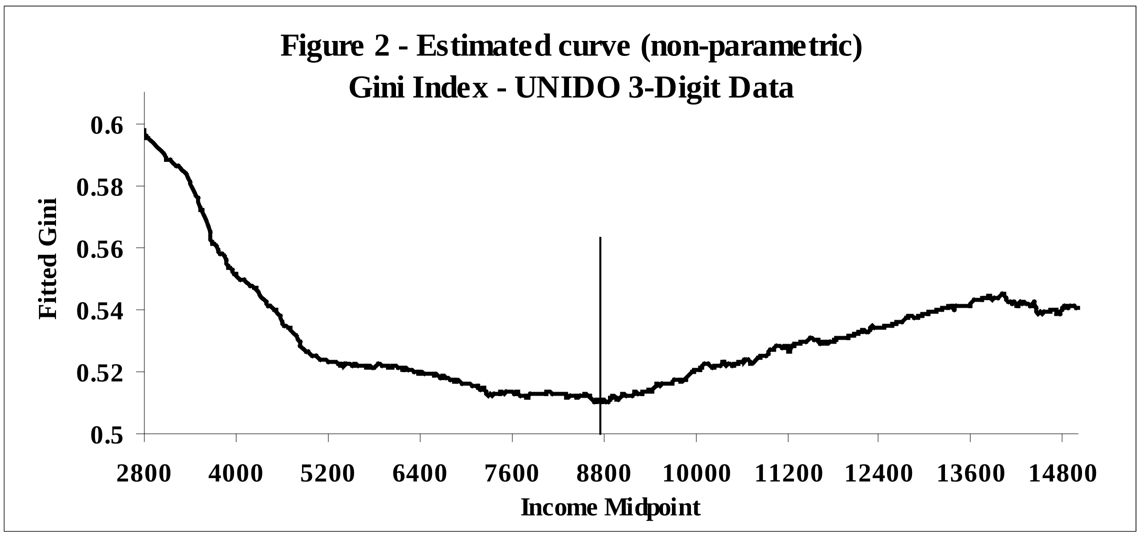


Figure 2 - Estimated curve (non-parametric)
Gini Index - UNIDO 3-Digit Data
Fitted Gini
0.6
0.58
0.56
0.54
0.52
0.5
2800
4000
5200
6400
7600
8800
10000
11200
12400
13600
14800
Income Midpoint

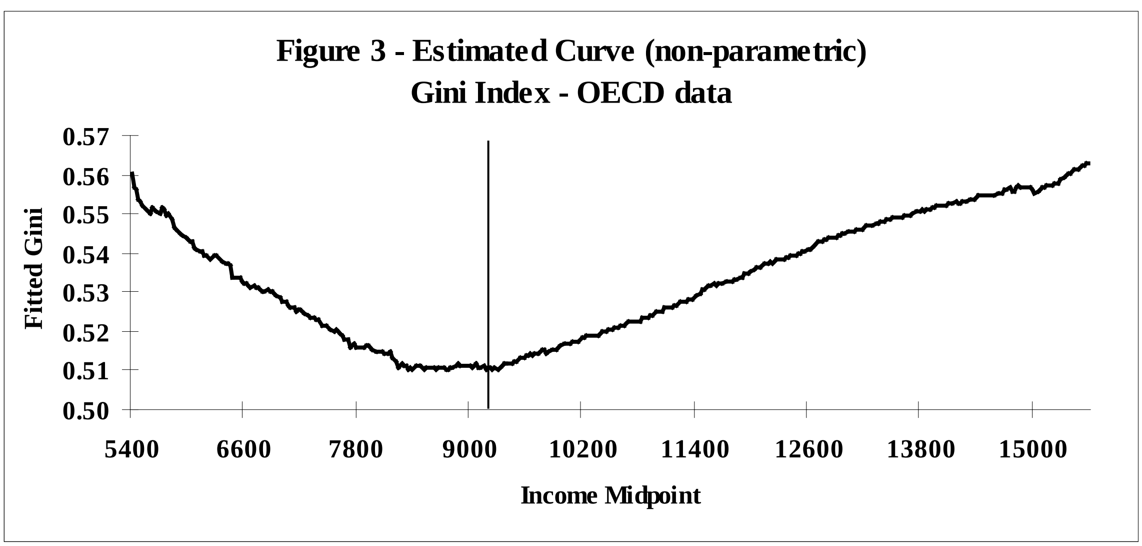


Figure 3 - Estimated Curve (non-parametric)
Gini Index - OECD data
Fitted Gini
0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.50
5400
6600
7800
9000
10200
11400
12600
13800
15000
Income Midpoint

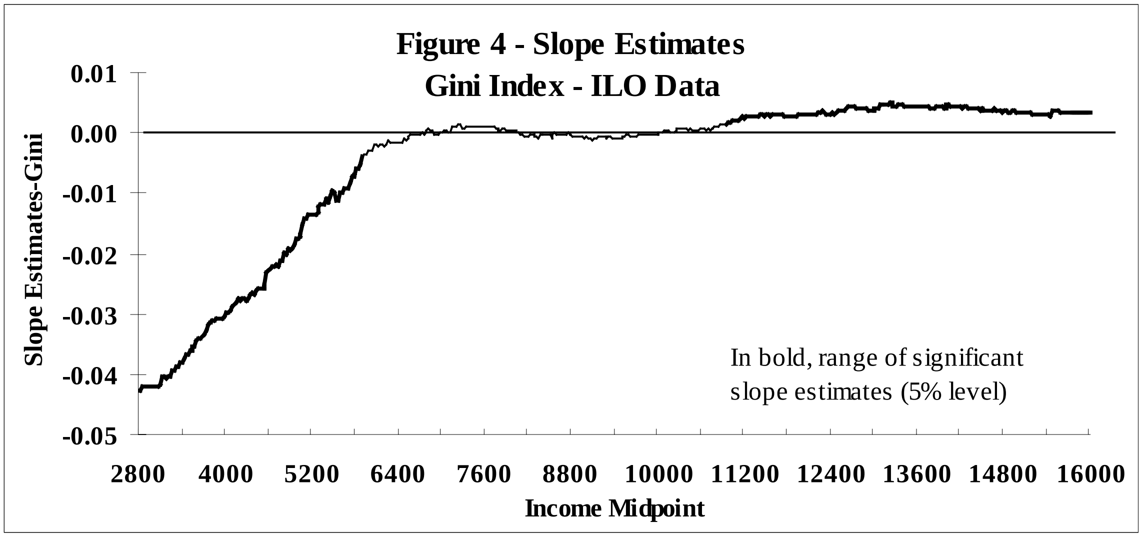


Figure 4 - Slope Estimates
Gini Index - ILO Data
Slope Estimates-Gini
0.01
0.00
-0.01
-0.02
-0.03
-0.04
-0.05
In bold, range of significant
slope estimates (5% level)
2800 4000 5200 6400 7600 8800 10000 11200 12400 13600 14800 16000
Income Midpoint

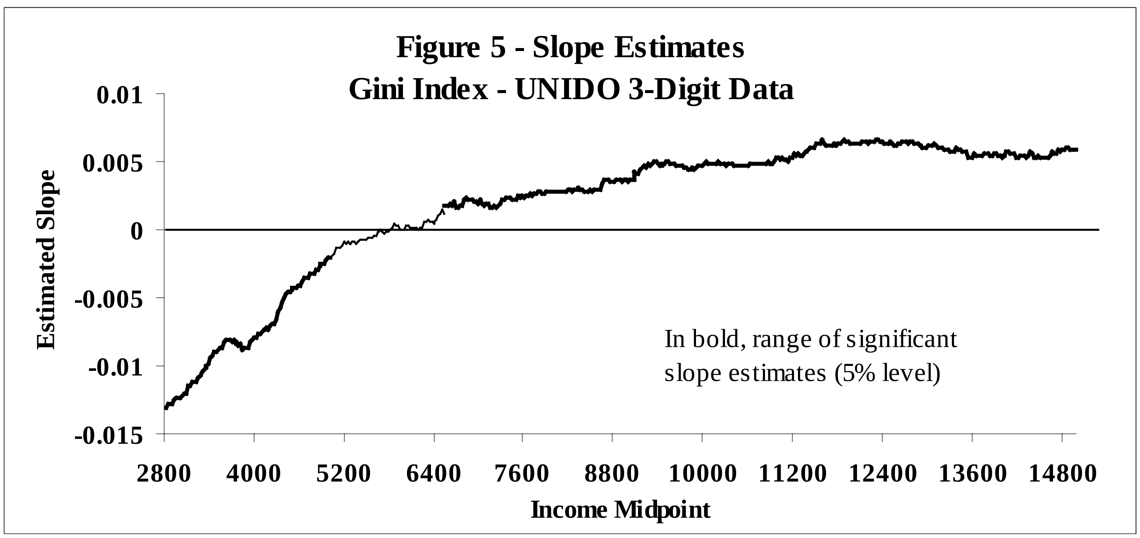


Figure 5 - Slope Estimates
Gini Index - UNIDO 3-Digit Data
Estimated Slope
0.01
0.005
0
-0.005
-0.01
-0.015
In bold, range of significant
slope estimates (5% level)
2800 4000 5200 6400 7600 8800 10000 11200 12400 13600 14800
Income Midpoint

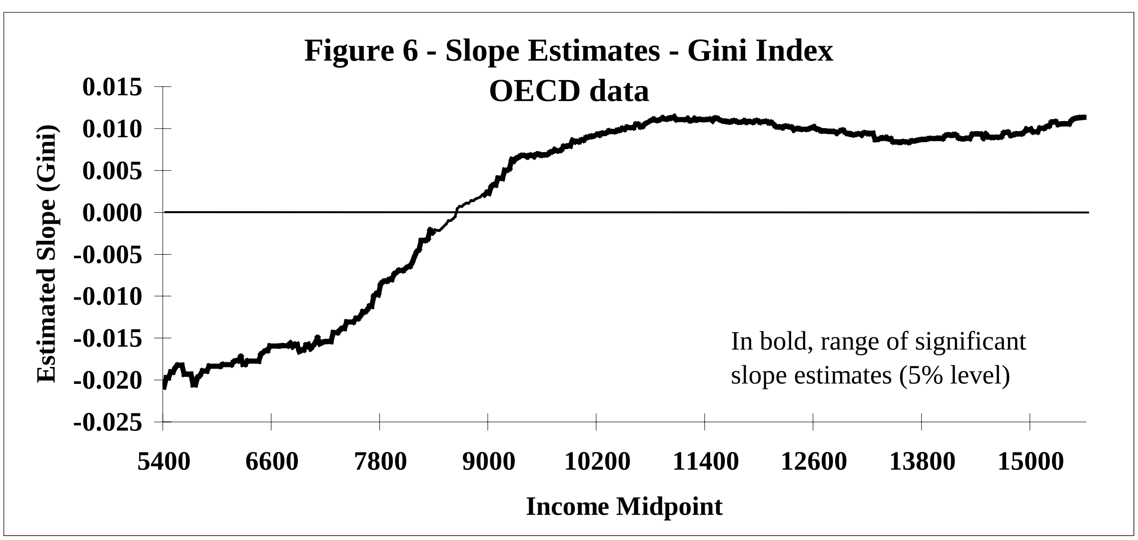


Figure 6 - Slope Estimates - Gini Index
OECD data
Estimated Slope (Gini)
0.015
0.010
0.005
0.000
-0.005
-0.010
-0.015
-0.020
-0.025
In bold, range of significant
slope estimates (5% level)
5400 6600 7800 9000 10200 11400 12600 13800 15000
Income Midpoint

Figure 7 - Within-country relationship - Gini Coefficient
ILO Data

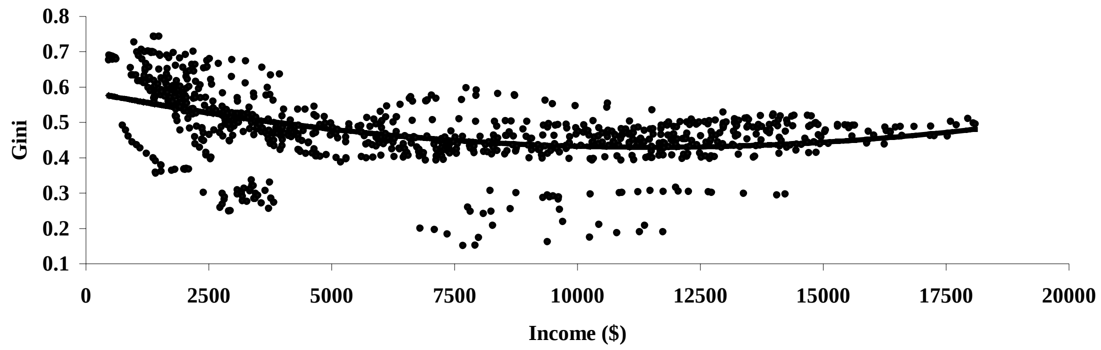


Figure 8 - Within-Country Relationship - Gini Coefficient
UNIDO 3-Digit Data

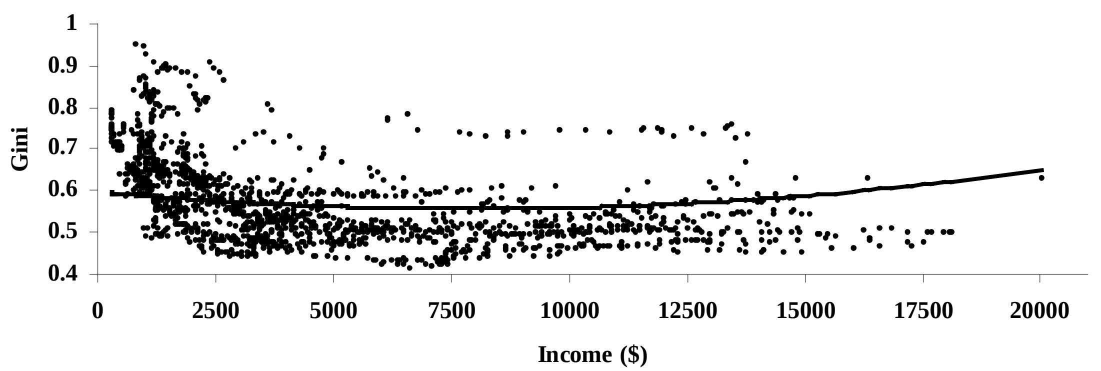
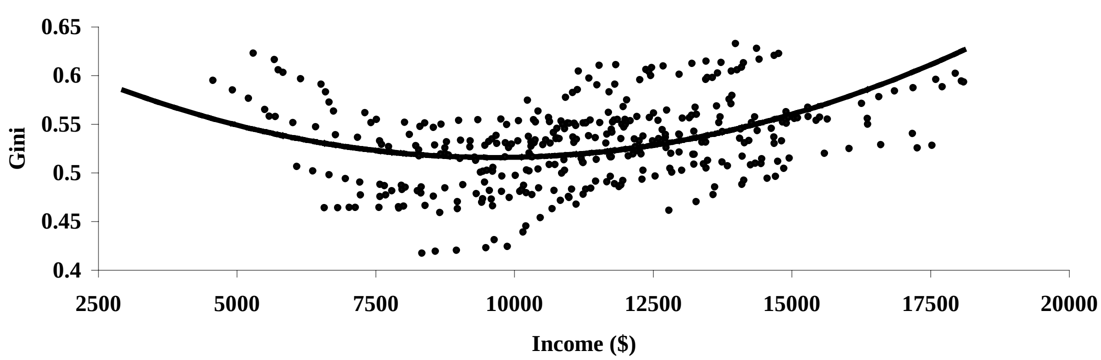


Figure 9 - Within-country relationship - Gini Coefficient
OECD Data



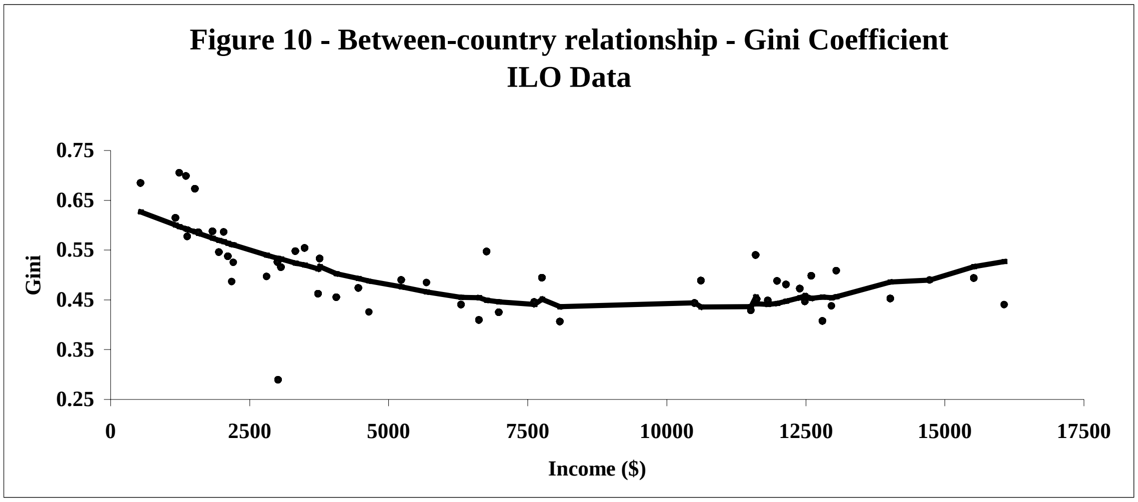


Figure 10 - Between-country relationship - Gini Coefficient
ILO Data
0.75
0.65
0.55
0.45
0.35
0.25
Gini
0
2500
5000
7500
10000
12500
15000
17500
Income ($)

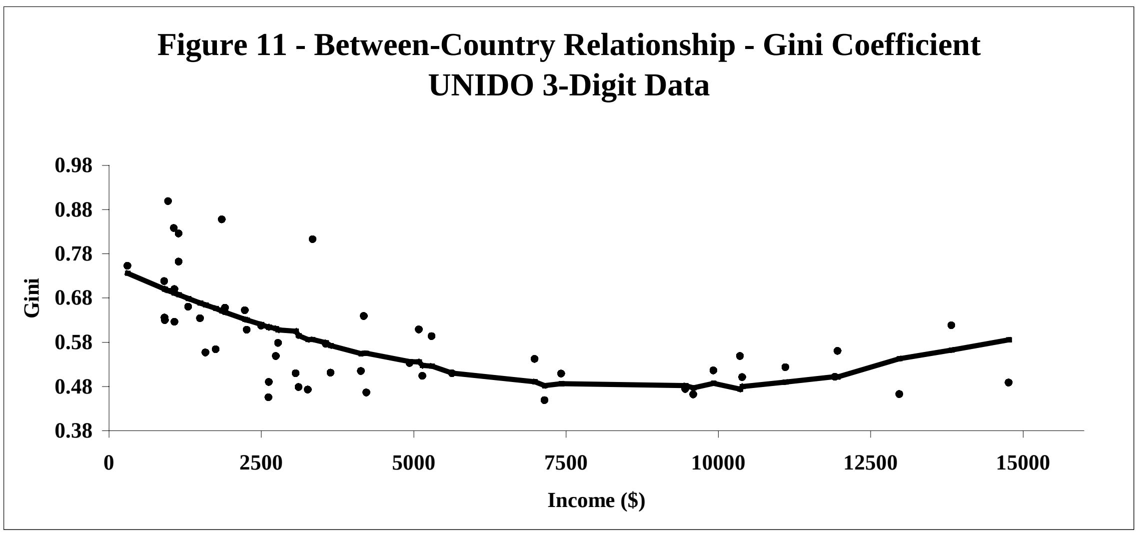
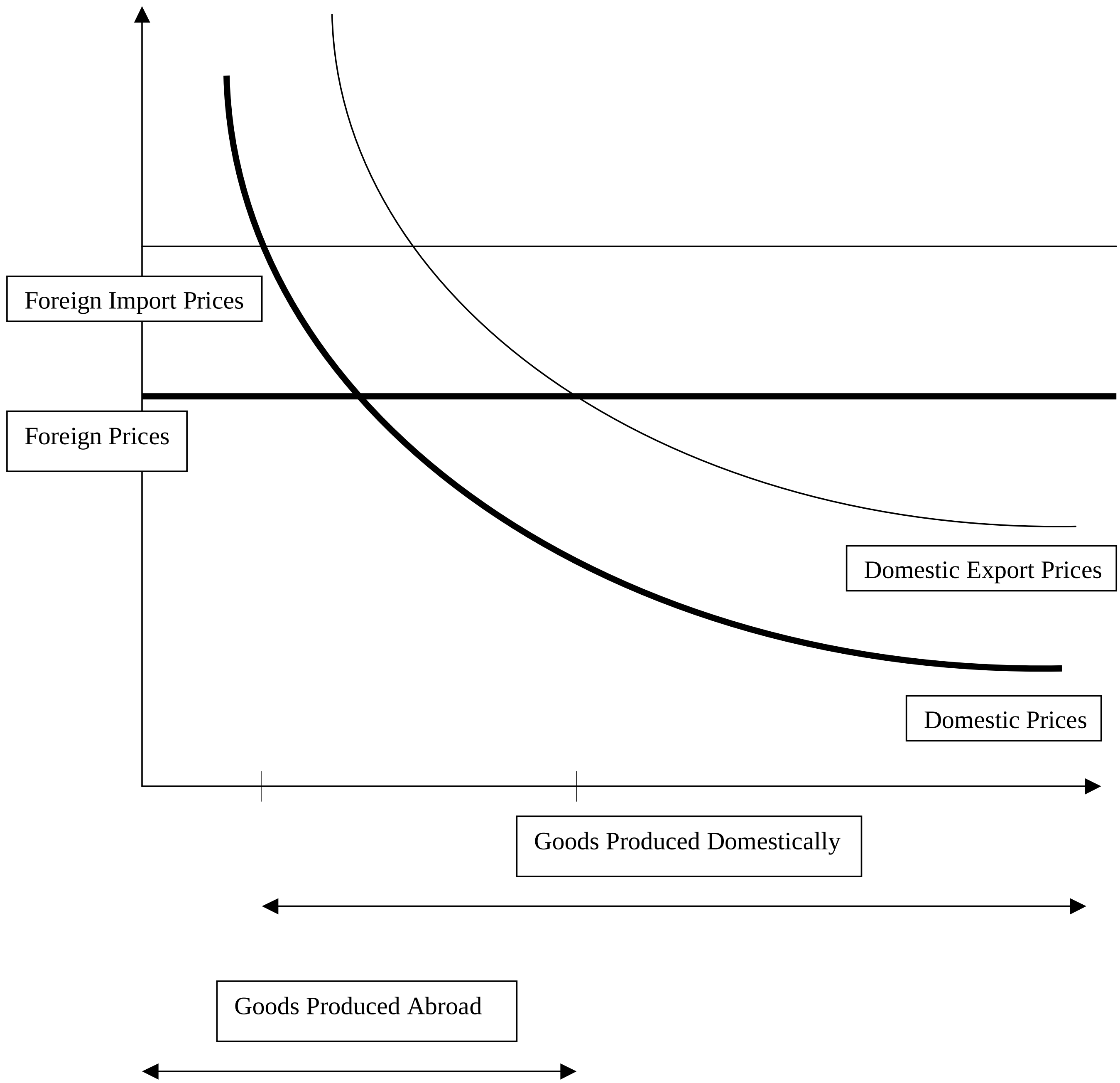


Figure 11 - Between-Country Relationship - Gini Coefficient
UNIDO 3-Digit Data
0.98
0.88
0.78
0.68
0.58
0.48
0.38
Gini
0
2500
5000
7500
10000
12500
15000
Income ($)

Figure 12: International Specialization Patterns



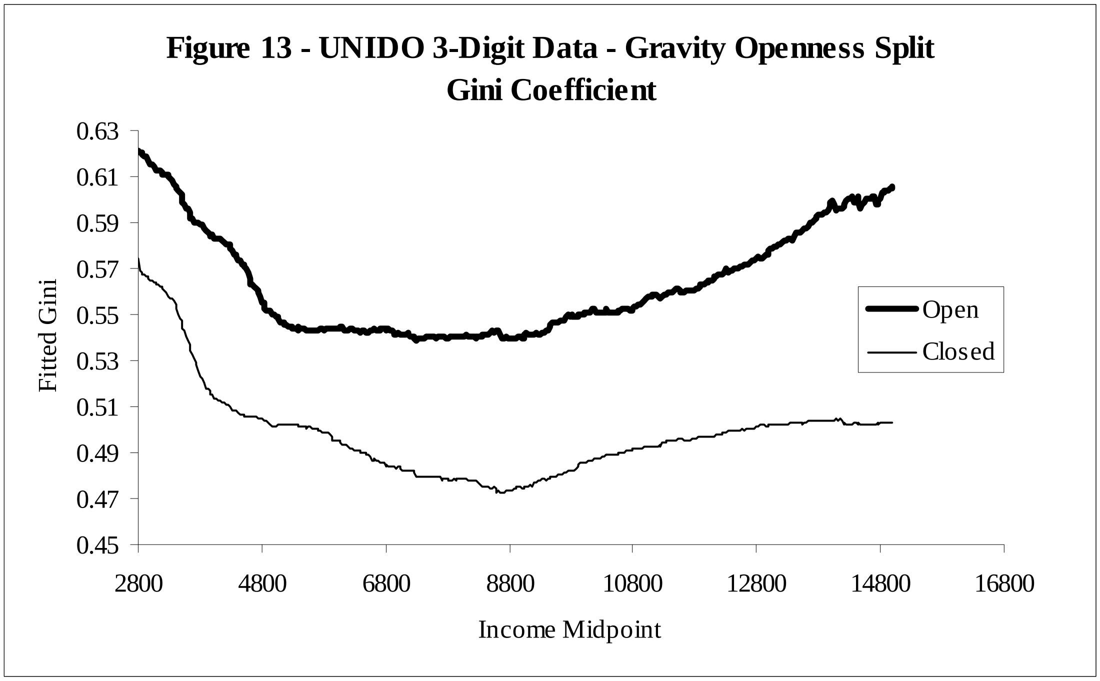


Figure 13 - UNIDO 3-Digit Data - Gravity Openness Split
Gini Coefficient
0.63
0.61
0.59
0.57
0.55
0.53
0.51
0.49
0.47
0.45
Fitted Gini
2800
4800
6800
8800
10800
12800
14800
16800
Income Midpoint
Open
Closed

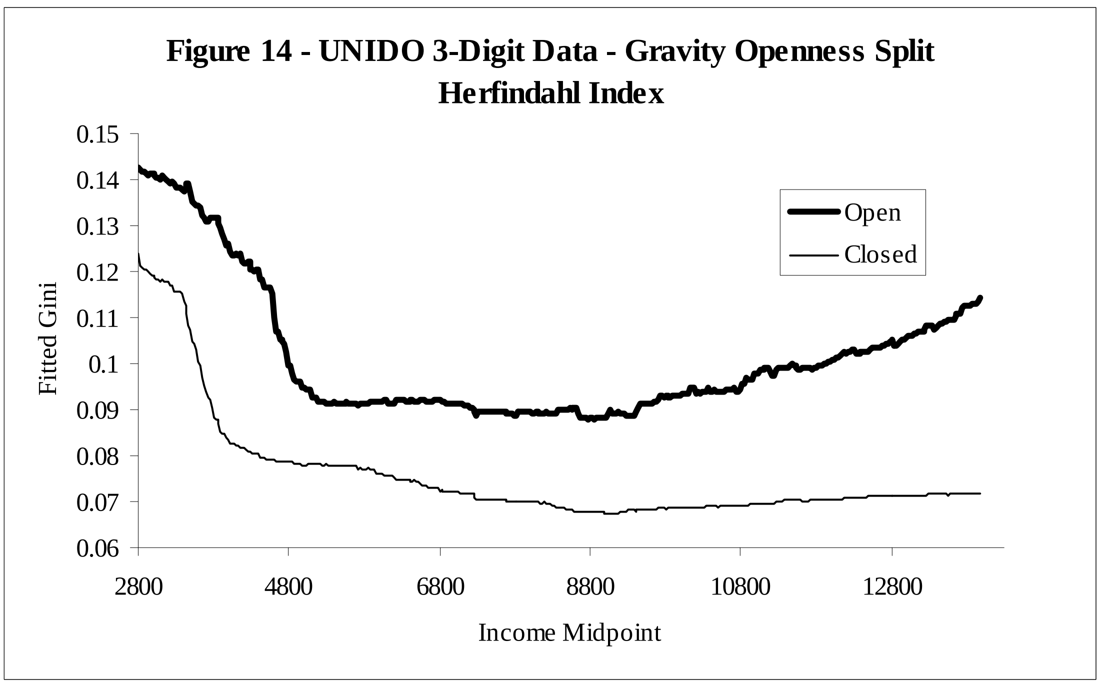


Figure 14 - UNIDO 3-Digit Data - Gravity Openness Split
Herfindahl Index
0.15
0.14
0.13
0.12
0.11
0.1
0.09
0.08
0.07
0.06
Fitted Gini
2800
4800
6800
8800
10800
12800
Income Midpoint
Open
Closed

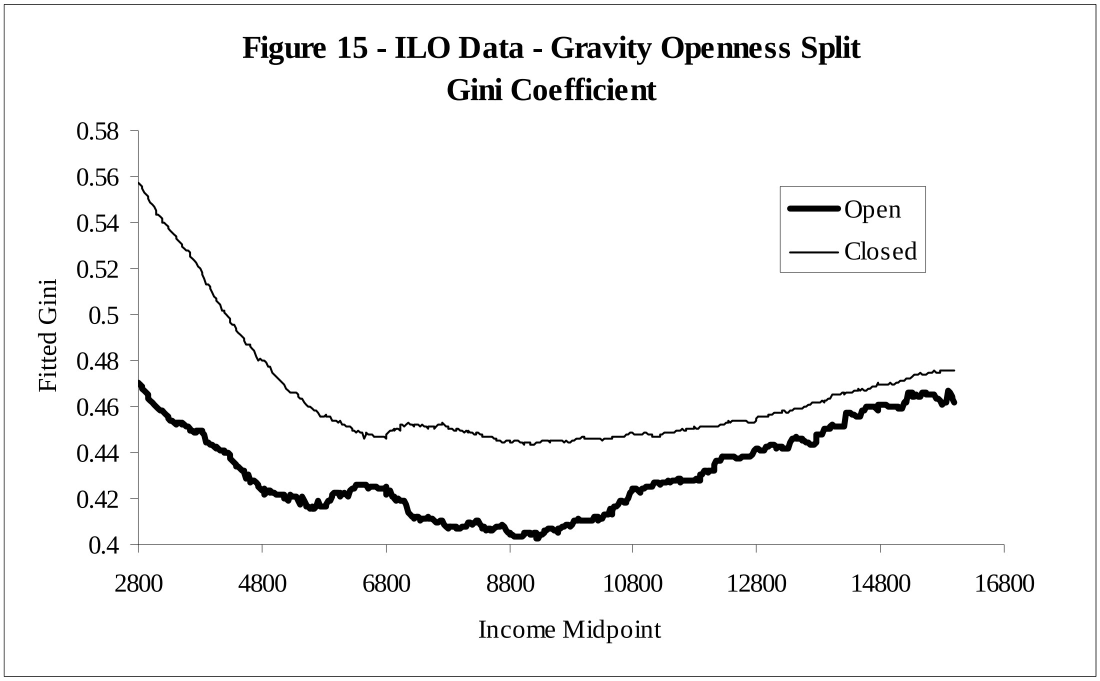


Figure 15 - ILO Data - Gravity Openness Split
Gini Coefficient
Fitted Gini
Income Midpoint
Open
Closed
0.58
0.56
0.54
0.52
0.5
0.48
0.46
0.44
0.42
0.4
2800
4800
6800
8800
10800
12800
14800
16800

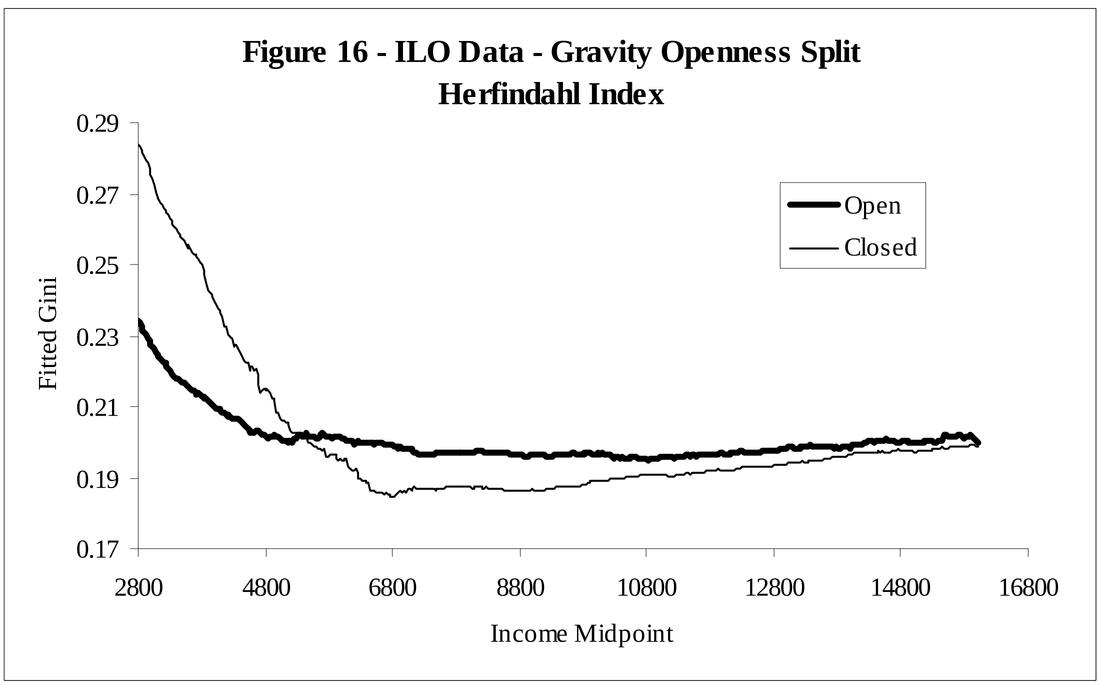


Figure 16 - ILO Data - Gravity Openness Split
Herfindahl Index
Fitted Gini
Income Midpoint
Open
Closed
0.29
0.27
0.25
0.23
0.21
0.19
0.17
2800
4800
6800
8800
10800
12800
14800
16800