Overall Specialization and Income:
Countries Diversify[*]

Luca De Benedictis[†] Marco Gallegati[‡] and Massimo Tamberi[§]

Abstract

This paper gives evidence to a stylized fact often disregarded in international trade empirics: countries' diversification. In the last fifteen years, the growth of world trade coexisted with the tendency of countries to reduce the specialization of their export composition along the development path. On average, countries do not specialize, they diversify. Our semiparametric empirical analysis shows how this result is robust to the use of different statistical indexes used to measure trade specialization to the level of sectoral aggregation and to the level of smoothing in the nonparametric term associated to income per capita. Using a General Additive Model (GAM) with country-specific fixed-effect, we show that, controlling for countries heterogeneity, sectoral export diversification increases with income.

Keywords: International Trade, Specialization, Development, Generalized Additive Models

JEL codes: C14, E32, F10

[†]DIEF - University of Macerata - Italy. debene@unimc.it

[‡]DEA - Università Politecnica delle Marche - Italy. marco.gallegati@univpm.it

[§]DEA - Università Politecnica delle Marche - Italy. m.tamberi@univpm.it

> *"... a country's economic growth may be defined as a long-term rise in capacity to supply increasingly diverse economic goods to its population ."*
> Simon Kuznets, Nobel Prize Lecture, 1971.

1 Introduction

In the last thirty years or so, the growth in the volume of world trade has been widely reported and dissected.[1] The number of countries trading with each others has increased, the number of categories of goods that has become internationally tradable has increased as well, and so has the volume of each category traded. Despite the debatable relative contribution of each component,[2] the coevolution of the *intensive margin*, the *extensive margin*, the variety/quality of internationally traded goods, and the number of countries actively trading internationally has modified the way countries contribute to the formation of international trade flows in terms of their own export structure. In essence, two opposite trends could have been resulting: on the one hand, the availability of a larger number of good through imports could have been pushing countries toward an higher degree of specialization of domestic productions and exports, on the other hand, also the reverse could have been possible, and the increase in the bundle of intermediate goods available through trade could have been encouraged the production and exports of new domestic goods, increasing the diversification of countries.

If one looks for some theoretical guidelines in international trade models the result would be mixed. Different theories do not share a common view about the way trade specialization evolves with per capita income along the development path. Indeed, close to models that predict that countries specialization should increase over the development path (Krugman, 1987), there are models that predict that countries economic development is associated with a low degree of specialization (Stokey, 1988), and more general frameworks (Peretto, 2003) have shown that both cases could happen according to the effect of international market integration on firms competition. Neither

---

[1]See the *American Economic Review Papers and Proceedings* 2003 section on international trade, and especially Broda and Weinstein (2003).

[2]See Helpman, Melitz and Rubinstein (2006) on the prevalence of the extensive margin, and Evenett and Venables (2002) on the role of new trading countries.

empirical works do help to clearly discriminate between alternative trade-based explanations of the link between specialization and development. In fact, the available evidence on this issue provides different answers depending on the data set, the measure of specialization and methodological analysis employed.[3]

For the purpose at hand, let us consider two countries - Italy and Indonesia -and observe their export structure in a certain time period. In the top panels of figure 1 we plotted the sectoral market shares of the two countries in 1985. Each single bar in the two panels identifies the total value of the country's sectoral export relative to the value of the world exports in that sector: *i.e.* over about 770 sectors (SITC rev.2 at 4-digits level) the Italian highest sectoral market share in 1985 was on "Fabrics, woven, of sheep's or lambs' wool or of fine animal hair" with a sectoral market share of 61 per cent; while Indonesian highest sectoral market share in the same year was on "Vegetable materials of a kind used primarily for plaiting" with a sectoral market share of 44 per cent. The horizontal line identifies the total export share of the country, *i.e.* in 1985 Italy's share of world exports was about 0.04, while Indonesia's was about 0.01. If we divide each sectoral value by the total export share of the country we obtain the quotients plotted in the bottom panels of figure 1. In that case the bars identify the value of the sectoral Balassa Index, and the horizontal lines depict the demarcation value of 1, above which a sector is characterized by revealed comparative advantages (RCA) (Balassa, 1965; De Benedictis and Tamberi, 2004).[4]

Three facts are worth noticing. First, the two countries export in sectors with and without RCA; second, Italian market shares are in general higher than Indonesia's; third, Italy does export in a larger set of sectors than Indonesia does. As far as our research question, the Italian export structure is characterized by an higher level of export diversification, while the Indonesian export structure is characterized by an higher level of export specialization. From an overall perspective export differentiation and export specialization are just antonyms, and Indonesia is more *overall specialized* than Italy just because it does export in a more *limited* number of RCA sectors than Italy does.[5] If we take this information as a simplistic cross-country induction,

[3] The set of papers that we will scrutiny in the following analysis includes Kim (1995), Amiti (1998), Brasili *et al.* (2000), Brülhart (2001), Proudman and Redding (2000), Redding (2002), Imbs and Wacziarg (2003), Kalemli-Ozcan, Sorensen and Yosha (2003), Koren and Tenreyro (2004), and De Benedictis, Gallegati and Tamberi (2006). A complementary strand of this literature points out that sectoral specialization patterns, rather than overall specialization itself, matter (Hausmann, Hwang and Rodrik, 2005).

[4]The two sectors previously quoted now take the values 15.25 for the Italian top sector, and 44.00 for Indonesia's.

[5]For a more broather discussion of the definition of Overall specialization and its rela-
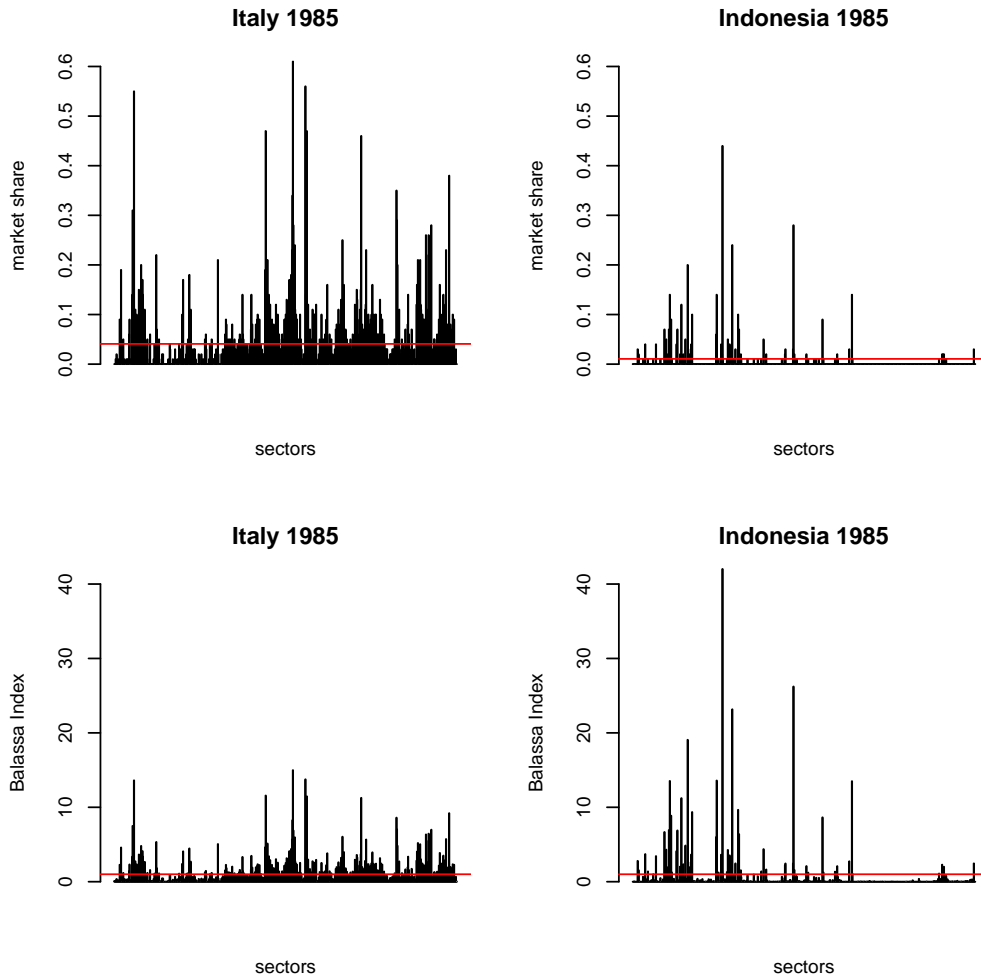
Figure 1: Sectoral specialization in Italy and Indonesia in 1985

what we will end up with is the proposition that countries diversify their export structure as they reach an higher level of income per capita. The same proposition is however more problematic if we analyze it along the time dimension.

In figure 2 we plotted the analogous of the bottom panels of figure 1 for the year 2001. From a simple visual inspection one can clearly state that Italy has not changed its exports structure much, on the contrary Indonesia has enormously increased the number of RCA sectors in which the country

---

tion with sectoral specialization see De Benedictis and Tamberi (2004).
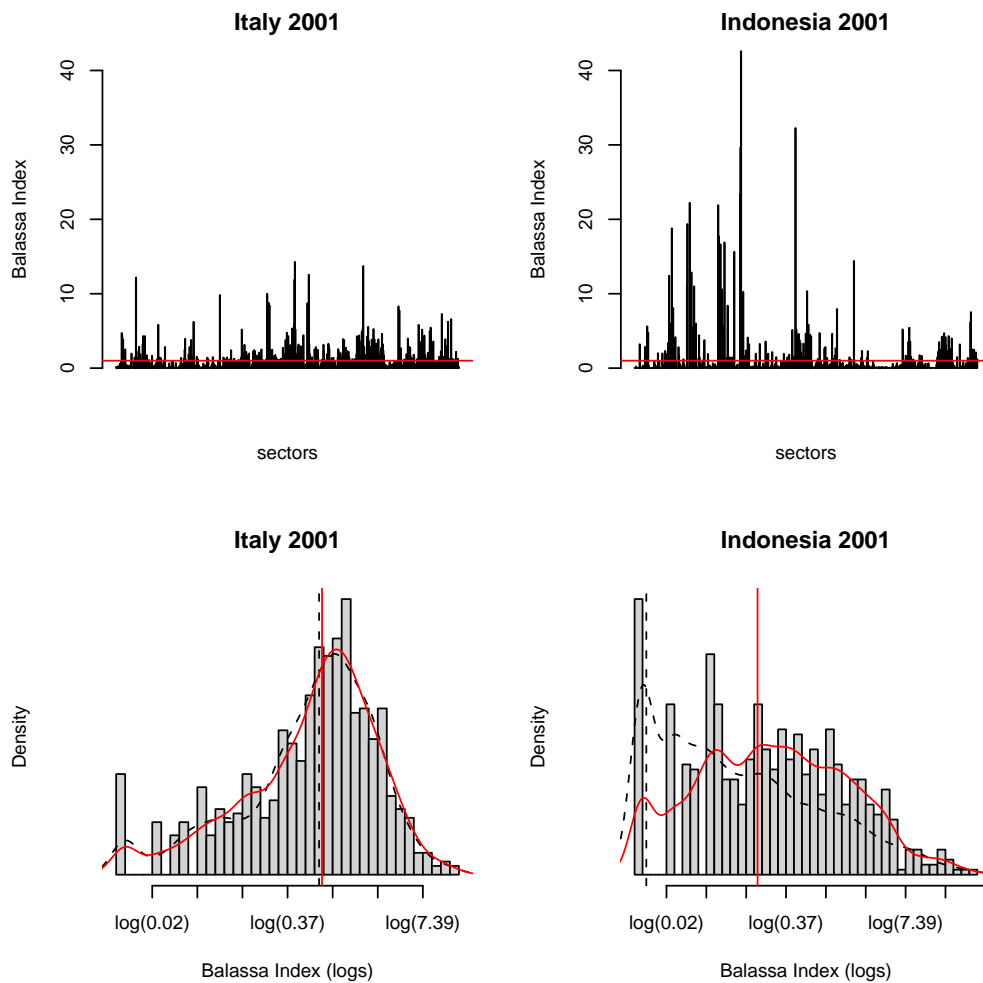
4

Figure 2: Overall specialization in Italy and Indonesia

exports, its level of overall specialization has reduced along time.[6] It is difficult to say if Italy still share a common trend with Indonesia in terms of increased differentiation. In order to make a definitive statement, we shall move to synthetic measures of the position of the distribution of the sectoral Balassa Index. In the bottom panels of figure 2 we plotted the histograms of the sectoral Balassa Index in 2001. The number of zeros (corresponding to

---

[6]Since the horizontal axis is ordered according to SITC codes, Italy is more sectorally specialized in sectors on the right of the SITC scale (manufactures) and Indonesia is more sectorally specialized in sectors on the left of the SITC scale (agricultural goods and basic manufactures).

the number of sectors in which the country does not export) is still higher for Indonesia, indicating the higher level of overall specialization previously emphasized. An even clearer picture emerges by looking at the estimated kernel densities for 1985 (the dotted bell-shaped lines) in comparison with the one estimated for 2001 (the continuous bell-shaped lines). The estimated distribution has moved to the right in the case of Indonesia, and has remained almost invariate in the case of Italy. If we look at the median of the distribution (the vertical line, dotted for 1985 and continuous for 2001) also Italy shows a marginal movement to the right. Therefore one can conclude that also along the time dimension countries tend to diversify their export structure (even if with different intensities).

This statement is however very strong and needs to be brought to more robust analysis. This is the task of this paper. We accomplish it exploring the link between overall specialization and per capita income using semiparametric panel techniques, and we perform a robustness analysis checking the sensitivity of our main result to alternative measures of specialization, to different levels of disaggregation within manufacturing exports, and to different smoothing parameters of the nonparametric term associated to income per capita. The remaining part of the paper is organized as follows. Section 2 reviews the existing empirical literature; section 3 presents our data set, the various measures of specialization and empirical methodology employed; section 4 provides empirical evidence on specialization dynamics and section concludes the paper.

2 Relation to existing empirical literature on specialization dynamics

Our paper relates to a particular brand of research that has adopted an empirical approach to analyze changes in countries' overall degree of specialization using *ad hoc* and/or atheoretic measures of specialization, such as the Revealed Comparative Advantage (RCA) or the location quotient, which represent, respectively, trade-based and production-based measures of specialization. Such studies on the specialization patterns differ according to several aspects: i) the measure of specialization, ii) the variables and level of aggregation of the data, and iii) the estimation methodology. A synopsis of the recent empirical literature on the relationship between specialization (diversification) and development is presented in table 1.

Columns 2 to 5 of Table 1 shows how the empirical analysis on the specialization-development relationship is not an homogeneous body of re-

Table 1: Recent empirical literature on specialization and development

Author and year	Time period	Variable	Number of Manufact. sectors	Countries	Index of specialization	Method	Degree of specialization
Kim (1995)	1960-1987	E	20	US states	Krugman	Par	Decrease
Amiti (1999)	1968-1990	P	65	5-10 EU	Gini	Par	Increase
Brasili *et al.* (2000)	1970-1995	T	66	14 Mix	RCAS	Par, Nonpar	Unchanged
Brülhart (2001)	1972-1996	P, E	32	13 European	Gini	Par	Mixed
Proudman and Redding (2000)	1970-1993	T	22	G-5	RCA	Par	Unchanged
Redding (2002)	1970-1990	P	20	7 OECD	GDP shares	Par	Unchanged
Imbs and Wacziarg (2003)	1969-1997	P, E	28	99 Mix	Gini and Herfindahl	Par,NonPar	U-shaped
Kalemli-Ozcan *et al.* (2003)	1977-1993	P	10	21 OECD	Square deviation	Par	Decrease
Koren and Tenreyro (2004)	1963-1998	P	28	42 Mix	Herfindahl (weighted)	Par, NonPar	U-shaped
De Benedictis *et al.* (2006)	1985-2001	T	539	40 Mix	BI median	Nonpar	Decrease

Note: E=Employment; P=Production; T=Trade. Par=Parametric; NonPar=Nonparametric. In the "Countries" column Mix means both OECD and non-OECD countries.
Brasili *et al.* (2000) use merchandise not manufacturing sectors.

7

search, as the employed data set differ in terms of: to: i) the time span, ii) the variables used to construct the measure of specialization (based on trade or production variables), iv) the set of sectors (economy-wide or manufacturing) and the level of disaggregation and, finally, iv) the set of countries considered n the analysis. The time span of the analyses covers the last 30 years of the XXth century (with the exception of the UNIDO database which starts from the early 60s) and with the exceptions of a small number of papers covering a broad sample of developed and developing countries[7] the samples generally include only a limited set of countries (usually a subset of developed countries). Both trade and production (and/or employment) data have been used in calculating indicators of specialization patterns in production. Indeed, if there are evident reasons to use trade data for analyzing intra-industry trade and production (employment) data for analyzing locational patterns, less consensus seems to characterize the empirical analysis of specialization patterns. Anyway, apart from Brasili, Epifani and Helg (2000), Proudman and Redding (2000), and De Benedictis, Gallegati and Tamberi (2006) which employs trade data, production and employment are the variables that have been used most frequently in constructing specialization indexes. These papers, as regards the sectors and the level of disaggregation of the data used to construct indicators of sectoral concentration, employ industry-level data at various levels of disaggregation (from 1 to 3-digit level depending on the database used, OECD-STAN, OECD-BTD, ILO, UNIDO).[8]

Column 6 of Table 1 shows that many different ways have been employed in the empirical literature to measure changes in the overall degree of specialization. Kim (1995) and Amiti (1999) use movements in the country specialization indexes, Gini and the weighted standard deviation of the Balassa index, to determine whether the industrial structure of each country has become more geographically concentrated. In Proudman and Redding (2000) and Redding (2002) the dynamics of a country's pattern of international specialization corresponds to the evolution of the external shape of the GDP shares distribution, represented by the entire cross-section distribution of the measure of specialization in an industry over time (the share of the industry in the country's GDP). Similarly, Brasili, Epifani and Helg (2000) analyze the dynamics of trade patterns of some developed and emerging countries studying the shape of the sectoral distribution and the evolution over time of an index of trade specialization (the symmetric revealed comparative advantage index, RCAS). Imbs and Wacziarg (2003) study the evolution

---

[7]Imbs and Wacziarg (2003), Koren and Tenreyro (2004), and De Benedictis, Gallegati and Tamberi (2006).

[8]The only exception is represented by Brasili, Epifani and Helg (2000) which uses data on total merchandise instead of manufacturing.

of sectoral concentration running simple fixed-effects linear regression of the measures of concentration on income (with country-specific effects) using, among a variety of measures of sectoral concentration, only those measures that make use of data on the entire distribution, such as the Gini coefficient, the Herfindahl index and the coefficient of variation of sector shares. Other papers[9] construct measures of specialization that take into account the potential relationship between risk and sectoral diversification. In particular, Kalemli-Ozcan, Sorensen and Yosha (2003) construct a measure of industrial specialization in production and calculate an index of specialization indexes, the "square deviations index", based on the difference between sector shares in a region and sector shares of the other regions in the same risk sharing group, while Koren and Tenreyro (2004) use both weighted and unweighted Herfindahl indexes of sectoral concentration, with the former index weighting sectors by their own volatilities (and in this way being not sensitive to the arbitrariness of sectoral classification). Finally, De Benedictis, Gallegati and Tamberi (2006) use the median of the sectoral empirical density function of the Balassa Index (BI) of Revealed Comparative Advantage as a summary measure of overall specialization.

Column 7 of Table 1 shows that, in order to characterize the evolution of the relationship between specialization and development both parametric and non-parametric techniques have been used. In particular, while the former strand of the empirical research has employed a parametric approach,[10] most recent empirical studies in this area have adopted nonparametric methods[11] The preference accorded to nonparametric techniques in recent empirical papers relates to the fact that nonparametric methods allows empirical researchers to explore the issues related to the shape and the statistical significance of the relashionship between specialization and income without making any a priori explicit or implicit assumption about the relationship.[12]

---

[9]Kalemli-Ozcan, Sorensen and Yosha (2003) and Koren and Tenreyro (2004).

[10]Kim (1995), Amiti (1999), Proudman and Redding (2000), Brasili, Epifani and Helg (2000), Redding (2002, Kalemli-Ozcan, Sorensen and Yosha (2003) and Koren and Tenreyro (2004).

[11]Imbs and Wacziarg (2003), Koren and Tenreyro (2004) and De Benedictis, Gallegati and Tamberi (2006).

[12]The non-parametric methodology employed in recent emprical studies (Imbs and Wacziarg, 2003, and Koren and Tenreyro, 2004) is a locally weighted scatterplot smoothing procedure called loess (Cleveland, 1979) a procedure that allows to determine a smoothed fitted nonparametric curve for representing the relationship linking sectoral concentration and income. A different nonparametric procedure, the Generalized Additive Model (from now on GAM), is employed in De Benedictis, Gallegati and Tamberi (2006). Such a model allows the empirical researcher to gain more flexibility, as it replaces the linearity assumption with some univariate smooth functions in a nonparametric setting, but retain the

Finally, the last column in Table 1 shows that such studies, not surprisingly, giving the large numbers of differing aspects, yielded conflicting results concerning the relationship between the evolution of international trade dynamics over time. The results span from a decrease in specialization in Kim (1995), Kalemli-Ozcan, Sorensen and Yosha (2003) De Benedictis, Gallegati and Tamberi (2006), to a lack of any increase in international specialization Proudman and Redding (2000) and Redding (2002), to a nonmonotonic relationship between diversification and development in Imbs and Wacziarg (2003) and Koren and Tenreyro (2004).[13]

In this paper we do not provide nor a comparative analysis between variables (employment and production vs. trade) or estimation methodologies (parametric vs. nonparametric), nor a comprehensive description of specialization trends. Rather we try to provide further empirical evidence on the evolution of international trade specialization using annual export trade data for a very large number of manufacturing sectors (as a high level of disaggregation let us capture a wide range of industrial specialization dynamics), a broad sample of countries and a 17 years time span. We analyze the evolving pattern of industrial specialization applying semiparametric panel techniques, as they let us estimate the shape of the relationship without making any a priori explicit or implicit assumption about it and take into account country-specific characteristics and, among the different indexes of statistical dispersion, we focus on measures of relative statistical dispersion because relative indexes, in contrast to absolute ones, are sensible to changes in the world structure even when the national distribution remains unchanged.[14] Moreover, as a robustness check, we perform the same exercise using alternative measures of overall specialization, different levels of disaggregation within manufacturing exports and different smoothing parameters of the nonparametric term associated to income per capita.

3   Methodological framework

The use of a particular dataset is not neutral with respect to the research question, thus the choice of sectoral and geographic aggregation level, of time length, and of the variable selected becomes relevant in order to evaluate the

---

additivity assumption.

[13]Only Amiti (1998) and Brülhart (1998), analyzing just a few sample of OECD countries, provides evidence that countries progressively increase their overall degree of specialization.

[14]Anyway, we applied our nonparametric analysis to both measures of specialization, *i.e.* relative vs. absolute indexes, finding some relevant differences. More details about such differences are provided in section 4.

generality of an empirical result. The main characteristic that differentiate our data set from the other ones used in empirical literature on specialization and development are: i) the type of data chosen, trade rather than production data, and ii) the very large number of sectors involved.

If, on the analysis of the relationship between specialization and development, we take the point of view of trade theory, and, thus, concentrate on trade-based explanations of the evolution of trade patterns, we should not consider production and export specialization, nor production and trade data, as equivalent. Moreover, there are some other reasons why trade data may be preferable to production data, as trade data are generally broadly available, more reliable and more finely disaggregated across industries than production data.[15] Thus, exports trade data may be considered a first best indicator of overall specialization dynamics.

As regards the choice between manufacturing and economy-wide sectoral data, we chose to work only with the manufacturing sector mainly so as to avoid biases that can be induced by the influence of strong specialization linked to geographical and geophysical characteristics. We prefer to concentrate our attention on the so called "foot-loose" sectors, because in this case efficiency in exporting is, broadly speaking, due to the same forces leading to economic growth.[16]

3.1 Alternative measures of specialization

Among the different measures of specialization employed in the literature, we focus on three relatives measures of Overall Specialization (OS). These indexes are all indexes of relative statistical dispersion, and they have the advantage to allow a simple passage from the RCA measure to the OS measure, since the second is built on the basis of the distribution of the first. If other traditional indexes of (absolute) statistical dispersion were used, like Gini or Herfindhal, this direct passage would be lost and, moreover, the distribution benchmark would be the equidistributional *loci*, whose interpretation is not, in this context, fully clear.

In the relative indexes both country and world data are relevant for the final result; this means that changes in the world distribution automatically reflect in the OS measures, even if the national distribution has not changed. This is an advantage. Consider the case of an unchanged national distribution. A Gini index obviously shows an unchanged situation, even if, in the meanwhile the world structure has changed in a significant manner, due to,

---

[15]Indeed, the more aggregate the data, the less information we are likely to obtain.

[16]A detailed specification of the data set employed is presented in the Appendix.

for example, changes in technology or demand structure. Instead, relative indexes, like those proposed in this paper, would be sensible to changes of that kind in the world structure, and it seems reasonable if the analysis of specialization is used to understand the position of a country in the world economy. Consequences in terms of economic growth, firm profitability and other economic variables depend a lot on the relative position of a country with respect to technological and demand dynamics at the world level.

Our common starting point for all the OS indexes is the widely used RCA index known as the Balassa Index, BI, from the sectoral distribution of which we derive three indexes of OS, described in the following:

- as a first step, a simple positional index of the distribution of sectoral BI seems a suitable measure of overall specialization. Since BI is an asymmetric index, the median (OS^{me}) more than the mean seems an appropriate positional index. OS^{me} is an inverse index of OS: a high OS^{me} says that there are many sectors with comparative advantages and this means that the country has a low OS (because it efficiently trades in many goods)

- a second index may be derived from previous literature (see Amiti, 1999) ant it has been called "country Gini"; in terms of the Lorenz curve, it is calculated ranking sectors according to their growing BI and measuring national shares (BI numerator) on the y axis and world shares (BI denominator) on the x axis. We prefer to call this the "relative Gini" index (OS^{rg}), since it measure the relative (to the world average) sectoral concentration of the trade structure of a country. With data ordered according to their growing BI, we compute the index according to:

$$OS^{rg} = \sum_{i=1}^{n-1}(p_i - q_i)/\sum_{i=1}^{n-1} p_i \qquad (1)$$

where q_i and p_i are, respectively, cumulated shares of the numerator and denominator of BI (that is: national and world sectoral shares), and i denotes sectors. Its minimum value (minimum OS) is $min(OS^{rg}) = 0$, when a country has the same export shares distribution than the world, that is when $q_i = p_i$ for all i. Its maximum is $max(OS^{rg}) = 1$, when the whole exports of a country are concentrated in only one sector, that is when $q_i = 0$ for $i = 1, ..., n-1$ and $q_n = 1$.[17]

Finally, it is possible to utilize an index, derived from Theil (1967) (OS^{th}), that is an entropic index where the numerator and denominator of BI are proportionally confronted:

---

[17]As a consequence OS^{rg} reduces to $OS^{rg} = \sum_{i=1}^{n-1}(p_i)/\sum_{i=1}^{n-1} p_i = 1$ (if the world structure is not perfectly concentrated in the n sector too.)

$$OS^{th} = \sum_{i=1}^{n}\{(z_i/z)ln[(z_i/z)/(Z_i/Z)]\} \qquad (2)$$

where z_i is country export in sector i, Z_i the equivalent for the world, z and Z are country and world total export.[18]

OS^{th}, as evident, is a weighted sum of the logs of sectoral BI, with weights represented by the country sectoral shares; from this point of view OS^{th} can be interpreted as a barycentre of the BI distribution. It ranges from 0 (minimum OS) when $z_i/z = Z_i/Z$ for all i, to ∞ (maximum OS), when at least one $(Z_i/Z) > (z_i/z) = 0$. In terms of the Theil approach, it can be interpreted as a measure of the "surprise" we would have if we predicted the trade structure of the country on the basis of the average world structure, or, in other terms, it is the information content of the message (when our starting information is the world structure).

Table 2 displays the Pearson's and the rank correlation coefficients between our concentration measures and per-capita income(using our pooled export data). It constitutes an interesting, more complete follow-up of previous themes, as well as an introduction to those in the sections to follow: it is evident that there is cause for reflection, given the high level of correlation between the variables.

Table 2: Correlation with per capita income

2 digit			
	OS^{me}	OS^{rg}	OS^{th}
Simple correlation	0.607	-0.603	-0.696
Rank correlation	0.608	-0.682	-0.711
4 digit			
	OS^{me}	OS^{rg}	OS^{th}
Simple correlation	0.713	-0.636	-0.708
Rank correlation	0.735	-0.688	-0.726

3.2 Empirical methodology

In order to investigate the specialization pattern of different countries two different econometric approaches may be employed: parametric or nonparametric. Parametric approaches by definition impose a structure on the functional

---

[18]In the calculation of OS^{th}, when z_i/z was equal to 0, we have taken into account that $lim_{z\to 0}zln(z) = 0$.

form representing the specialization-development relationship. On the other hand, nonparametric methods avoid imposing any particular functional form to the estimated relationship. Our preference goes to nonparametric methods, as they let us estimate the shape of the relationship without making any a priori explicit or implicit assumption about it.

There are several approaches available to estimate nonparametric regression models, like kernel smoothing regressions, locally weighted polynomial regressions and generalized additive models,[19] and most of these methods assume that the nonlinear functions of the independent variables to be estimated by the procedures are *smooth* continuos functions.

The use of kernel smoothing techniques offers two main alternatives (Bowman and Azzalini, 1997). The first one is to fit a *local linear* regression, implying the alternative least squares problem:

$$\min_{\alpha,\beta} = \sum_{i=1}^{n} [y_i - \alpha - \beta(x_i - x)]^2 \cdot w(x_i - x; h). \tag{3}$$

where y_i and x_i are the ith measurement of the response and explanatory variables, respectively, for $i = 1, ..., n$. The kernel function $w(x_i - x; h)$ is in general a positive symmetric function with a maximum at 0 and it decreases monotonically as the distance between each observation x_i increases with respect to the point of interest, x. It follows that more weight is given to the observation close to x; the most widely used functions are the triangular, gaussian, and tricube functions. The *fixed* smoothing parameter h controls the bandwidth of the kernel function, selecting the number of observations around ypc to be included in the local mean estimation or the local regression.

Replacing the fixed bandwidth of equation 3 with a *variable bandwidth*, as in equation 4 the least square problem becomes:

$$\min_{\alpha,\beta_1,...,\beta_p} = \sum_{i=1}^{i} [y_{ct} - \alpha - \beta_1(x_i - x) - ... - \beta_p(x_i - x)^p]^2 \cdot w(x_i - x; h_i). \tag{4}$$

Equation 4 is the least squares problem of a locally weighted polynomial regression (a.k.a. *loess*), of p-degree, due to Cleveland. The variable bandwidth, h_i, reflects the density of the data through the nearest neighbor distance, $d_k(x)$, which is the distance to the kth nearest neighbor of the covariate value x_i. The *span* of the estimator is the parameter $k/n \in [0, 1]$ that describes the proportion of the sample which contributes a positive weight to

---

[19]See Fox (2000a, 2000b) for a discussion on nonparametric regression methods.

each local polynomial regression. The smoothness of the regression is therefore dependent on the two parameters p and k. Finally, the *loess* estimator also incorporates robustness in the fitting procedure, which may be appealing in cases where specific observations can exert a significant influence on the fit.

Generalized additive regression models (GAM)[20] extend the traditional linear statistical models by flexibly modeling additive linear relationships as a combination of smooth nonparametric functions and parametric forms, where the smooth functions are estimated using nonparametric smoothers like spline or loess functions. hus, we may take into account country-specific characteristics inside a nonparametric framework running semiparametric models, where both parametric and nonparametric components are jointly at work. Such semiparametric models, where both parametric and nonparametric components are jointly at work, are particularly to evaluate the statistical significance of the nonparametric fitted functions estimated using different smoothing parameters.

They re-cast the standard linear regression set-up by modeling the dependent variable y_i as an additive combination of a parametric component α, a nonparametric component $f_j(x_{i,j})$, and an i.i.d. disturbance term ε_i with zero mean and variance σ^2, that is

$$y_i = \alpha + \sum_{j=1}^{k} f_j\left(x_{ij}\right) + \varepsilon_i \tag{5}$$

where the functions $f_j(.)$ are *smooth* regression functions to be estimated from the data, and the estimates of $f_j(x_{ij})$ for every value of x_{ij}, written as $\hat{f}_j(x_{ij})$, are obtained using a fitting alghoritm known as *backfitting*.[21] Such a model allows us to gain more flexibility, as it replaces the linearity assumtpion with some univariate smooth functions in a nonparametric setting, but retain the additivity assumption. Moreover, an important advantage of GAMs with respect to other nonparametric methods is the possibility to evaluate the statistical significance of the smooth nonparametric components.

Two smoothing functions are available to estimate these partial-regression functions $f_j(.)$, that is spline and locally-weighted regression smoothers. Both smoothers have similar fits with the same equivalent number of parameters,

---

[20]GAMs were introduced by Hastie and Tibshirani (1986) and are described in detail in Hastie and Tibshirani (1990).

[21]A full description of how the alghoritm works in GAMs is available in Hastie and Tibshirani (1990).

but the local regression (*loess*) method developed by Cleveland (1993) provides robust fitting when there are outliers in the data, support multiple dependent variables and computes confidence limits for predictions when the error distribution is symmetric, but not necessarily normal. In the *loess* method the regression function is evaluated at each particular value of the independent variable, x_i, using a local neighborhood of each point and the fitted values are connected in a nonparametric regression curve. In fitting such a local regression, a fixed proportion of the data is included in each given local neighborhood, called the *span* of the local regression smoother (or the smoothing parameter), and the data points are weighted by a smooth function whose weights decrease as the distance from the center of the window increases.

4 Empirical evidence on specialization and development

In this section we analyze the shape of the relationship between the overall degree of specialization and per capita income inside a nonparametric framework and test the robustness of our results using different measures of specialization and different levels of disaggregation of the data.

A time-period of 17 years could be interpreted as a too short one in order to draw any conclusions about the relationship between specialization and the level of development. But the countries included in our data set are countries that, given the very different values of their own level of per capita income, belongs to different stages of the economic development, and that differ significantly on many accounts, such as the size, the degree of openess, the quality of the institutions, etc. Thus, once we take into account such country-specific characteristics of our data set through country fixed effects, all pairs of *(specialization, income)* may be considered equivalent and interpreted as the values of an hypothetical country at different stages of development .

We try to evaluate the nature of the relationship between overall specialization and per capita income taking into account country-specific characteristics inside a nonparametric framework, running, as said, a generalized additive model (GAM).

Thus, we estimate the following generalized additive model where both income and country-specific effects contribute to understand the evolution of sectoral concentration along the development path, that is

$$OS_{ct}^x = \alpha_c + g_j(ypc_{ct,j}) + \epsilon_{ct} \qquad (6)$$

where x is the index of specailization, c is the number of countries (c=1,...,C) and t the number of years (t=1,...,T). In equation 6 the parametric component is represented by a set of dummy regressors corresponding to the number of countries, and the nonparametric component is given by a smooth term for per capita income. Among the various general scatterplot smoothers considered in the literature for the $g_j(.)$ function we choose, in conformity with the analysis of the previous section, a locally weighted regression smoother.

It may be useful to summarize our procedure of estimation. It can be synthesized in a two general points : what is in common and what differs among the various estimations.

For the first of them, as already anticipated, we have some common characteristics involving all our estimates: we always use trade data for measuring diversification for whole period covered by the analysis, we always use the same methodological (semi-parametric) framework.

Our estimates differ because of four aspects:

a) we employ three different indicators of Overall Specialization (as already illustrated);

b) we employ data at different level of sectoral disaggregation (2 and 4 SITC digits, that is to say about 30 and 500 manufacturing sectors, respectively);

c) we use three different spans in the non-parametric component (k equal to 0.25, 0.5 and 0.75), so more or less stressing "local" information;

d) we also use two different degree (first and second) for the polynomial of the non-parametric component, thus allowing for linear and non linear local regressions.

Despite these differences, our results are very stable, as we synthesized in the paper title: "countries diversify".

Results are presented in a graphical version only for the estimations using the 4 digit level of disaggregation, to avoid a too heavy presentation; these results are reported in figures 3 and 4, where it is possible to visualize the non parametric component of the estimation, or, in other words, the marginal effects of ypc on the OS measures. Since cross-sectional effects are captured by the country dummies, the marginal $ypc - OS$ relationship, showed in the figures, is to be interpreted as the relationship along the time-path of modern economic growth of a "typical" country.

First, let's observe and comment figures 3 and 4, still remembering that OS^{me} is an inverse index of overall specialization, while OS^{rg} and OS^{th} are direct indexes.

17

Figure 3: Nonparametric fitted functions from fixed effects GAM regression
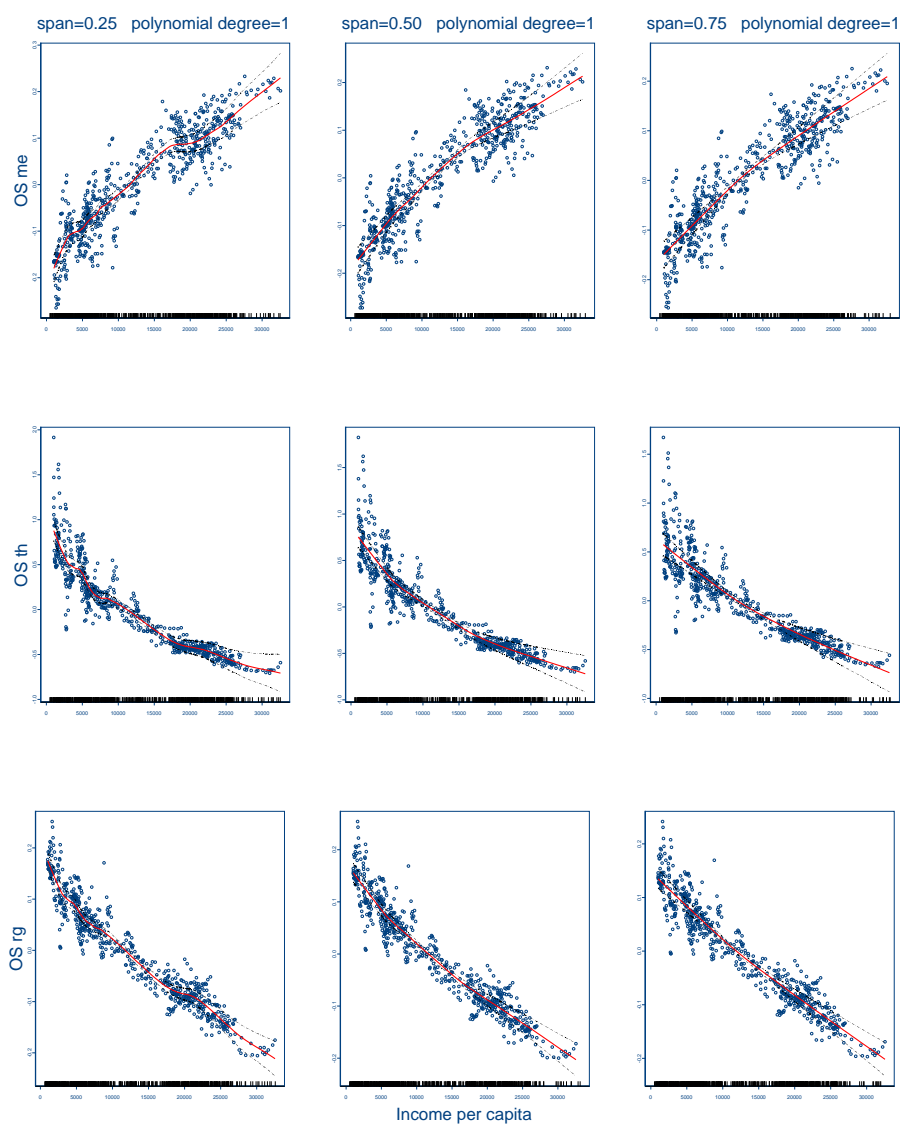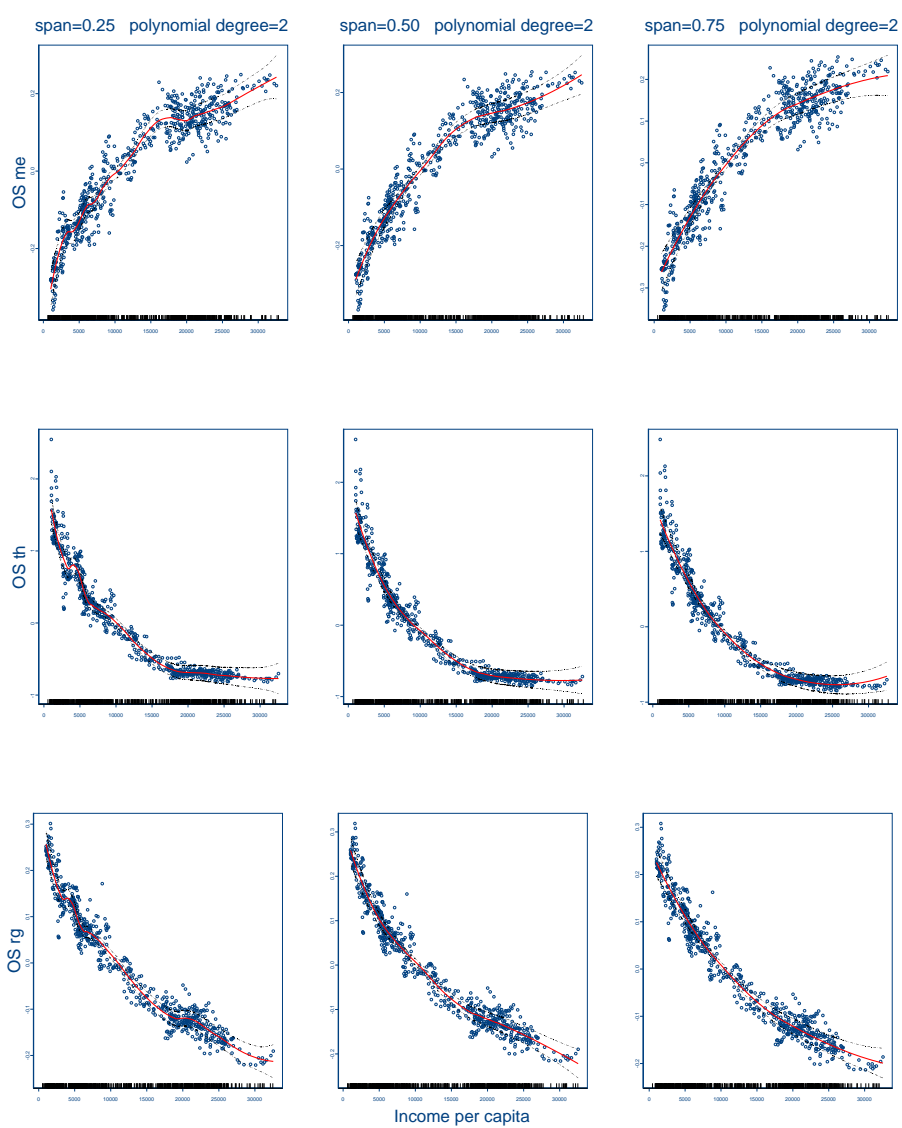- Degree 1 polynomial

Figure 4: Nonparametric fitted functions from fixed effects GAM regression - Degree 2 polynomial

All the estimations (3 indexes, 3 spans, 2 polynomial degree) show a negative, monotonic relationship between ypc and OS. While the effect of ypc on OS^{me} is always positive, it is always negative on the remaining two indexes. Anticipating a point that will be discussed a few lines below, an analogous result of negative, monotonic relationship is obtained at the 2 digit level. This seems to us the strongest result of our analysis. As a consequence, our clear interpretation is that countries always diversify along their path of "modern economic growth".

A second observation is about the changing intensity of the evidenced relationship. This derives from the presence of certain degree of non linearity almost always present in the $ypc-OS$ relationship, since, generally, all curves tend to flatten at higher level of ypc. Consider that our OS indexes are both normalized (0-1 in the case of OS^{rg}) and not (as OS^{th}), so that this tendence to flatten does not depend on the characteristics of the indexes.

A further note on this point is that the nonlinearity just discussed is much more evident if a second degree polynomial is used (see figure 4) in the non parametric component of the estimations, while, with the use of a first degree polynomial, a quasi-linear relation is evidenced in some cases (see figure 3). In particular the $ypc - OS$ linkage seems near to be linear in the case of OS^{rg}, with all spans.

Finally, but marginally, in the case OS^{me}, and with lower span levels, some "disturbances", at the otherwise remarkable smoothness of the curve, can be evidenced, but without significantly changing the general result.

Table 3: F-values for nonparametric effects (degree 1 polynomial)

$OS_{ct} = \alpha + \widehat{lo(ypc_{ct})} + \epsilon_{ct}$						
4 digit						
$\widehat{lo(ypc_{ct})}$	span=0.25		span=0.50		span=0.75	
	Nonpar. F	P(F)	Nonpar. F	P(F)	Nonpar. F	P(F)
OS^{me}	6.9	4.73e-008	12.2	7.63e-007	14.7	4.13e-005
OS^{th}	13.5	2.22e-016	30.4	8.81e-016	28.6	9.48e-009
OS^{rg}	10.0	5.67e-012	18.0	9.92e-010	15.6	2.34e-005
2 digit						
$\widehat{lo(ypc_{ct})}$	span=0.25		span=0.50		span=0.75	
	Nonpar. F	P(F)	Nonpar. F	P(F)	Nonpar. F	P(F)
OS^{me}	6.6	1.41e-007	6.3	0.00	5.5	0.01
OS^{th}	8.4	6.80e-010	17.5	1.80e-009	19.1	2.79e-006
OS^{rg}	8.6	4.78e-010	15.6	1.35e-008	17.3	8.38e-006

Table 4: F-values for nonparametric effects (degree 2 polynomial)

$OS_{ct} = \alpha_c + \widehat{lo(ypc_{ct})}^2 + \epsilon_{ct}$						
4 digit						
$\widehat{lo(ypc_{ct})}$	span=0.25		span=0.50		span=0.75	
	Nonpar. F	*P(F)*	*Nonpar. F*	*P(F)*	*Nonpar. F*	*P(F)*
OS^{me}	6.1	1.44e-009	11.7	9.78e-009	15.1	1.60e-006
OS^{th}	10.0	0.00	13.9	2.72e-010	25.3	3.22e-010
OS^{rg}	8.5	3.11e-014	12.9	1.48e-009	21.2	9.24e-009
2 digit						
$\widehat{lo(ypc_{ct})}$	span=0.25		span=0.50		span=0.75	
	Nonpar. F	*P(F)*	*Nonpar. F*	*P(F)*	*Nonpar. F*	*P(F)*
OS^{me}	6.3	6.16e-010	10.7	1.74e-007	11.8	0.00
OS^{th}	6.1	1.36e-009	5.1	0.00	9.6	0.00
OS^{rg}	7.4	5.87e-012	11.0	3.54e-008	16.0	7.66e-007

As said, an analogue picture is obtainable if the 2 digit level is used; we do not show a figure for this case for brevity, as all our previous results are confirmed.[22] This is also stressed by tables 3 and 4, which provide a complete information on the significance tests on the $ypc - OS$ relationship (non-parametric F-tests), both for 2 and 4 digit levels of disaggregation.

The major evidence of the tables is that these F-tests, for all indexes, all spans, all polynomial degrees and all disaggregation levels, invariably show that the $ypc - OS$ relationship, with these data set and in this time span, is significant. This is also another reason to be confident about the last two words of the title of this paper.[23]

There is also a general tendence for the F-tests to be higher with higher levels of the span. This is not always true, but most of the estimations follow this pattern of significativity. This may be interpreted, even if with more than a word of caution, in the sense that the $ypc - OS$ relationship is a deep-rooted one, depending on the basic forces of modern economic growth,

---

[22]On the contrary, our empirical evidence displayed by absolute indexes does not provide a uniform pattern of the specialization-income relationship. Anyway, in comparison to relative indexes, absolute indexes tend to display a less pronounced pattern of diversification (the results of nonparametric analysis with absolute indexes are available from the authors upon request).

[23]Albeit always significant, we can see that F-tests are usually, but not always, higher when the OS^{th} and the OS^{rg} indexes are used, a little bit lower for OS^{me}; they are often higher when a 4 digit sectoral disaggregation is used. The only exceptions to this "rule" being OS^{me} (at 2 digit, first degree polynomial), and, partially, OS^{th} (at 2 digit, first degree polynomial).

then better captured when many data are used for the local estimations.

Finally, it is possible to observe that there are neither evident nor systematic differences in F-tests values if first or second polynomial degrees are used: we are not in the position to prefer a more or less pronounced non-linearity in the $ypc - OS$ relationship.

5   Conclusions

International trade theories do not share a common view about the way trade specialization evolves with per capita income along the development path. This subject has been analyzed, more or less explicitly, in different strands of the recent empirical literature on specialization. Nevertheless, as we showed in the first part of this work, no clear conclusion seems to emerge, since the results are generally different and not directly comparable, due to the different context in which they are developed from the point of view of the kind of data utilized, the time span, the geographic extension covered and the methodology used.

Our objective has been to do a step in the direction of a better understanding of the relationship between development and diversification, trying to make a systematic analysis, also if in a limited context. Using trade data, in a context of semi-parametric analysis, we perform different regressions using a panel data set of seventeen years and a large set of countries at different stages of development, comparing different indexes of overall specialization (all synthesized from the whole distribution of a measure of comparative advantages), and controlling both for countries specificities and for different non-parametric coefficients.

On the whole, results are quite clear and robust, in the sense that they point out that countries continuously diversify along their path of economic development, even if this process seems to be more intense in the first phases of development than in the late ones.

Appendix

Our data set is based on trade data and consists of a balanced panel stemming from two different sources: exports come from ECLAC-UN CAN2003 (ECLAC, 2003), and per capita income data are from the World Development Indicators (World Bank, 2005). Specifically, our data set consists of:

- export data based on the SITC rev.2 classification at the 2 and 4-digit level (about 30 and 500 manufacturing sectors, respectively);[24]

- annual observations over the 1985-2001 period;

- 39 countries selected on the basis of total GNP (> 100 billions as in WB WDR data set);[25]

- per capita income (ypc) is measured in PPP constant 2000 international dollars.

Table 5: List of countries, ordered according to average percapita income

INDIA	PAKISTAN	BANGLADESH	CHINA
PHILIPPINES	EGYPT	INDONESIA	ALGERIA
SOUTH AFRICA	COLOMBIA	TURKEY	THAILAND
BRAZIL	POLAND	CHILE	MALAYSIA
ARGENTINA	MEXICO	VENEZUELA	GREECE
PORTUGAL	KOREA	SPAIN	ISRAEL
ITALY	AUSTRIA	FINLAND	BELGIUM
UK	NETHERLAND	FRANCE	SWEDEN
JAPAN	DENMARK	AUSTRALIA	SWITZERLAND
NORWAY	CANADA	USA	

The 39 countries ordered according to average per capita income are listed in table 5, while table 6 presents the summary statistics of the variables used in nonparametric analysis for the whole period, the first year and last year of the sample, respectively.

---

[24]Manufacturing is defined as the sum of sectors from code 5 to 9, included. The total amount of sectors included in the database is 786 total sectors.

[25]The choice of total income as a basis for the selection of countries was made to avoid possible distortions due to the presence of too small economies; nevertheless relative per capita incomes range from 3 (Bangladesh) to 100 (USA) in 2001 as it is shown in table 6.

Table 6: Summary statistics

	Minimum	Maximum	Mean	Std. dev.
Full sample	2-digit sectors			
OS^{me}	0.023	1.065	0.583	0.252
OS^{rg}	0.064	2.729	0.725	0.614
OS^{th}	0.177	0.968	0.553	0.198
Full sample	4-digit sectors			
OS^{me}	0.002	0.950	0.365	0.256
OS^{rg}	0.190	4.426	1.232	0.859
OS^{th}	0.311	0.991	0.682	0.170
Full sample	Income			
ypc	1022	32554	12632	0.671
y	49639	9013924	672846	1.854
1985	2-digit sectors			
OS^{me}	0.062	1.065	0.545	0.282
OS^{rg}	0.077	2.389	0.775	0.573
OS^{th}	0.185	0.921	0.581	0.201
1985	4-digit sectors			
OS^{me}	0.063	1.050	0.644	0.262
OS^{rg}	0.093	2.577	0.656	0.603
OS^{th}	0.217	0.966	0.532	0.200
1985	Income			
ypc	1022	23919	10668	0.693
y	49639	5563406	490169	1.903
2001	2-digit sectors			
OS^{me}	0.002	0.923	0.321	0.277
OS^{rg}	0.220	4.426	1.443	0.976
OS^{th}	0.316	0.970	0.721	0.176
2001	4-digit sectors			
OS^{me}	0.003	0.947	0.398	0.259
OS^{rg}	0.210	3.395	1.114	0.812
OS^{th}	0.355	0.984	0.660	0.173
2001	Income			
ypc	1479	32554	14667	0.665
y	115417	9013924	879354	1.847

Note: For per capita income, ypc, and absolute income levels, y, the coefficient of variation (and not the standard deviation) is reported.

6 References

Acemoglu D. and Zilibotti F. (1997), "Was Prometheus Unbound by Chance? Risk, Diversification, and Growth", *Journal of Political Economy*, August, 105(4), 709–51.

Amiti M. (1999), "Specialisation Patterns in Europe", *Weltwirtscahftliches Archiv*, 134(4), 573-593.

Balassa B. (1965), "Trade Liberalization and Revealed Comparative Advantage", *Manchester School of Economics and Social Studies*, 33, 99-123.

Bowman A. W. and A. Azzalini (1997), *Applied Smoothing Techniques for Data Analysis*, Clarendon Press, Oxford.

Brasili A., Epifani P. and R. Helg (2000), "On the Dynamics of Trade Patterns", *De Economist*, 148 (2), 233-258.

Broda C. and D. Weinstein 2004, "Variety Growth and World Welfare", *American Economic Review Papers and Proceedings*, 94 (2), 139-144.

Brülhart M. (1998), "Trading Places: Industrial Specialization in the European Unio", *Journal of Common Market Studies*, 36(3), 319-346.

Brülhart M. (2001), "Evolving Geographical Specialization of European Manufacturing Industrie", *Weltwirtschaftliches Archiv* 137 (2), 215-243..

Cleveland W. S. (1993), *Visualizing Data*, Hobart Press, Summit, New Jersey.

De Benedictis L. and M. Tamberi (2004), "Overall Specialization Empirics: Techniques and Application", *Open Economies Review*, 15 (4), 323-346.

De Benedictis L., Marco Gallegati and M. Tamberi 2006, '"Semiparametric Analysis of the specialization-income relationship", *Applied Economics Letters*, forthcoming.

Di Nardo J. and J.L: Tobias (2001), "Nonparametric Density and Regression Estimation", *Journal of Economic Perspectives*, 15(4), 11-28.

ECLAC (2000), *CAN2000*, Economic Commission for Latin America and the Caribbean.

Evenett S.J. and A.J. Venables 2002, "Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows", *mimeo*.

Fan J. and I. Gijbels (1996), *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Feenstra R. and A. Rose (2000), "Putting Things in Order: Trade Dynamics and Product Cycles", *Review of Economics and Statistics* 82, 369-382.

Fox J., 2000a, *Nonparametric Simple Regression: Smoothing Scatterplots*, Sage, Thousands Oaks CA.

Fox J., 2000b, *Multiple and Generalized Nonparametric Regression*, Sage, Thousands Oaks CA.

Grossman G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, MIT Press.

Hastie T. and R. Tibshirani (1986), "Generalized Additive Models", *Statistical Science*, 1, 297-318.

Hastie T. and R. Tibshirani (1990), *Generalized Additive Models*, London, Chapman and Hall.

Hausmann R., Hwang J. and D. Rodrik (2005), "What You Export Matters", *NBER Working Paper no. 11905*, .

Helpman E., M. Melitz and Y. Rubinstein 2006, "Trading partners and Trading Volume", *mimeo.*

Imbs J. and Wacziarg R. (2003), "Stages of Diversification", *American Economic Review*, 93(1), 63-86.

Kalemli-Ozcan S., Sørensen B. and O. Yosha (2003), "Risk sharing and industrial specialization: Regional and international evidence", *American Economic Review*, June, 93 (3), 903–918.

Koren M. and S. Tenreyro (2004), "Diversification and Development", *FRB Boston Series*, paper no. 03-3, Revised 2004.

Kim S. (1995), "Expansion of Markets and the Geographic Distribution of Economic Activities: the Trends in U.S. Regional Manufacturing Strucuture, 1860-1987 ",*The Quarterly Journal of Economics*, 110(4), 881-908.

Krugman P. 1987, "The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher: Notes on the Presence of Dynamic Scale Economies", *Journal of Development Economics*, 27, 41-55.

Peretto P.F. 2003, "Endogenous Market Structure and the Growth and Welfare Effects of Economic Integration", *Journal of International Economics*, 60 (1), 177-201.

Proudman J. and S. Redding, 2000, "Evolving patterns of International trade", *Review of International Economics*, 8(3), 373-396.

Redding S., 2002, "Specialization Dynamics", *Journal of International Economics*, 58(2), 299-334.

Ruhashyankiko J.F. (2000), "How is the World Specializing?", *mimeo.*

Stockey N., 1988, "Learning by Doing and the Introduction of New Goods", *Journal of Political Economy*, 96(4), 701-717.

World Bank (2005) *Global Development Network Growth Database*, World Bank, http://www.worldbank.org/research/growth/GDNdata.htm.