ORIGINAL PAPER

Overall trade specialization and economic development: countries diversify

**Luca De Benedictis · Marco Gallegati ·
Massimo Tamberi**

Abstract This paper provides evidence for an aspect of trade often disregarded in international trade research: countries' sectoral export diversification. The results of our semiparametric empirical analysis show that, on average, countries do not specialize; on the contrary, they diversify. Our results are robust for different statistical indices used to measure trade specialization, for the level of sectoral aggregation, and for the level of smoothing in the nonparametric term associated with per capita income. Using a generalized additive model (GAM) with country-specific fixed effects it can be shown that, controlling for countries' heterogeneity, sectoral export diversification increases with income.

Keywords International trade · Specialization · Development ·
Generalized additive models

JEL Classification C14 · E32 · F10

L. De Benedictis
Istituto di Istituzioni Economiche e Finanziarie,
Università di Macerata, Macerata, Italy
e-mail: debene@unimc.it

M. Gallegati · M. Tamberi (✉)
Dipartimento di Economia, Facoltà di Economia "G. Fuà",
Università Politecnica delle Marche, Piazzale Martelli 8, 60121 Ancona, Italy
e-mail: m.tamberi@univpm.it

M. Gallegati
e-mail: marco.gallegati@univpm.it

1 Introduction

In recent years, an increasing number of theoretical and empirical studies have focussed on the relationship between economic diversification[1] and development across countries, but neither theoretical models nor empirical evidence has come to a clear conclusion on this subject.

Trade theories disagree about the way trade specialization evolves with per capita income; indeed, along with models suggesting that countries' specialization should increase over the development path (Krugman 1987), there are models indicating that countries' economic development is associated with a low degree of specialization (Stockey 1988), and more general frameworks (Peretto 2003) showing that both cases could happen, creating international market integration. Nor do empirical works help to clarify alternative trade-based explanations of the link between specialization and economic development. In fact, the available evidence on this issue provides different answers depending on the data set, on the measure of specialization, and on the methodological approach employed.

The empirical literature on specialization patterns analyses changes in countries' overall degree of specialization using both absolute and relative measures of specialization (e.g. Gini, Herfindahl, and Balassa indices). Such measures are inherently static, therefore, in order to analyse issues related to changes in countries' overall degree of specialization, the evolution of these distribution indices over time has to be analysed. As data sources rarely go back to the pre-1970s period, it may be difficult to get robust evidence on long run patterns of specialization given the relatively short time span available. To date, there have not been many proponents of empirical solutions to this question: Kalemli-Ozcan et al. (2003), Imbs and Wacziarg (2003), and Koren and Tenreyro (2007) are the few exceptions. The Imbs and Wacziarg (2003) research is the most relevant to this analysis, as this paper investigates the link between overall specialization and growth analysing the relationship between a meaningful synthetic indicator of overall specialization and the level of per capita income.

In this paper sectoral diversification evolving along the development path is investigated inside a semiparametric framework using detailed manufacturing trade export data (2- and 4-digit SITC) over the period 1985–2001. The preference for nonparametric techniques in recent empirical papers reflects the fact that nonparametric methods allow empirical researchers to explore the issues related to the shape and the statistical significance of the relationship between specialization and income without making any explicit or implicit assumptions about that relationship.

With a sample covering a time span of 17 years, the possibility of drawing any long-run conclusions about the relationship between specialization and the level of development relies on considering all per capita income–specialization pairs as the values of a hypothetical country at different stages of development. A crucial point in analysing the relationship between specialization patterns and economic

---

[1] In this paper the word diversification, in accordance with most of the empirical literature on the specialization–development relationship, is used just as an antonymous of specialization rather than in association with the concept of risk.

development is the role of country-specific effects, as countries may differ significantly in aspects such as size, geography, and institutional and political features; as a consequence, a fixed effects estimator in the context of generalized additive models (GAM) is used to account for these differences.

Moreover, to check robustness, the sensitivity of the main result is tested against (i) alternative measures of export specialization, (ii) different levels of disaggregation within manufacturing exports, and (iii) different smoothing parameters of the nonparametric term associated with per capita income. After controlling for cross-country heterogeneity, there is robust evidence that countries monotonically diversify with the level of development.

The paper is organized as follows. Section 2 reviews the existing empirical literature; Sect. 3 presents the data set, the various measures of specialization, and the empirical methodology employed; Sect. 4 provides the empirical evidence on specialization dynamics and, finally, Sect. 5 concludes the paper.

2 Empirical literature on specialization dynamics

What is the evolution of trade patterns over time? How does sectoral diversification evolve with income growth? As an example, observe the export structure of two countries, Italy and Indonesia, over a certain period of time.

In the top panels of Fig. 1, the world sectoral market shares of the two countries in 1985 is plotted, with each single bar in the two panels identifying the total value of the country's sectoral exports relative to the value of world exports in that sector. Of approximately 770 sectors (SITC rev. 2 at the 4-digit level), Italy's highest sectoral market share in 1985 is in "Fabrics, woven, of sheep's or lambs' wool or of fine animal hair" with a sectoral market share of 61%, while Indonesia's highest sectoral market share in the same year is in "Vegetable materials of a kind used primarily for plaiting" with a sectoral market share of 44%. The horizontal line identifies the total world export share for the country: in 1985 it was about 0.04 for Italy and about 0.01 for Indonesia. In the bottom panels of Fig. 1, the value of the sectoral Balassa Index (BI) is obtained by dividing each sectoral share by the total export share of the country.

In this case 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).[2]

Three facts are worth noticing: first, the two countries export in sectors with and without RCA; second, Italy's market shares are generally higher than Indonesia's; and third, Italy exports in a larger set of sectors than Indonesia does.

The Italian export structure is characterized by a higher level of export diversification, while the Indonesian export structure is characterized by a higher level of export specialization. From a broad perspective (De Benedictis and Tamberi 2004) Indonesia's exports are more specialized than Italy's just because it exports

---

[2] The two sectors previously quoted take the values 15.25 for the Italy's top sector, and 44.00 for Indonesia's.
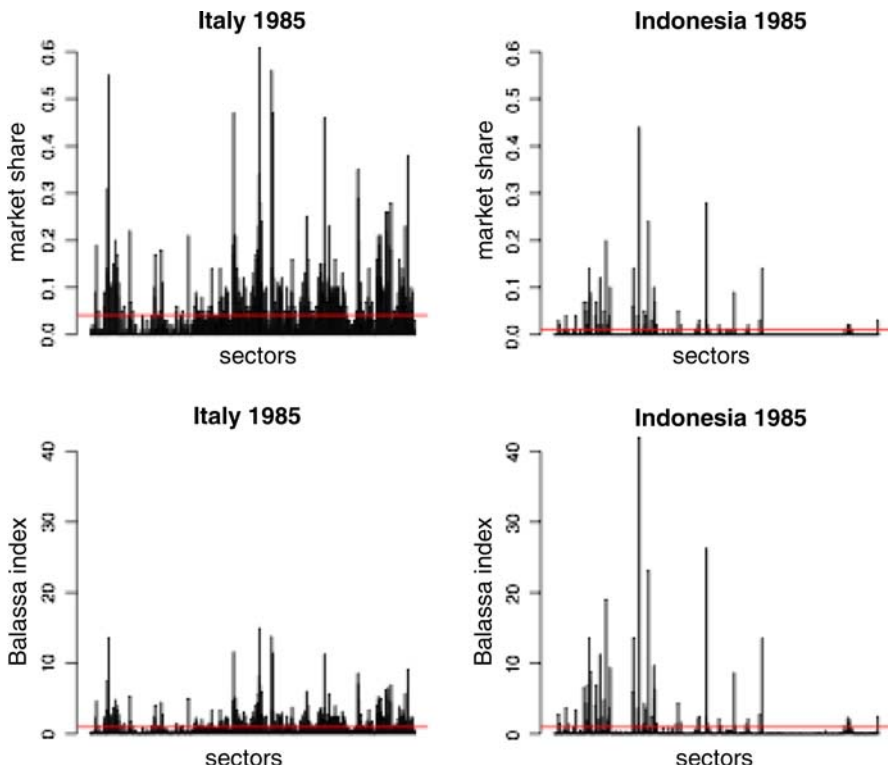
Fig. 1 Sectoral specialization in Italia and Indonesia in 1985

from a more limited number of RCA sectors than Italy does. A simplistic induction would let to the proposition that countries diversify their export structure as they reach a higher level of per capita income.

The evolution of sectoral specialization patterns along the development path has been analysed in many empirical papers using, for the most part, trade-based and production-based measures of specialization, like RCA or the location quotient, respectively.[3] A synopsis of the recent empirical literature is presented in Table 1.

The column labels of Table 1 show that studies of specialization patterns differ according to: (i) the time span; (ii) the kind of variables, trade or production, used to construct the measure of specialization; (iii) the set of sectors, economy-wide or manufacturing, and the level of disaggregation, from 10 to 539; (iv) the set of countries considered in the analysis;[4] (v) trade and/or

---

[3] A complementary strand of this literature points out that sectoral specialization patterns, rather than overall specialization itself, matter (Weinhold and Rauch 1999; Bensidoun et al. 2001; and Hausmann et al. 2005).

[4] This set is generally, but not always, only a limited set of countries (usually a subset of developed countries). Only a small number of papers cover a broad sample of developed and developing countries. See, for example, Imbs and Wacziarg (2003), Koren and Tenreyro (2007), and De Benedictis et al. (2008).

Table 1 Recent empirical literature on specialization and development

Author and year	Time period	Variable	No. of manuf. 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 (2007)	1963–1998	P	28	42 Mix	Herfindahl (weighted)	Par, Nonpar	U-shaped
De Benedictis et al. (2008)	1985–2001	T	539	40 Mix	BI median	Nonpar	Decrease

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 and not manufacturing sectors

production (employment) data;[5] and, finally, (vi) the index of specialization employed.

As a consequence, these studies yield conflicting results concerning the evolution of specialization patterns over time which range from a decrease in specialization (Kim 1995; Kalemli-Ozcan et al. 2003; De Benedictis et al. 2008), to a lack of any increase in international specialization (Proudman and Redding 2000; Redding 2002), to a nonmonotonic relationship between diversification and development (Imbs and Wacziarg 2003; Koren and Tenreyro 2007).[6]

Finally, both parametric and nonparametric techniques have been used in empirical studies; recent empirical studies give preference to nonparametric methods as they allow empirical researchers to explore this issue without making any a priori assumptions about the relationship.[7]

This paper provides further empirical evidence on the evolution of international trade specialization using annual export trade data for a large number of manufacturing sectors (as a high level of disaggregation captures a wide range of industrial specialization dynamics), a broad sample of countries, and a 17 year time span. The evolving pattern of industrial specialization is analysed applying semiparametric panel techniques that estimate the shape of the relationship without making any a priori explicit or implicit assumptions about it and that take into account country-specific characteristics. In addition, measures of relative statistical dispersion are focused on because relative indices, in contrast to absolute indices, reflect changes in the world structure even when the national distribution remains unchanged.[8]

As a test of robustness, the same exercise is performed using alternative measures of overall specialization, different levels of disaggregation within manufacturing exports, and different smoothing parameters for the nonparametric term associated with per capita income.

A major limitation of this approach is that that it does not provide a comparative analysis between variables (employment and production vs. trade) or estimation methodologies (parametric vs. nonparametric); such features, here unexplored, represent opportunities for future research.

---

[5] Indeed, if there are evident reasons to use trade data for analysing intra-industry trade and production (employment) data for analysing location patterns, even less consensus seems to characterize the empirical analysis of specialization patterns.

[6] Only Amiti (1999) and Brülhart (1998), analysing just a small sample of OECD countries, provide evidence that countries progressively increase their overall degree of specialization.

[7] In particular, while the former strand of empirical research has employed a parametric approach (Kim 1995; Amiti 1999; Proudman and Redding 2000; Brasili et al. 2000; Redding 2002; Kalemli-Ozcan et al. 2003; Koren and Tenreyro 2007), most recent empirical studies in this area have adopted nonparametric methods (Imbs and Wacziarg 2003; Koren and Tenreyro 2007; and De Benedictis et al. 2008).

[8] Anyway, the nonparametric analysis was applied to both measures of specialization, i.e. relative and absolute indices, finding some relevant differences. More details about such differences are provided in Sect. 4.

3 Methodological framework and data set

The choice of a particular data set is critical to the research question. There are two main characteristics that differentiate this data set from the ones used in the empirical literature on specialization and development: (i) the type of data chosen, that is, trade instead of production data, and (ii) the very large number of sectors involved.

Taking the view of trade theory, and thus concentrating on trade-based explanations of the evolution of trade patterns, production and export specialization should not be considered as equivalent, nor should production and trade data. Moreover, there are other reasons why trade data may be preferable to production data, as trade data are generally more broadly available, more reliable, and more finely disaggregated across industries than production data.[9] Thus, export trade data may be considered the best indicator of overall specialization dynamics.

This study works only with manufacturing, instead of economy-wide sectoral data, so as to avoid the bias that can be introduced by geographical and geophysical characteristics. Indeed, strong advantages in sectors linked to these (exogenous) characteristics would result in an algebraic distortion of the BI for other sectors (as in the case of oil exporting countries).

As a consequence, this research focuses on the so-called "footloose" sectors, because in this case efficiency in exporting is, broadly speaking, due to the same forces that lead to economic development.[10] In particular, this data set consists of export data for 539 manufacturing sector (on the SITC rev. 2 classification at the 2- and 4-digit level) over the period 1985–2001 for 39 countries whose per capita income is measured in purchasing power parity (PPP) constant 2,000 international dollars. A detailed explanation of the data set is presented in the Appendix.

3.1 Alternative measures of overall specialization

Both absolute and relative measures of specialization have been employed in the empirical literature. Traditional indices of (absolute) statistical dispersion, like the Gini or Herfindhal, compare the domestic distribution against a hypothetical and unrealistic equi-distribution,[11] while indices of relative statistical dispersion have the ability to measure the relative efficiency of sectors, through the direct comparison of country and world shares in the sectors being considered. In relative indices both country and world data are relevant, hence changes in the world distribution are automatically reflected in Overall Specialization (OS) measures, even though the national distribution is unchanged. In terms of economic growth, firm profitability, and other economic variables, it is a country's position relative to technological and demand dynamics at the world level that matters. Thus, it seems reasonable to analyse specialization using indices that, being sensible to changes in

---

[9] As a matter of fact, the more aggregated the data, the less information is likely to be obtained.

[10] Nevertheless, the links between primary and manufacturing sectors are not exclusively limited to algebraic questions (as emphasized by the literature on the Dutch Disease).

[11] Since this study uses trade data, there is no particular reason to assume that the equi-distribution should be considered as the benchmark.

the world structure, are likely to interpret the position of a country in the world economy.

A very common index of sectoral specialization is the BI of RCA (Balassa 1965), that can be expressed as:

$$BI_{[cst|w]} = \frac{X_{cs}}{X_{ws}} \Big/ \frac{X_c}{X_w},$$ (1)

where X denotes exports, c is a specific country, w is the world economy, s is a specific sector, and t is the time period considered. The $BI_{[cst|w]}$, as a measure of sectoral specialization relative to the world exports, reveals that country c has a comparative disadvantage in sector s if $0 < BI < 1$, and a comparative advantage in sector s if $1 < BI < X_w/X_c$. Since the Index has a fixed lower bound and a variable upper bound—with a fixed demarcation value of 1—the BI follows an asymmetric distribution.

Among the different measures of OS employed in the literature, three relatives measures are the focus here, all derived from the sectoral distribution of the BI: OS^{me}, OS^{rg}, and OS^{th}.[12]

As the BI measures, in some way, the relative efficiency of sectors in an economy, a simple positional index of the distribution of the sectoral BI may be a suitable measure of overall specialization; since the BI is an asymmetric index, the median (OS^{me}), rather than the mean, is an appropriate positional index of overall specialization. OS^{me} is an inverse index of OS: a high OS^{me} says that there are many sectors with comparative advantages, meaning that the country has a low OS (because it trades efficiently in many goods).

A second index, known as "country Gini", may be derived from previous literature (Amiti 1999); in terms of the Lorenz curve, it is calculated by ranking sectors according to their growing BI and measuring national shares (the BI numerator) on the y axis and world shares (the BI denominator) on the x axis. We prefer to call it "relative Gini" index (OS^{rg}), as it measures the relative (to the world average) sectoral concentration of the country's trade structure.

With data ordered by increasing BI results, the index is computed as:

$$OS^{rg} = \sum_{i=1}^{n-1} (p_i - q_i) \Big/ \sum_{i=1}^{n-1} p_i,$$ (2)

where q_i and p_i are, respectively, cumulate shares of the numerator and denominator of the BI (that is: national and world sectoral shares), and i denotes sectors. Its minimum value is $min (OS^{rg}) = 0$, when a country has the same export share distribution as the world, that is, when $q_i = p_i$ for all i. Its maximum value is $max (OS^{rg}) = 1$, when the total exports of a country are concentrated in only one sector, that is, when $q_i = 0$ for $i = 1,...,n-1$, and $q_n = 1$.[13]

---

[12] As evidenced in Koren and Tenreyro (2007) such indices are always sensitive to classification; nonetheless, we believe, that our definition of sectors can be useful to look at the time series evidence about countries' evolving specialization.

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

Table 2 Correlation with per capita income

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

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

$$OS^{th} = \sum_{i=1}^{n} \{z_i/z \ln[(z_i/z)/(Z_i/Z)]\}, \qquad (3)$$

where z_i is country exports in sector i, Z_i is world exports in sector i, and z and Z are country and world total exports.[14] OS^{th} is a weighted sum of the logs of the sectoral BI, with weights represented by the country sectoral shares, and, from this point of view, can be interpreted as a barycenter 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" incurred if the country's trade structure were predicted on the basis of the average world structure, or, in other words, it is the information content of the message (when the starting point is the world structure).

Table 2 displays the Pearson's and the rank correlation coefficients between these concentration measures and per-capita income (using the pooled export data). All overall specialization measures display a high level of correlation with per capita income at both 2- and 4-digit levels, with OS^{th} being preferred at the 2-digit level and OS^{me} at the 4-digit level.

3.2 Empirical methodology

In investigating the specialization pattern of countries, two different econometric approaches may be employed: parametric or nonparametric. Parametric approaches, by definition, impose a structure on the functional form representing the specialization–development relationship, while by contrast, nonparametric methods avoid imposing any particular functional form on the estimated relationship. Nonparametric methods are therefore preferred, as they allow for estimating the shape of the relationship without making any a priori explicit or implicit assumptions about it.[15]

There are several approaches available to estimate nonparametric regression models (kernel smoothing regressions, locally weighted polynomial regressions and generalized additive models) and most of these methods assume that the nonlinear

---

[14] In the calculation of OS^{th}, when z_i/z was equal to 0, $lim_{z \to 0} z \ln(z) = 0$.

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

functions of the independent variables to be estimated are smooth continuous functions.[16]

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:

$$\min_{\alpha,\beta} = \sum_{i=1}^{n} [y_i - \alpha - \beta(x_i - x)]^2 \cdot \omega(x_i - x; h), \tag{4}$$

where y_i and x_i are the ith measurement of the response and explanatory variables, respectively, for $i = 1,\ldots,n$. The kernel function $w(x_i - x; h)$ is a positive symmetric function (with a maximum at 0) that decreases monotonically as the distance between each observation x_i increases with respect to the point of interest, x.[17] The fixed smoothing parameter h controls the bandwidth of the kernel function, selecting the number of observations around x_i to be included in the local mean estimation or in the local regression.[18]

Replacing the fixed bandwidth h in Eq. 4 with a variable bandwidth h_i, the least squares problem becomes:

$$\min_{\alpha,\beta_1,\ldots,\beta_p} = \sum_{i=1}^{n} \left[y_i - \alpha - \beta_1(x_i - x) - \ldots - \beta_p(x_i - x)^p\right]^2 \cdot \omega(x_i - x; h_i). \tag{5}$$

Equation 5 is the least squares problem of a locally weighted polynomial regression (loess) of p-degree (Cleveland [1993]). 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, where the span of the estimator is the parameter $k/n \in [0,1]$ describing the proportion of the sample that contributes a positive weight to each local polynomial regression. As a consequence, 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)[19] 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

---

[16] The nonparametric methodology employed in recent emprical studies (Imbs and Wacziarg [2003]; Koren and Tenreyro [2007]) is a locally weighted scatterplot smoothing procedure called loess (Cleveland [1979]). This procedure allows for determining a smoothed, fitted nonparametric curve to represent the relationship linking sectoral concentration and income. A different nonparametric procedure, the generalized additive model (GAM), is employed in De Benedictis et al. ([2008]). 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 retains the additivity assumption.

[17] The most widely used functions are the triangular, gaussian, and tricube functions.

[18] If the data are unevenly distributed or some outliers are present, it could be convenient to use a larger h where the data are sparser and a smaller h when the data are denser.

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

functions. Thus, country-specific characteristics may be accounted for by running semiparametric models, where both parametric (country dummies) and nonparametric (the relationship between overall specialization and the level of development) components are involved.

GAM models 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 $f_j(x_{ij})$, and an i.i.d. disturbance term ε_i with zero mean and variance σ^2, that is

$$y_i = \alpha + \sum_{j=1}^{k} f_j(x_{ij}) + \varepsilon_i \qquad (6)$$

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.[20] Such a model allows more flexibility replacing the linearity assumption with some univariate smooth functions in a nonparametric setting, while retaining the additivity assumption. Moreover, an important advantage of GAMs with respect to other nonparametric methods is the possibility of evaluating the statistical significance of the smooth nonparametric components.

Two smoothing functions are available to estimate these partial-regression functions $f_j(.)$: spline and locally weighted regression smoothers. Both smoothers have similar fits with the same equivalent number of parameters, but the local regression (loess) method developed by Cleveland (1993) provides robust fitting when there are outliers in the data, supports multiple dependent variables, and computes confidence limits for predictions when the error distribution is symmetric, but not necessarily normal.

4 Empirical evidence on specialization and development

In this section, the shape of the relationship between the overall degree of specialization and per capita income inside a nonparametric framework is analysed and the robustness of the results is tested using three different measures of specialization and two different levels of disaggregation of the data.

A period of 17 years could be interpreted as too short to draw any conclusions about the relationship between specialization and the level of development. However, the countries included in this data set are countries that, given the very different values of their per capita incomes, are in different stages of economic development, and also differ significantly in many respects, such as size, degree of openness, quality of institutions, etc. Thus, once such country-specific characteristics have been taken into account through country-specific fixed effects, all pairs of (specialization, income) may be considered equivalent and interpreted as the values of a hypothetical country at different stages of development.

---

[20] A full description of how the algorithm works in GAMs is available in Hastie and Tibshirani (1990).
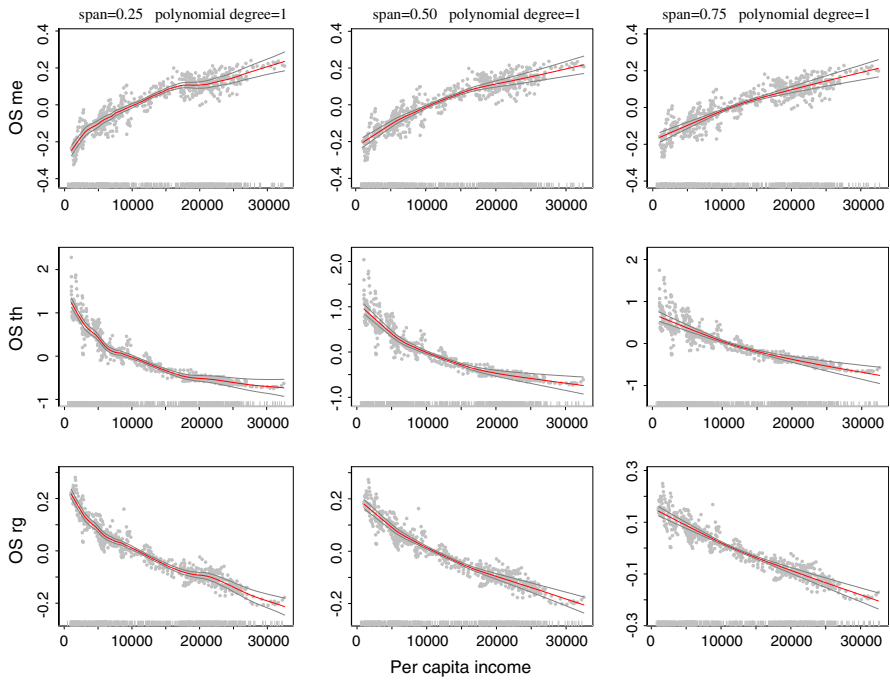
Fig. 2 Nonparametric fitted functions from fixed effects GAM regression (4 digit level—Degree 1 polynomial)

The following generalized additive model with income and country-specific effects is estimated in order to understand the evolution of sectoral concentration along the development path:

$$OS_{ct}^{x} = \alpha_{c} + g_{j}(ypc_{ctj}) + \varepsilon_{ct}, \tag{7}$$

where x is the index of specialization, c is the number of countries ($c = 1,…,C$) and t is the number of years ($t = 1,…,T$). In Eq. 7 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 smoothing term for per capita income. Among the various general scatterplot smoothers considered in the literature for the $g_{j}(.)$ function a locally weighted regression smoother was chosen, based on the analysis of the previous section.

In order to analyse the sensitivity of these results to some specific changes along several dimensions, such as indicators, level of aggregation, and span and degree of polynomial, a sequence of GAM regressions is run that, inside a common framework represented by trade data and nonparametric methodology, differ in: (i) the indicator of OS employed (OS^{me}, OS^{rg} and OS^{th}); (ii) level of sectoral disaggregation (2- and 4-SITC digits, or about 30 and 500 manufacturing sectors, respectively); (iii) the span in the nonparametric component (k equal to 0.25, 0.5, and 0.75), or more or less emphasis on "local" information; and (iv) the degree
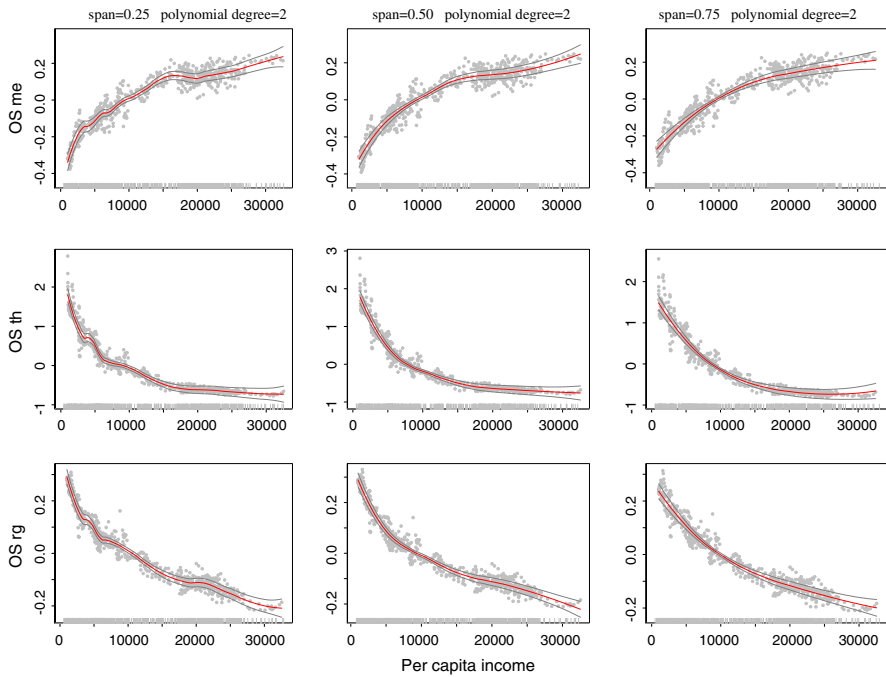
Fig. 3 Nonparametric fitted functions from fixed effects GAM regression (4 digit level—Degree 2 polynomial)

(1 and 2) for the polynomial of the nonparametric component (thus allowing for linear and non linear local regressions).

Figures 2 and 3 display the nonparametric fitted functions from fixed effects GAM regressions representing the marginal effects of *ypc* on the *OS* measures.[21] Indeed, since cross-sectional effects are captured by the country dummies, the marginal *ypc-OS* relationship has to be interpreted as the relationship along the time-path of modern economic growth of a "typical" country.

Despite the differences, two main results emerge from the visual inspection of Figures 2 and 3: first, there is a positive (negative) monotonic relationship between *ypc* and OS^{me} (OS^{rg} and OS^{th})[22] that suggests that "countries diversify" along their development path, and, second, the changing intensity of the evident relationship suggests the existence of a certain degree of nonlinearity in the *ypc-OS* relationship, since, generally, all curves tend to flatten as the level of *ypc* increases.

---

[21] For brevity, nonparametric fitted functions from fixed effects GAM regressions at the 2-digit level of disaggregation are not presented here, given the similarity of the results. On the contrary, the empirical evidence displayed by absolute indices does not provide a uniform pattern of the specialization–income relationship. In particular, in comparison to relative indices, absolute indices tend to display a less pronounced pattern of diversification (the results of nonparametric analysis with absolute indices are available from the authors upon request).

[22] As outlined in Sect. 3, OS^{me} is an inverse index of overall specialization, while OS^{rg} and OS^{th} are direct indices of overall specialization.

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

$OS_{ct} = \alpha_c + lo(ypc_{ct}) + \varepsilon_{ct}$

$lo(ypc_{ct})$	span = 0.25		span = 0.50		span = 0.75	
	Nonpar. F	Prob	Nonpar. F	Prob	Nonpar. F	Prob
4-digit sectors						
OS^{me}	6.9	1.44E−09	11.7	9.78E−09	15.1	1.60E−06
OS^{th}	13.5	0.00E+00	13.9	2.72E−10	25.3	3.22E−10
OS^{rg}	10	3.11E−14	12.9	1.48E−09	21.2	9.24E−09
2-digit sectors						
OS^{me}	6.3	6.16E−10	10.7	1.74E−07	11.8	0.00E+00
OS^{th}	6.1	1.36E−09	5.1	0.00E+00	9.6	0.00E+00
OS^{rg}	7.4	5.87E−12	11	3.54E−08	16	7.66E−07

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

$OS_{ct} = \alpha_c + lo(ypc_{ct})^2 + \varepsilon_{ct}$

$lo(ypc_{ct})$	span = 0.25		span = 0.50		span = 0.75	
	Nonpar. F	Prob	Nonpar. F	Prob	Nonpar. F	Prob
4-digit sectors						
OS^{me}	6.1	1.44E−09	11.7	9.78E−09	15.1	1.60E−06
OS^{th}	10	0.00E+00	13.9	2.72E−10	25.3	3.22E−10
OS^{rg}	8.5	3.11E−14	12.9	1.48E−09	21.2	9.24E−09
2-digit sectors						
OS^{me}	6.3	6.16E−10	10.7	1.74E−07	11.8	0.00E+00
OS^{th}	6.1	1.36E−09	5.1	0.00E+00	9.6	0.00E+00
OS^{rg}	7.4	5.87E−12	11	3.54E−08	16	7.66E−07

With regard to the nonlinearity of the *ypc-OS* relationships, two points merit
further discussion: first, given that the *OS* indices are both normalized (between 0
and 1 for OS^{rg}) and not normalized (OS^{th}), the observed tendency to flatten does not
depend on the characteristics of the indices. Second, the observed nonlinearity is
much more evident if a second degree polynomial is used (see Fig. 3) in the
estimated nonparametric component, while a quasi-linear relationship is evident if a
first degree polynomial is used (see Fig. 2).[23]

The results provided by Figs. 2 and 3 are confirmed by Tables 3 and 4, which
provide complete information on the significance tests on the *ypc-OS* relationship
(nonparametric *F*-tests), for both 2- and 4-digit levels of disaggregation. The major
evidence from the tables is that all *F*-tests, for all different indices, spans,

---

[23] In particular the *ypc-OS* link seems to be nearly linear in the case of OS^{rg}, with all spans.

polynomial degrees, and disaggregation levels, invariably show that the *ypc-OS* relationship, with this data set and in this time span, is significant.[24]

There is also a general tendency for the results of *F*-tests to be higher with the level of the span; this is not always true, even if most of the estimations follow this pattern. Such a result may be interpreted, even with more than a word of caution, to mean that the *ypc-OS* relationship is a deep-rooted one which, depending on the basic forces of modern economic growth, may be 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*-test values when first or second polynomial degrees are used; thus, there is no reason to prefer a more or less pronounced nonlinearity in the *ypc-OS* relationship.

5 Conclusions

The aim of this research is to make a progress towards a clearer understanding of the relationship between economic development and overall specialization. The systematic semiparametric analysis uses trade data relative to a panel data set of a large set of countries at different stages of development over a period of seventeen years.

Comparing different indices of overall specialization derived from the distribution of comparative advantages, and controlling for both countries' specificities and their different nonparametric smoothing parameters, there are clear and robust results. The results point out that countries diversify along their path of economic development, with a trend that seems to be more pronounced in the early phases of economic development than in its later stages.

There are limitations to this empirical analysis: first of all, regarding countries' fixed effects it would be interesting to try to identify these effects, at least partially: likely candidates are variables related to country size, geographical and institutional characteristics. Second, specialization indices are always sensitive to classification. Even if this study has tried to reduce the occurrence of such a problem by using data at different levels of sectoral disaggregation (2- and 4-digits SITC), nevertheless, the problem remains.[25] Third, different results could, in theory, derive from production and/or employment data, but a direct comparison between trade and production data is not easy, since they are based on different systems of classification (although there are recent improvements) and also because they tend to react to (partially) different forces. Finally, as the

---

[24] Albeit always significant, *F*-tests are usually, but not always, higher when the OS^{th} and the OS^{rg} indices 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" are OS^{me} (at 2-digit, first degree polynomial), and, partially, OS^{th} (at 2-digit, first degree polynomial).

[25] Besides, the inclusion or exclusion of agricultural sectors can have its relevance (even if in Imbs and Wacziarg (2003) this seems not to be the case).

process of product diversification along the development path involves one of the main aspects of structural change in an economy, i.e. the creation of new goods,[26] a (non-existing) data set with an expanding (not fixed) number of sectors would be required. This limitation is common to all studies, but is likely to be more severe for those analyses using a long time span and a limited number of sectors.

Nonetheless, this contribution offers suggestions for future empirical research, as well as for future theoretical advances addressing trade and growth. In particular, future contributions could analyse the difference between absolute and relative indices of concentration, and, given the contradictory results obtained in the literature about employment, output, and export specialization trends, could provide a comparative analysis of international specialization patterns using both trade data and production or employment data.

Appendix

Our data set is based on trade data and consists of a balanced panel stemming from two different sources: exports come from CAN2003 (ECLAC-UN 2003), and per capita income 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);[27]
– annual observations over the 1985–2001 period;
– 39 countries selected on the basis of total GNP (>100 billions as in WB WDR data set);[28]
– per capita income (*ypc*) is measured in PPP constant 2,000 international dollars.

The 39 countries included in the sample, ordered according to average per capita income,[29] 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 and last year of the sample, respectively.

---

[26] New goods sometimes replace old ones, but in other cases they simply add to them.

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

[28] 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.

[29] Countries in Table 5 are presented in ascending order as to per capita income starting with the lowest income country at the top left of the first column (Bangladesh) and ending with the highest income country at the bottom of the last column (the United States).

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

Bangladesh	Colombia	Portugal	Australia
Pakistan	Venezuela	Greece	Austria
India	Poland	Spain	Belgium
China	Chile	Israel	Denmark
Indonesia	Brazil	United Kingdom	Japan
Egypt	Malaysia	Finland	Canada
Philippines	Mexico	Italy	Norway
Thailand	South Africa	Sweden	Switzerland
Algeria	Argentina	France	United States
Turkey	Korea	Netherlands	

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	0.95	0.365	0.256
OS^{rg}	0.19	4.426	1.232	0.859
OS^{th}	0.311	0.991	0.682	0.17
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	0.2
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}	0.22	4.426	1.443	0.976
2001	*4-digit sectors*			
OS^{me}	0.003	0.947	0.398	0.259

Table 6 continued

	Minimum	Maximum	Mean	Std. dev.
OS^{rg}	0.21	3.395	1.114	0.812
OS^{th}	0.355	0.984	0.66	0.173
2001	*Income*			
ypc	1479	32554	14667	0.665
y	115417	9013924	879354	1.847

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

References

Amiti, M. (1999). Specialization patterns in Europe. *Weltwirtschaftliches Archiv/Review of World Economics, 135*(4), 573–593.

Balassa, B. (1965). Trade liberalization and revealed comparative advantage. *Manchester School of Economics and Social Studies, 33*(2), 99–123.

Bensidoun, I., Gaulier, G., & Unal-Kesenci, D. (2001). The nature of specialization matters for growth: An empirical investigation. CEPII working paper n. 2001–13. Centre d'Etudes Prospectives et d'Informations Internationales, Paris.

Bowman, A. W., & Azzalini, A. (1997). *Applied smoothing techniques for data analysis*. Oxford: Clarendon Press.

Brasili, A., Epifani, P., & Helg, R. (2000). On the dynamics of trade patterns. *De Economist, 148*(2), 233–258.

Brülhart, M. (1998). Trading places: Industrial specialization in the European Union. *Journal of Common Market Studies, 36*(3), 319–346.

Brülhart, M. (2001). Evolving geographical specialization of European manufacturing industries. *Weltwirtschaftliches Archiv/Review of World Economics, 137*(2), 215–243.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association, 74*(368), 829–836.

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

De Benedictis, L., Gallegati, M., & Tamberi, M. (2008). Semiparametric analysis of the specialization–income relationship. *Applied Economics Letters, 15*(4), 301–306.

De Benedictis, L., & Tamberi, M. (2004). Overall specialization empirics: Techniques and application. *Open Economies Review, 15*(4), 323–346.

ECLAC-UN (Economic Commission for Latin America, the Caribbean–United Nations) (2003). *Competitiveness Analysis of Nations (CAN2003)*. Santiago de Chile: ECLAC-UN. CD-ROM.

Fox, J. (2000a). *Nonparametric simple regression: Smoothing scatterplots*. Thousand Oaks, CA: Sage.

Fox, J. (2000b). *Multiple and generalized nonparametric regression*. Thousand Oaks, CA: Sage.

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science, 1*(3), 297–318.

Hastie, T., & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman and Hall.

Hausmann, R., Hwang, J., & Rodrik, D. (2005). What you export matters. NBER working paper 11905. National Bureau of Economic Research, Cambridge, MA.

Imbs, J., & Wacziarg, R. (2003). Stages of diversification. *American Economic Review, 93*(1), 63–86.

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

Kim, S. (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, 110*(4), 881–908.

Koren, M., & Tenreyro, S. (2007). Volatility and development. *Quarterly Journal of Economics, 122*(1), 243–287.

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*(1–2), 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., & Redding, S. (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.

Stockey, N. (1988). Learning by doing and the introduction of new goods. *Journal of Political Economy, 96*(4), 701–717.

Theil, H. (1967). *Economics and information theory*. Amsterdam: North-Holland Publishing Company.

Weinhold, D., & Rauch, J. E. (1999). Trade, specialization, and productivity growth. *Canadian Journal of Economics, 32*(4), 1009–1027.

World Bank. (2005). *Global development network growth database*. Washington, DC: World Bank. CD-ROM.