Graduate Institute of International Studies

From the SelectedWorks of Luca De Benedictis

Spring March, 2008

Semiparametric analysis of the specialization-income relationship

Luca De Benedictis

Available at: https://works.bepress.com/luca_de_benedictis/8/

Submitted Manuscript

For Peer Review

Semiparametric analysis of the specialization-income relationship

Journal:	*Applied Economics*
Manuscript ID:	AEL-05-0287
Journal Selection:	Applied Economics Letters
Date Submitted by the Author:	30-May-2005
JEL Code:	C14 - Semiparametric and Nonparametric Methods < C1 - Econometric and Statistical Methods: General < C - Mathematical and Quantitative Methods, F10 - General < F1 - Trade < F - International Economics, F43 - Economic Growth of Open Economies < F4 - Macroeconomic Aspects of International Trade and Finance < F - International Economics, O57 - Comparative Studies of Countries < O5 - Economywide Country Studies < O - Economic Development, Technological Change, and Growth
Keywords:	specialization, semiparametric regression

powered by ScholarOne
Manuscript Central™

Semiparametric analysis of the specialization-income relationship

Luca De Benedictis* Marco Gallegati† Massimo Tamberi‡

May 30, 2005

Abstract

In this paper we investigate the empirical relationship between overall specialization and per capita income. The metric we use to measure overall specialization is the median of the sectoral distribution of the Balassa index of Revealed Comparative Advantages applied to 4-digit (SITC rev.2) sectoral export manufactured data for 39 countries over the period 1985-2001. Once we take account of country specificities using fixed effects GAM, the results of the semiparametric analysis indicate that sectoral diversification of manufactured exports increases monotonically with development, and thus the reconcentration observed at high levels of income in the recent literature may not be linked to trade-induced specialization

Keywords: Specialization, Semiparametric regression
JEL codes: C14, F10, F43, O57

*University of Macerata, Macerata, Italy
†DEA and SIEC, Politechnic University of Marche, Ancona, Italy
‡DEA and SIEC, Politechnic University of Marche, Ancona, Italy

For Peer Review

Submitted Manuscript

1 Introduction

The analysis of sectoral specialization and industrial diversification across countries and along time has recently gained momentum. The extension of new growth theories to a multisectoral framework (Grossman and Helpman, 1991; Acemoglu and Zilibotti, 1997) and the availability of internationally comparable sectorally disaggregated data sets generate a new stream of applied research on the behaviour of industrial specialization along the development path (Imbs and Wacziarg, 2003; Kalemli-Ozcan, Sørensen and Yosha, 2003; Koren and Tenreyro, 2004).

In spite of largely relying on theoretical arguments related to international trade specialization and the evolution of comparative advantages, all the above mentioned empirical studies use employment, production or value added sectoral data in their analysis of industrial diversification.[1] The contribution of the present paper is to fully discuss the question under scrutiny from the point of view of international export flows. More specifically, we provide some evidence about sectoral diversification without making any *a priori* assumption about the specific functional form of the relationship between specialization and development.

2 Semiparametric regression models

In order to investigate the specialization pattern of different countries we analyse the shape of the relationship between per capita income and the median of the BI[2] using semiparametric regression models,[3] as such a methodology allows us exploring the issues related to the shape and the statistical significance of the relashionship between export diversification and income without making any *a priori* explicit or implicit assumption about the relationship, that is nonparametric regression.

Our sample is based on a balanced panel of exports data, based on the SITC rev.2 classification at the four digit level for 539 manufacturing sector[4] (source: ECLAC-UN CAN2000), and per capita income[5] (source:

[1]The only noticeable exception is Feenstra and Rose (2000). They notice that "... there is a strong negative correlation between the number of goods that a country exports and its ranking [income level]. ..., because rich countries tend to be open and diversified exporters, while poor countries tend to be closed and specialized exporters."

[2]Following De Benedictis and Tamberi (2004) we consider the median of the sectoral empirical density function of the Balassa (1965) Index (BI) of Revealed Comparative Advantages as a summary measure of overall specialization

[3]We also compare the median of the BI to three other popular indexes of overall specialization: the Gini and Herfindahl indexes, and the so-called Country Gini. All the indexes are correlated to each other, with the only exception of the Herfindahl index that appears to be different from the others in some respects).

[4]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.

[5]The choice of total income as a basis for the selection of countries was made to avoid

For Peer Review

PENN World Table 5.6) for 39 countries from 1985 to 2001, with countries selected on the basis of total GNP PPP 1998 (> 100 billions as in WB WDR data set).

There are several approaches available to estimate nonparametric regression models.[6] One such model is the *generalized additive regression model* (GAMs), [7] a combination of smooth nonparametric functions and parametric forms where it is possible to evaluate the statistical significance of the smooth nonparametric components. The result is an additive semiparametric model where the standard linear regression set-up is re-casted by modeling the dependent variable y_t 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,

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

where the functions $f_j(.)$ are *smooth* regression functions to be estimated from the data, like *loess* or spline functions, and the estimates of $f_j(x_{ij})$ for every value of x_{ij}, written as $\widehat{f}_j(x_{ij})$, are obtained using a fitting algorithm known as *backfitting*.

In semiparametric regressions like equation (2) the partial regression function $f(.)$ is estimated by the data through a scatterplot smoother like

$$\widehat{f}\left(t\right) = \sum_{j=1}^{k} S_\lambda\left(t, x_{ij}\right) y_i, \tag{2}$$

where S_λ is a weight function parametrized by the smoothing parameter λ which describes the proportion of the sample used in estimating $f\left(t\right)$ (with t being the target point). The span of the estimator λ controls the degree of smoothing applied to the data: the larger the span, the smoother the results, and vice versa. The choice of the smoother, even if it is in someway subjective, may be related to theoretical considerations or to the characteristics of the data. In what follows we employ the *loess* for the estimation of the partial-regression function $f_j(.)$.

2.1 Simple GAM

For our analysis a suitable model may be represented by the following nonparametric simple GAM regression

possible distortions due to the presence of too small economies; nevertheless relative per capita incomes range from 7 (Pakistan) to 100 (USA) in 1998.

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

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

For Peer Review

$$m_{ct} = \alpha + lo(ypc_{ct}) + \varepsilon_{ct}, \tag{3}$$

where m_{ct}, is the median of the BI, ypc_{ct} per capita income, $lo(ypc_{ct})$ some unknown function of per capita income, α the intercept, ε_{ct} are zero mean independently distributed errors, and indexes $c = 1, ..., C$ and $t = 1, ..., T$ refer to countries and time, respectively.

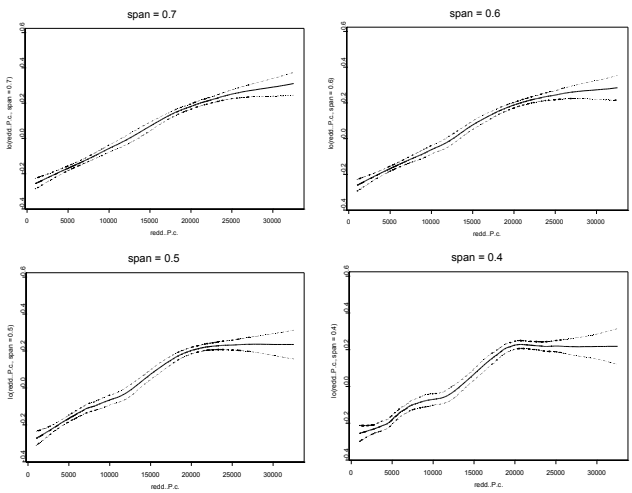

Figure 1: Nonparametric fitted functions from simple GAM regression

Figure 1 displays the nonparametric fitted functions, $lo(\widehat{ypc_{ct}})$, and the 95-percent confidence bands obtained estimating equation (4) using different span parameters in order to control for the robustness of the relationship linking sectoral export concentration and per capita income (the values of the span parameters from the right bottom panel to the left top panel are 0.4, 0.5, 0.6 and 0.7, respectively). These smooth plots are drawn by connecting the points in plots of the fitted values for each function against its regressor, while the dashed lines above and below the smooth curves are constructed, at each of the fitted values, by adding and subtracting two pointwise standard-error.[8] Thus, the plots of ypc versus $lo\,(\widehat{ypc})$ from a GAMs may be interpreted as the analogous of the coefficients in a linear regression. The shapes of the fitted functions $lo(\widehat{ypc_{ct}})$ in figure 1 indicate

[8]Under additional assumptions of no bias these upper and lower curves can be viewed as approximate 95% pointwise confidence intervals bands.

two main findings: i) the relationship between sectoral export diversification and per capita income decreases moving from low to high levels of per capita income, and ii) there is some evidence of a nonlinear relationship (with both findings being much more evident as the span parameter of the loess function gets smaller). Thus, the results from simple GAMs seem to suggests a piecewise increasing relationship where sectoral diversification of manufactured exports increases strongly initially, and then flattens out at the later part of the development process (a result that is particularly evident at the smallest span parameter values).

The nonparametric fitted curves in figure 1 reflect the relationship between sectoral concentration and income for an average country and, thus, implicitly hypothesises the equivalence between all pairs (m_{ct}, ypc_{ct}) regardless of any considerations about country-specific characteristics. But, in reality, as countries may differ significantly on many accounts, *i.e.* size, degree of openness and the quality of institutions (just to mention the main variables around which the empirical growth literature has debated in the recent years), the failure to consider these country-specific characteristics can lead to errors in estimating the specialization-income relationship. A rapid glance at figure 2 confirms such a impression as each country shows a particular trend that is independent from the initial level of its per capita income or its initial level of m.

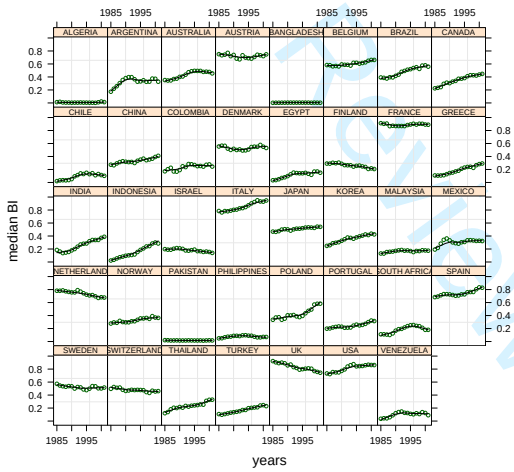

Figure 2: The median of the Balassa Index for the 39 countries of the sample

2.2 Fixed-effects GAM

In this subsection we take into account the role of country-specific effects for the relationship between the pattern of overall specialization and the

For Peer Review

level of per capita income, *i.e.* the marginal link between ypc and m net of countries specificities. A way to include such country-specific effects inside the GAM framework is to modify the parametric component of equation (4) through the introduction of an intercept term for each country of the sample capturing country-specific time-invariant shifts in the level of sectoral diversification. Hence, equation (4) becomes

$$m_{ct} = \alpha_c + lo(ypc_{ct}) + \epsilon_{ct}, \tag{4}$$

with the parametric component α_c indexes $c = 1, ..., C-1$ representing a set of dummy regressors corresponding to the number of countries of the sample minus one, and the nonparametric component given by a smooth term for per capita income as in equation (4).

Figure 3 displays the shapes of the smooth curves obtained by fitting the generalized additive model of equation (5). The picture of the relationship between sectoral diversification and income changes significantly once we isolate the influence of country specificities. Indeed, a monotonic linear relationship between income and diversification emerges in all the cases, thus showing that, when country specificities are taken into account, sectoral diversification of manufactured exports always increases with development. Such a result contrasts with the evidence stemming from other recent studies about the stages of sectoral specialization along the development path (Imbs and Wacziarg, 2003, and Koren and Tenreyro, 2004) where a non-monotonic specialization patterns is found. [9] Thus, our results with manufactured exports data seems to indicate that the two stages of diversification documented in these studies using employment and value added data are not to linked to trade-induced specialization or agglomeration.

Moreover, in order to evaluate the statistical significance of the nonparametric fitted functions estimated using different smoothing parameters in equation (4) and (5) we report in Table 1 the results of a type of score test. The column headed *"Nonpar d.f."* contains the nonparametric degrees of freedom used up by the fit. The column headed *"Nonpar* χ^2*"* contains an approximate χ^2-statistic (see Bowman and Azzalini, 1997, p.163) and represents a type of score test to evaluate, through the p-values reported in the last column, the contribution of the nonparametric term in equations (4) and (5) and may be used as a specification test for the significance of the estimate of $lo(\widehat{ypc}_{ct,j})$. The results in Table 1 indicate that per capita income contributes significantly to the fitted semiparametric regression both in simple and fixed-effects GAM and that its importance slightly decrease as the span gets smaller.

[9] For example, Imbs and Wacziarg (2003) provided empirical evidence of a U-shaped pattern using data on sector level employment and value added with sectoral diversification followed by sectoral concentration.

For Peer Review

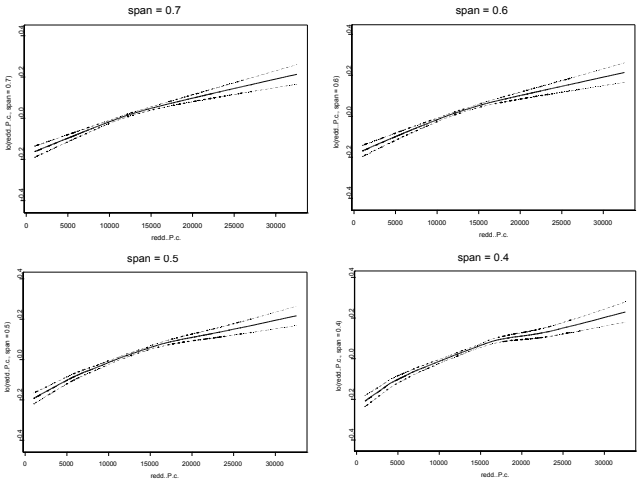

Figure 3: Nonparametric fitted functions from fixed-effects GAM regression

Table 1: Degrees of freedom and F-values for nonparametric effects

$lo(\widehat{ypc_{ct}})$	$m_{ct} = \alpha + lo(ypc_{ct}) + \epsilon_{ct}$			$m_{ct} = \alpha_c + lo(ypc_{ct}) + \epsilon_{ct}$		
span	*Nonpar. d.f.*	*Nonpar.* χ^2	$P(\chi^2)$	*Nonpar. d.f.*	*Nonpar.* χ^2	$P(\chi^2)$
0.7	1.3	20.76	0.00	1.3	22.07	0.00
0.6	1.7	20.40	0.00	1.7	21.41	0.00
0.5	2.5	18.62	0.00	2.5	18.26	0.00
0.4	3.8	15.84	0.00	3.8	13.19	0.00

3 Conclusion

In this paper we look at exports trade manufactured data in order to shed some light on the link between specialization and per capita income using semiparametric regression methods. A simple GAM of the median of the BI index on per capita income suggests a piecewise increasing relationship where sectoral diversification of manufactured exports increases strongly initially, and then flattening out at higher levels of income. But the picture of the relationship changes significantly once we isolate the influence of country

Peer Review

specificities using fixed effects GAM, as a monotonic linear relationship between income and diversification emerges along the development path. Thus, as our results indicate, once we account for country specificities, sectoral diversification of manufactured exports always increases with development, the reconcentration observed at high levels of income in some studies (Imbs and Wacziarg, 2003) may not be linked to trade-induced specialization.

4 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.

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

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

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

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

Feenstra R. and Rose A. (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 Helpman E. (1991), *Innovation and Growth in the Global Economy*, MIT Press.

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

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

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

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

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

For Peer Review