Discussion Papers
Collana di
E-papers del Dipartimento di Economia e Management – Università di Pisa

Davide Fiaschi - Andrea Mario Lavezzi - Angela Parenti

On the Determinants of Distribution Dynamics

Discussion Paper n. 165
2013

Discussion Paper

n. 165

Davide Fiaschi - Andrea Mario Lavezzi - Angela Parenti

On the Determinants of Distribution Dynamics

Abstract

In this paper we propose a novel approach to identify the impact of growth determinants on the distribution dynamics of productivity. Our approach integrates counterfactual analysis with the estimation of stochastic kernels. The counterfactuals are constructed from a semi-parametric growth regression, in which the cross-section heterogeneity in the growth determinants is removed. The methodology also allows us to test for potential distributional effects in the residuals. We illustrate the usefulness of the proposed methodology by an application to a cross-section of countries, which highlights the significant impact on inequality and polarization in the world productivity distribution of growth determinants from an augmented Solow model.

Classificazione JEL: C14; C21; O40; O50

Keywords: Convergence, inequality, polarization, distribution dynamics, counterfactual analysis.

Contents

I. Introduction

The world income distribution has been the subject of many studies over the last two decades, leading to the identification of a stylized fact: the distribution of per capita income underwent a substantial change from a unimodal shape in the 1960s to a twin-peaked shape in the 1990s (see, e. g., Quah, 1996a, and Durlauf et al., 2005). The same twin-peaked distribution also characterizes the regional distribution of productivity in Europe (see, e. g., Fiaschi and Lavezzi, 2007). However, it is still unclear whether the observed twin-peaks are a persistent phenomenon (see Galor, 2007), and which factors drive the formation of the two peaks (Quah, 1996b, 1997, Beaudry et al., 2005).

In this paper we address the latter issue, and propose a new methodology to measure the *distributional effect* of individual growth determinants, that is their role in favoring convergence or divergence in the distribution dynamics. Our approach combines semiparametric growth regressions (see, e. g. Liu and Stengos, 1999) with the estimation of stochastic kernels, i. e. of the operators mapping current distributions into future distributions of income or productivity.[1] Specifically, to evaluate the distributional impact of a given variable, we provide a generalized method of counterfactual analysis, based on the comparison between *actual* and *counterfactual* distributions to estimate short-run effects, and between actual and counterfactual *ergodic* distributions to identify long-run tendencies.

The counterfactuals are based on the cross-sectional heterogeneity of growth determinants, i. e. on how the distribution would have looked like if there were no heterogeneity in a specified variable across countries. In particular, we are able to identify for different ranges of per capita (or per worker) income the *direction* of the effect of a variable, i. e. whether the variable favors convergence or divergence in the distribution, an aspect which cannot be captured by the estimation of a single parameter as in standard growth regressions. In addition, the proposed methodology suggests a test

[1]See, e.g., Quah (1997) for details.

for the presence of distributional effects in the residuals of a growth regression, which provides a goodness-of-fit test of the specification of the semiparametric model we utilize in the analysis.

To illustrate the usefulness of our methodology, we apply our techniques to a sample of countries similar to the one used by Beaudry et al. (2005). Differently from Beaudry et al. (2005) we find that the cross-country *distribution* of investment ratios, labor force growth and human capital plays a role in the observed tendency to polarization of the distribution.

The paper is organized as follows: Section II. describes the methodology for the empirical analysis and clarifies the relation with existing approaches; Section III. presents the empirical application to the sample of Beaudry et al. (2005); Section IV. concludes. The appendices contain details on data and on the methodology.

II. Methodology

In this section we present the method for the empirical analysis and clarify the aspects of its novelty with respect to other approaches. Our method can be summarized as follows: we first estimate a semiparametric growth regression. Then we utilize the results to estimate counterfactual distributions with respect to individual variables of interest, in order to identify their contribution to convergence or divergence. The latter effect is denoted *marginal growth effect*.

II.A. Related Literature

Two main approaches to study convergence exist in the literature: the "growth regression approach" (GRA) and the "distribution dynamics approach" (DDA). By applying the GRA, it is possible to analyze whether economies are, on average, converging towards their steady-state level of per capita income or productivity, and to identify the average effect of growth determinants. The DDA, instead, aims at understanding how the whole cross-section distribution evolves over time.[2]

The most representative examples of the GRA are the so-called "Barro regressions" (see, e.g., Barro, 1991, and Barro and Sala-i Martin, 2004). Many applications of this method showed evidence of conditional convergence across different economies, that is of a negative relation between the growth rate and initial income levels, after controlling for other growth determinants.[3] De La Fluente (2003), in the spirit of the present paper, extends the GRA approach by decomposing the measures of σ and β-convergence (Barro and Sala-i Martin, 2004) into sums of *partial* σ and β-convergence measures, in order to assess the individual contribution to convergence of the explanatory variables included in a growth regression. De La Fluente (2003) defines such methodology "convergence accounting".

The alternative DDA, proposed by Danny Quah in a number of papers (see, e. g., Quah, 1993, 1996a,b, 1997) stems from criticism to the GRA for its inability to capture phenomena such as *mobility*, *stratification* and *polarization* in the world income distribution. On the contrary, operators such as stochastic kernels (or transition matrices) may reveal information on such aspects of the growth process.[4] Quah (1996b, 1997) takes a further step, with the aim of investigating the caues of the observed polarization in the distribution dynamics, introducing to this purpose *conditioned* stochastic kernels. In particular, in Quah (1996b) conditioned stochastic kernels are based on residuals from two-sided regressions of labor productivity on human capital, physical capital, and country dummies. Differently, Quah (1997) defines conditioned stochastic kernels as operators mapping *unconditioned* income levels into *conditioned* income levels, that is incomes normalized: "on the basis of incomes relative to one's neighbours appropriately weighted" (Quah, 1997,

[2]See Quah (1997) for a more detailed discussion, and Durlauf et al. (2005) for an exhaustive survey of the different methodologies adopted in the empirical analysis of economic growth.

[3]Uncertainty in the choice of explanatory variables in the "Barro regressions" is one of the main difficulties of the GRA. See Durlauf et al. (2005) for a thorough analysis of this issue.

[4]In addition to these types of criticisms, Bernard and Durlauf (1996) show that a negative sign of the coefficient of initial income in a growth regression does not imply absolute or conditional convergence, as the data-generating process may be characterized by multiple, locally stable, equilibria.

p. 47), where weights are calculated with respect to one of the variables affecting the income dynamics. The assessment of the effect of explanatory variables on the income dynamics is based on the comparison of actual and conditioned stochastic kernels.[5]

Another strand of literature proposes counterfactual analysis as an alternative methodology to identify the impact of individual explanatory variables on distributions (see e. g. DiNardo et al., 1996 and Machado and Mata, 2005). Beaudry et al. (2005) apply this analysis in a study of economic growth. In particular, they analyze in a cross-country setting the distributional effects of some growth determinants in the comparison of two periods, 1960-1978 and 1978-1998, as the second period is characterized by a tendency to polarization. By estimating linear growth regressions as a first step, they build counterfactual distributions for the second period assuming that a factor of interest (a coefficient of the estimated growth regression, or the distribution of a variable, e. g., investment ratios) maintains in the second period the same value taken in the first.

Finally, our methodology shares similarities with Cheshire and Magrini (2005), who combine the GRA with the DDA in the analysis of factors driving convergence in a large cross-section of European urban regions in the period 1978-1994. In particular, they estimate a *linear* growth regression model, compute counterfactual distributions under different assumptions on explanatory variables, and compare a "predicted" stochastic kernel (computed on the basis of fitted values of growth regression) with a "simulated" stochastic kernel (computed on the basis of alternative values of the explanatory variables in the growth regression).

Our method represents a generalization of existing methods in different aspects. i) With respect to the estimation of "conditioned" stochastic kernels of Quah (1997), our method is based on a multi-

[5] Another approach to study the causes of polarization is taken by Johnson (2005) and Feyrer (2008), who compare the observed per capita income dynamics to the dynamics of explanatory variables such as human capital, physical capital and TFP. By this approach, however, it is not possible to *directly* evaluate the effects of the variables on the income distribution dynamics, in particular the possible presence of divergence or convergence effects in different income ranges.

variate analysis to identify the effect of a specific variable, and not on the consideration of one variable at time. We are therefore able to control more precisely for the effects on growth of other variables, different from the "conditioning" ones, avoiding the omitted variable bias. Quah (1996b) performs a multivariate analysis, but only considers the residuals from this analysis to condition the stochastic kernel, and therefore may only obtain an estimate of the joint effect of the variables included in the regression. In addition, we not only estimate actual and counterfactual stochastic kernels, but also the implied ergodic distributions, an aspect so far neglected in the literature. ii) With respect to Cheshire and Magrini (2005), we use a more general semiparametric specification, instead of a linear regression, for the baseline estimation. iii) The same remark applies to Beaudry et al. (2005), who perform a counterfactual analysis in a linear regression framework. The advantage of using a nonparametric analysis with respect to Beaudry et al. (2005) is first of all that we can avoid the possible misleading evidence of instability in time of coefficients of linear regressions, which can instead depend on nonlinarities of the underlying model. In addition, to focus on the effect of the distribution of a variable, Beaudry et al. (2005) build a counterfactual in which the distribution of the first period affects income in the second period through the coefficients estimated in the latter. This procedure is valid only if the underlying models are linear or, if nonlinear, if they are stable across periods. Our approach to study the effect of the distribution of a variable, instead, is based on simply assuming away (see below) heterogeneity in a variable to build the counterfactual, and does not require extra hypotheses on the linearity or the stability of the growth model.

In the following we detail our methodology, which is based on six steps: i) estimation of a semiparametric growth regression model (Section II.B.); ii) computation of counterfactual productivity (Section II.C.i.); iv) estimation of counterfactual stochastic kernels (Section II.C.i.); v) estimation of counterfactual ergodic distributions (Section II.C.i.); vi) evaluation of the distributional effects of a variable and estimation of its marginal growth effect (Section II.C.ii.);

ii) test on the distributional effects of growth residuals (Section II.D.).

II.B. Modeling Productivity Growth

Define $y_i(t)$ labour productivity of country i ($i = 1, ..., N$) at time t. Labour productivity of country i at time $T > 0$, therefore, can be expressed as:

$$y_i(T) = y_i(0)e^{g_i T}, \tag{1}$$

where g_i is the annual rate of growth of productivity in country i, between periods 0 and T.

Assume that g_i is a function of K explanatory variables, collected in vector $\mathbf{X}_i = (X_{i,1}, ..., X_{i,K})$, and of a residual component υ_i accounting for unobservable factors, that is:

$$g_i = \varphi(\mathbf{X}_i, \upsilon_i). \tag{2}$$

Differently from other approaches to counterfactual analysis, we model the determinants of the growth rate by a semiparametric specification, that is:[6]

$$g_i = m(\mathbf{X}_i) + \upsilon_i = \alpha + \sum_{j=1}^{K} \mu_j(X_{i,j}) + \upsilon_i \tag{3}$$

where α is a constant term, $\mu_j(\cdot)$ are one-dimensional nonparametric functions operating on each of the K elements of $\mathbf{X}_i$, and υ_i is an error term with the properties: $E(\upsilon_i|\mathbf{X}_i) = 0$, $var(\upsilon_i|\mathbf{X}_i) = \sigma^2(\mathbf{X}_i)$ (i.e. the model allows for heteroskedasticity).[7]

[6] Notation refers to Härdle et al. (2004).

[7] Durlauf et al. (2001) consider a growth regression framework in which the impact of the explanatory variables is nonlinear. Specifically, they condition the marginal impact of a variable to the initial level of per capita income (as we do in the following), and find significant nonlinearities. However, the main difference with respect to the present analysis is that Durlauf et al. (2001) do not embed this exercise into a counterfactual analysis of the distribution dynamics of labour productivity.

II.C. Distributional Effects of Individual Variables

Denote by $\mathbf{X}_{i,\underline{k}}$ the vector of all explanatory variables but $X_{i,k}$ for country i, i. e.:

$$\mathbf{X}_{i,\underline{k}} = (X_{i,1}, ..., X_{i,(k-1)}, X_{i,(k+1)}, ..., X_{i,K}).$$

Eq. (3) can be rewritten as:

$$g_i = \alpha + \mu_k(X_{i,k}) + \sum_{j \neq k} \mu_j(X_{i,j}) + \upsilon_i. \tag{4}$$

Substituting Eq. (4) into Eq. (1) leads to the following expression for productivity at time T:

$$\begin{aligned} y_i(T) &= y_i(0)e^{[\alpha+\mu_k(X_{i,k})+\sum_{j\neq k}\mu_j(X_{i,j})+\upsilon_i]T} = \\ &= \underbrace{y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}}_{y_{i,\underline{k}}(T)} \underbrace{e^{\mu_k(X_{i,k})T}}_{e^{g^M_{i,k}T}} \underbrace{e^{\upsilon_i T}}_{e^{g^R_i T}}, \end{aligned} \tag{5}$$

where $y_{i,\underline{k}}(T) = y_i(0)e^{[\alpha+\sum_{j\neq k}\mu_j(X_{i,j})]T}$ is the level of productivity in period T obtained by "factoring out" the effect of $X_{i,k}$; $g^M_{i,k} = \mu_k(X_{i,k})$ is the part of the annual growth rate of y_i explained by $X_{i,k}$, capturing the "marginal" effect of $X_{i,k}$ on g_i and, finally, $g^R_i = \upsilon_i$ is the annual "residual growth", not explained by the variables in $\mathbf{X}_i$. The modelling of growth in Eq. (5) will be the basis for the identification of the distributional effects of the *k-th* variable.

II.C.i. Counterfactual Stochastic Kernels and Ergodic Distributions

We define the *counterfactual productivity* $y^{CF}_{i,k}(T)$, the productivity level that a region would attain at time T if there were no differences within the sample in terms of the *k-th* variable (whose values are collected in the N-dimensional vector $\mathbf{X}_k$). That is, the computation of the values of $y^{CF}_{i,k}(T)$ aims at capturing the effect on the productivity distribution of the cross-sectional distribution of the *k-th* variable. To isolate this effect, we will impose to each country the *cross-section average value* of the variable.[8]

Hence, the *counterfactual growth rate* of country i with respect to the k-th variable, $g_{i,k}^{CF}$, is defined as:

$$g_{i,k}^{CF} \equiv \hat{\alpha} + \sum_{j \neq k} \hat{\mu}_j(X_{i,j}) + \hat{\mu}_k(\bar{X}_k), \tag{6}$$

where $\bar{X}_k = N^{-1} \sum_{j=1}^{N} X_{k,j}$, and $\hat{\mu}_k(\cdot)$ is the estimated smoothed function relative to the *k-th* variable, obtained from the estimation of Eq. (3). The counterfactual productivity of country i in period T, relative to variable k, is therefore defined as:

$$y_{i,k}^{CF}(T) \;\equiv\; y_i(0)e^{g_{i,k}^{CF}T} = y_i(0)e^{[\hat{\alpha}+\sum_{j\neq k}\hat{\mu}_j(X_{i,j})+\hat{\mu}_k(\bar{X}_k)]T}. \tag{7}$$

Counterfactual productivities are the bases to compute *counterfactual stochastic kernels*. Specifically, the *actual* and *counterfactual* stochastic kernels are respectively defined as $\phi(\mathbf{y}(T)|\mathbf{y}(0))$ and $\phi^{CF}(\mathbf{y}_k^{CF}(T)|\mathbf{y}(0))$, where $\mathbf{y}(0)$, $\mathbf{y}(T)$ and $\mathbf{y}_k^{CF}(T)$ are the vectors collecting productivity levels at times 0 and T.[9]

The actual stochastic kernel $\phi(\cdot)$ maps the distribution of (relative) productivity in period 0 into the distribution of (relative) productivity in period T. The counterfactual stochastic kernel $\phi^{CF}(\cdot)$, instead, maps the distribution of (relative) productivity in period 0, into the distribution of counterfactual relative productivities in period T. Therefore, the counterfactual stochastic kernel shows,

[8] We use the *average value* because we are interested in considering the possible differences from the *average impact* of the variable (usually measured by the OLS coefficients). Alternatively, if the variable of interest were characterized by the presence of outliers, the median of the distribution could be preferable. Other counterfactuals could be built using quantiles of the distribution.

[9] In general, a stochastic kernel is an operator mapping the density of a variable at time t into its density at time $t+\tau$, $\tau > 0$, and indicates for each level of the variable in period t its the probability distribution in period $t+\tau$ over the possible values of the variable. The relation between the densities and the stochastic kernel is: $f_{t+\tau}(z) = \int_0^\infty g_\tau(z|x) f_t(x)\, dx$, where z and x are values of the variable, and $g_\tau(z|x)$ is the stochastic kernel. To estimate the stochastic kernel $g_\tau(z|x) = g(z,x)/f(x)$ we estimated the joint density of z and x, $g(z,x)$, and the marginal density of x, $f(x)$. In the estimation of $g(z,x)$ we followed Johnson (2005), who used the *adaptive kernel estimator* discussed by Silverman (1986, p. 100), in which the window of the kernel (Gaussian in our case) decreases when the density of observations increases (see Appendix F for more details).

for every initial productivity level, the probability distribution over productivity levels at time T had the cross-country heterogeneity in the variable k been absent. This implies that the possible differences with respect to the probability distribution based on the actual stochastic kernel depends on the *k-th* variable, in particular on its distribution across countries.

For actual and counterfactual stochastic kernels we estimate the corresponding ergodic distributions, i.e. the *actual* and the *counterfactual ergodic distribution*, following the procedure proposed by Johnson (2005) modified to take into account that we are dealing with normalized variables.[10] The ergodic distribution highlights whether the estimated distribution dynamics over the period of interest has completely exhausted its effects or, otherwise, significant distributional changes are expected in the future.

II.C.ii. The Distributional Effect of Individual Variables and the Marginal Growth Effect

To evaluate the distributional effect of individual variables: i) we assess the capacity of an individual variable to make actual and counterfactual stochastic kernels differ; and, ii) we highlight its *marginal growth effect* with respect to initial productivity. This will allow us to identify whether a variable is a source of convergence or divergence, in particular by identifying which ranges of the productivity distribution are affected.

To identify differences between actual and counterfactual kernels, we express the value of (log) actual productivity in period T, $y_i(T)$, in terms of the counterfactual productivity, $y_{i,k}^{CF}(T)$:

$$\begin{aligned}\log\left(y_i(T)\right) &= \log\left(y_{i,k}^{CF}(T)\right) + \alpha T + \sum_{j\neq k}\mu_j(X_{i,j})T + \mu_k(X_{i,k})T + \\ &- \hat{\alpha}T - \sum_{j\neq k}\hat{\mu}_j(X_{i,j})T - \hat{\mu}_k(\bar{X}_k)T + \upsilon_i T. \qquad (8)\end{aligned}$$

[10] See Appendix H.

The expected value of (the log of) actual productivity of country i in period T conditional to actual productivity in period 0, $E[log(y_i(T))|y_i(0)]$, is obtained from the actual stochastic kernel setting $\tau = T$. In particular, its relation with the expected value from the counterfactual kernel can be expressed as:

$$\begin{aligned} E\left[\log\left(y_i(T)\right)|y_i(0)\right] &= E\left[log\left(y_{i,k}^{CF}(T)\right)|y_i(0)\right] + E\left[\alpha - \hat{\alpha}|y_i(0)\right]T + \\ &+ \sum_{j\neq k} E\left[\mu_j(X_{i,j}) - \hat{\mu}_j(X_{i,j})|y_i(0)\right]T + \\ &+ E\left[\mu_k(X_{i,k})|y_i(0)\right]T - E\left[\hat{\mu}_k(\bar{X}_k)|y_i(0)\right]T + E\left[\upsilon_i|y_i(0)\right]T \end{aligned} \tag{9}$$

If $\hat{\alpha}$ and $\hat{\mu_j}$ $(j = 1, ..., K)$, are conditional unbiased estimators of α and μ, and $E\left[\upsilon_i|y_i(0)\right] = 0$, Eq. (9) reduces to:[11]

$$E\left[\log\left(y_i(T)\right)|y_i(0)\right] - E\left[log\left(y_{i,k}^{CF}(T)\right)|y_i(0)\right] = \left(E\left[\mu_k(X_{i,k})|y_i(0)\right]T - \mu_k(\bar{X}_k)\right)T \tag{10}$$

From Eq. (10), we can derive a condition for the equality of the expected values of productivity based on actual and counterfactual kernels. Specifically, these values are equal, i. e.:

$$E\left[\log\left(y_i(T)\right)|y_i(0)\right] = E[\log(y_{i,k}^{CF}(T))|y_i(0)] \tag{11}$$

if:

$$E[\mu_k(X_{i,k})|y_i(0)] = \mu_k(\bar{X}_k). \tag{12}$$

The result in Eq. (12) depends on the fulfilment of the following two conditions:

1. $E[\mu_k(X_{i,k})|y_i(0)] = E[\mu_k(X_{i,k})]$, i. e. $\mu_k(X_{i,k})$ and $y_i(0)$ are independent, that is the impact of the *k-th* variable on productivity in region i is independent from the initial productivity level.

2. $E[\mu_k(X_{i,k})] = \mu_k(E[X_{i,k}]) = \mu_k(\bar{X})$, i. e. $\mu_k(\cdot) = \beta_k X_{i,k}$, that is the *marginal* impact of the *k-th* variable is constant, i.e. the term $X_{i,k}$ has a linear effect on growth.

[11]Notice that $E\left[\hat{\mu}_k(\bar{X}_k)|y_i(0)\right] = E\left[\hat{\mu}_k(\bar{X}_k)\right]$.

Therefore, if Conditions 1 and 2 hold, we obtain the condition in Eq. (12), i. e.:

$$E[\mu_k(X_{i,k})|y_i(0)] = E[\mu_k(X_{i,k})] = \mu_k(E[X_{i,k}]) = \mu_k(\bar{X}_k). \quad (13)$$

Eq. (13) represents a necessary condition for the equality of the actual and counterfactual stochastic kernels and, therefore, for the absence of distributional effects of the *k-th* variable.

In growth empirics violations of Conditions 1. and 2. are common; for example (Liu and Stengos, 1999) find violations of Condition 2., while (Durlauf et al., 2001) find violations of Condition 1. In Section III. below we find a confirm of these violations.

As a second step to evaluate the impact of an individual variable on distribution dynamics, in particular whether it is a source of convergence or divergence, we need to identify the specific relation between the contribution of that variable to productivity growth and initial productivity levels. To this purpose, we define $g^M_{i,k} = \mu_k(X_{i,k})$ the *marginal growth effect* of the *k-th* variable in Eqq. (3)-(5). It may be observed that the estimation of Eq. (3) must include all the explanatory variables in order to avoid omitted-variable problems and obtain unbiased estimates.

The marginal effect of the *k-th* variable on the distribution dynamics is identified by estimating marginal growth conditioned on the initial level of productivity, i. e. by estimating $\phi^M(\mathbf{g}^M_k|\mathbf{y}(0))$, where $\mathbf{g}^M_k$ is the vector collecting the $g^M_{i,k}$'s. If the estimate of the marginal effect does not result statistically different from its unconditional mean, i. e. if $\phi^M(\mathbf{g}^M_k|\mathbf{y}(0)) = E[\mathbf{g}^M_k]\ \forall \mathbf{y}(0)$, then the *k-th* variable has no distributional effects. On the contrary, if $\phi^M(\mathbf{g}^M_k|\mathbf{y}(0))$ is statistically different from its unconditional mean and, in particular, it is an increasing (decreasing) function of $\mathbf{y}(0)$, then the *k-th* variable is a source of divergence (convergence).

Since the estimation of the marginal effect in semiparametric models is performed through the backfitting technique, it requires as identification assumption that: $E_{\mathbf{X}_k}[\mu_k(\mathbf{X}_k)] = 0$ (see Härdle et al., 2004, pp. 212-222). Therefore, the unconditional mean of marginal growth will always be equal to zero in the estimation of the semiparametric terms in the growth regression.

II.D. Misspecification Test of Distributional Effects of Residual Growth

As a final step, we propose a test for the presence of possible misspecifications of the model for different levels of initial productivity. In particular, Eq. (5) suggests to consider $\hat{\mathbf{g}}^R$, defined as $\hat{\mathbf{g}}^R \equiv log\left(\frac{\mathbf{y}(T)}{\hat{\mathbf{y}}(T)}\right)$, to test that:

$$E[\hat{\mathbf{g}}^R|\mathbf{y}(0)] = E[\hat{\mathbf{g}}^R] = 0 \ , \forall \ \mathbf{y}(0). \tag{14}$$

If $\mathbf{y}(0)$ is included in the set of regressors, the condition in Eq. (14) ensures that there is no omitted variable inconsistency related to $\mathbf{y}(0)$ (see Wooldridge, 2002, pp. 61-63). Eq. (14) suggests to test the null hypothesis that $E[\hat{\mathbf{g}}^R|\mathbf{y}(0)] = 0$ for each $\mathbf{y}(0)$. We will use a bootstrap procedure to compute the confidence interval of $\hat{\mathbf{g}}^R$ conditioned to $\mathbf{y}(0)$.

III. An Empirical Application to a Cross-Section of Countries

[12] To illustrate the practical use of our methodology, we provide an empirical application on the distribution dynamics of labour productivity, comparing our results to those of Beaudry et al. (2005). The main result of the counterfactual analysis of Beaudry et al. (2005) is that the recent tendency to polarization is mainly explained by a change in the effects on growth of the accumulation factors, i.e. the investment ratio and the rate of growth of population. This claim is supported by evidence of instability of the estimated parameters for these variables in linear regressions on two subperiods: 1960-1978 and 1978-1998. In a different framework (based on data envelopment analysis), Henderson and Russell (2005) find that polarization is brought about by technological catch-up.

[12] Dataset and codes are available at author's website `http://dse.ec.unipi.it/~fiaschi/WorkingPapers.html`.

In our application, we do not exactly replicate the sample and time period of Beaudry et al. (2005). In particular, we consider data from PWT 7.1, while Beaudry et al. (2005) utilize PWT 6.0. This allows us to extend the analysis to the period 1960-2008. Moreover, while Beaudry et al. (2005) utilize data on human capital from Barro and Lee (1993), we choose the dataset on education from Cohen and Soto (2007).[13] Overall, we consider 73 (61 with data on education) out of 75 (68 with data on education) countries of the sample of Beaudry et al. (2005).[14]

Specifically, in Section III.A. we estimate the growth model of Eq. (3); in Section III.B. we test for the presence of distribution effects in residual growth; in Section III.C. we study the unconditional distribution dynamics of labour productivity that we will use as benchmark; finally, in Section III.D. we present the distributional impact of regressors.

III.A. The Estimation of a Growth Model for Countries

Following Mankiw et al. (1992) and Beaudry et al. (2005), in the estimation of Eq. (3) the annual average growth rate of per worker GDP (g) of a country is regressed on: i) the (log) initial level (**y0**); ii) the (log) average annual employment growth rate (**n**); iii) the (log) average annual investment ratio at constant price (**i/y**); and, iv) its (log) average years of schooling (**Edu**).[15] Results of the estimated models are reported in Table 1. All regressors initially enter the specification as nonparametric terms. However, they are substituted by linear terms if their effect results to be linear.[16]

[13]Cohen and Soto (2007) present compelling evidence that the accuracy of their database is superior to that of Barro and Lee (1993).

[14]See Appendix A for details on the sample and for the list of countries.

[15]Appendix B contains the definitions and the descriptive statistics of the variables. As in Mankiw et al. (1992) we augment the employment growth rate assuming a depreciation rate $\delta = 0.03$ and a rate of growth of technical progress $g = 0.02$.

[16]The semiparametric estimation is carried out following the approach describe in Wood (2006) based on penalized regression splines. In particular, we use *mgcv* packages in R Development Core Team (2012), with the option "REML" discussed in Wood (2011). Appendix C contains a brief description of the method.

	Model I		Model II	Model III	Model IV
Dep. Var: *g*	GAM		GAM	GAM	GAM
	1960-1978	**1978-1998**	**1960-1998**	**1960-2008**	**1960-2008**
Parametric coefficients:	Estimate		Estimate	Estimate	Estimate
const	0.0325***	0.0105	0.0210***	0.0209***	0.0796***
Non parametric coefficients:	EDF		EDF	EDF	EDF
y0	2.343***	1.000**	2.016***	1.409***	(-0.0083***)
n	2.139***	2.307***	2.511***	3.235***	2.565***
i/y	1.000***	1.630**	3.000***	3.220***	4.016***
Edu	-	-	-	-	(0.0091**)
AICc	-409.24	-393.86	-449.85	-479.46	-416.51
Dev. expl.	0.49	0.34	0.58	0.66	0.73
REML score	-187.75	-180.85	-206.48	-220.00	-184.76
Scale est.($*10^{-5}$)	18.72	27.09	10.30	6.81	4.90
Obs.	73	73	73	73	61

Table 1: Estimates of semiparametric Model (3) for different time periods. Significance codes: 0.01"***" 0.05"**" 0.1"*". Terms in parenthesis enter linearly in the preferred specification (i.e. the estimated degrees of freedom (EDF) are equal to one)

Following Beaudry et al. (2005), in Model I we check the stability of the growth regression in the period 1960-1978 versus 1978-1998,[17] by focusing on the model with only **y0**, **n** and **i/y**, which correspond to the "Solovian" growth determinants. The model displays significant nonlinearities in both periods. The estimates of the two subperiods, however, turn to be not statistically different at 95% significance level for each variable, according to a bootstrap test reported in Appendix D. Therefore, in contrast with Beaudry et al. (2005) the growth model appears stable over the two subperiods. Given that we do not find evidence of instability of the model, in Model II we consider the whole period 1960-1998.

In Models III and IV we extend the analysis to the period 1960-2008. In particular, the comparison of the estimate of the Models II and III shows that we can actually extend the period of the analysis. In Model IV we add education to the explanatory variables.[18] In the

[17]Beaudry et al. (2005, fig. 2) show that polarization increases after 1978.

[18]Beaudry et al. (2005) consider human capital, institutions and possible nonlinear effects of initial income as robustness checks. They find that their main results are not affected.

preferred specification the initial level of productivity and education enter linearly. In particular, **y0** has a negative and significant effect, while **Edu** has a positive and significant effect.[19] The estimates of additive components related to **n** and **i/y** are shown as reported in Figures 1 and 2. In Appendix E we check that these results are robust to endogeneity.

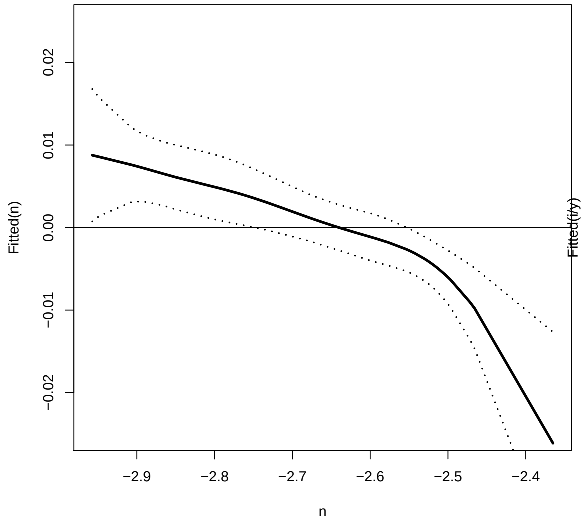

Figure 1: Estimate of additive component related to **n**. Thick line: estimates; dotted lines: 95% confidence bands.

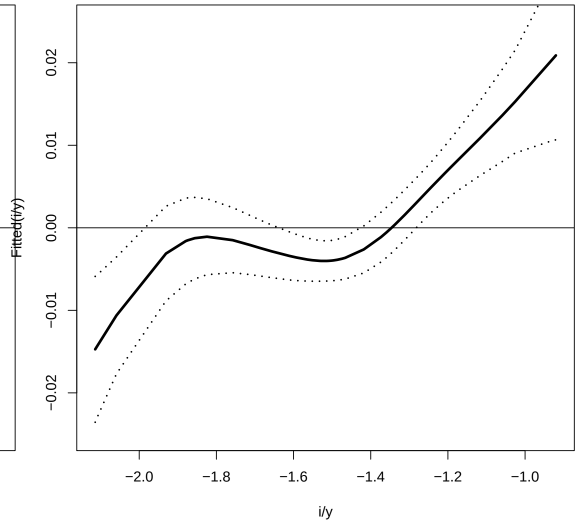

Figure 2: Estimate of additive component related to **i/y**. Thick line: estimates; dotted lines: 95% confidence bands.

Overall, from this initial part of the analysis, we do not find corroboration of the main result of Beaudry et al. (2005), that is a significant instability in the coefficients of the investment ratio and the population growth rate, that Beaudry et al. (2005) claim to be the main factors explaining the increase in polarization after 1978.

III.B. Misspecification Test of Residual Growth

Figure 3 reports the estimated distribution of the annual residual growth $\hat{\mathbf{g}}^R$ conditioned on the initial level of productivity $\mathbf{y}(\mathbf{0})$.

[19]Beaudry et al. (2005) generally also find a negative and significant coefficient for **y0** in both subperiods, but no evidence of instability in the estimated coefficient. The effect of **Edu**, instead, is generally nonsignificant in the second superiod and no evidence of instability in this coefficient is found.

We also report the conditional mean (thick line) with the corresponding confidence bands obtained by a bootstrap procedure, and a vertical line representing the unconditional mean, which is approximately zero as expected. Figure 3 shows that for any initial level of productivity most of the mass of the conditional distribution of residual growth is concentrated around the unconditional mean, and that the conditional mean is never statistically different from the unconditional mean. We then conclude that the residual growth of Model (3) has not significant distributional effects, i.e. the estimated model appears correctly specified, at least conditioning on the initial level of productivity (see Eq. 14).

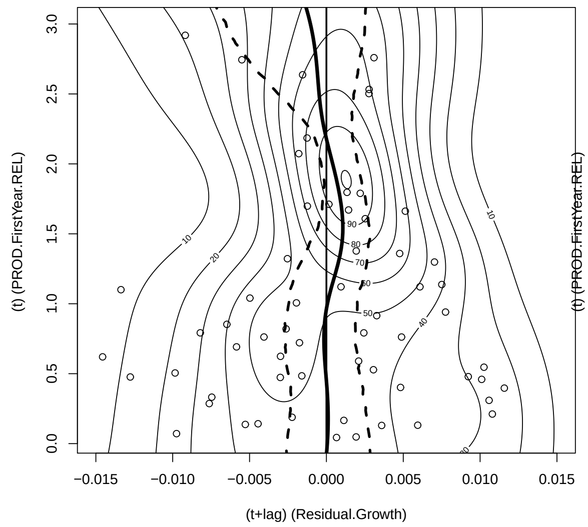

Figure 3: Conditional distribution of residual growth, the conditional mean (thick line), its confidence bands (dotted lines) and the unconditional mean (thin vertical line).

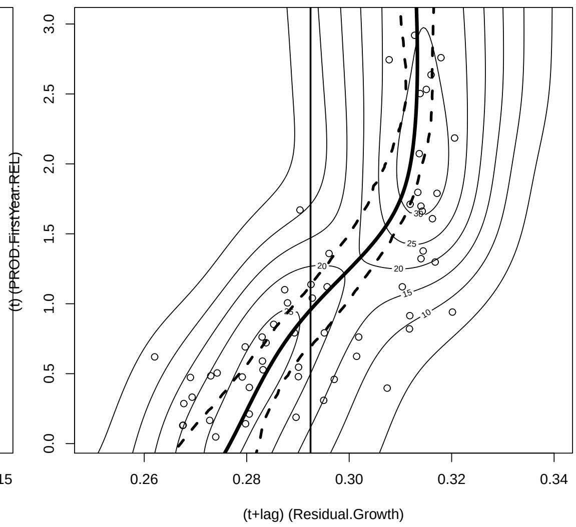

Figure 4: Conditional distribution of residual growth with *bias*, the conditional mean (thick line), its confidence bands (dotted lines) and the unconditional mean (thin vertical line).

In Figure 4 we report the result of the test for the model where regressors include only a constant, which represents the extreme omitted-variable case; as expected the test highlights the presence of omitted variables.

III.C. The Unconditional Distribution Dynamics

In this section we study the unconditional distribution dynamics of labour productivity. All stochastic kernels are estimated considering a time lag of τ =49 years, i. e. the whole period. In each figure displaying the estimate of the stochastic kernel we report: a solid line representing the estimated median value of productivity at $t + \tau$ conditioned on the productivity level at time t; the corresponding confidence band at 95% significance level (indicated by dotted lines) obtained by a bootstrap procedure,[20] and the 45° line.

Figure 5 reports the actual stochastic kernel of productivity, while Figure 6 the actual distributions (AD) of productivity in 1960 and 2008, along with the actual ergodic distribution (AED).

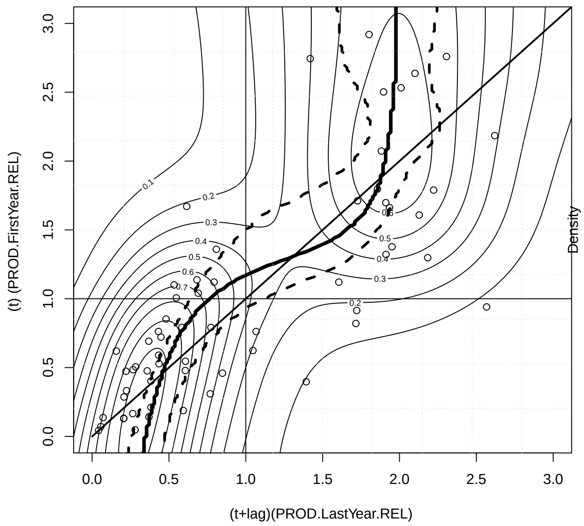

Figure 5: Actual stochastic kernel of productivity (lag=49). Thick line: median of the stochastic kernel; dotted lines: 95% confidence bands.

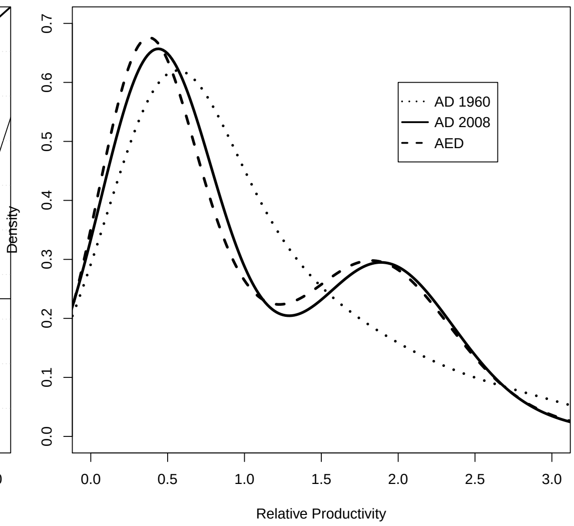

Figure 6: AD 1960 (dotted line), AD 2008 (solid line) and AED (dashed line) distributions of productivity.

Figure 5 shows that most of the mass is concentrated around the 45° line and, in particular, the median value crosses the 45° line from below in two points. This is reflected in the 2008 distribution,

[20]The procedure is illustrated in Appendix I.

showing two peaks in the proximity of the values of 0.5 and 2.[21] The ergodic distribution reflects this tendency (see Figure 6). This sample, therefore, confirms the presence of a tendency to polarization in the cross-country income distribution.

III.D. Conditional Distribution Dynamics

Given the results of the estimation of the growth model in Eq. (3), reported in Table 1, and once controlled for the potential presence of distributional effects in residual growth, the analysis proceeds by calculating and discussing the distributional impact of the variables present in the preferred specification of Table 1.

III.D.i. Initial Productivity

In Figure 7 we present the MGE for initial productivity. The identified pattern is consistent with conditional convergence, as the result in Table 1: the conditional mean of MGE is above the unconditional mean for countries with an initial productivity below the average, while the opposite holds for countries with above-average initial productivity.

The overall distributional impact seems sizeable, as highlighted by the comparison between the AD and CD in 2008 (see Figure 8) and the AED and CED (see Figure 9). If each country had had the same level of productivity in 1960 the distribution would have been more dispersed. This is already evident in the CD in 2008, but it is much more evident in the CED. Gini indexes reported in Table 2 quantify the reduction of inequality from the AD in 2008 to the CD in 2008 in about 10 base points (about the same holds

[21] Silvermans bootstrap tests for multimodality (see Appendix G for details) show that the null hypothesis of unimodality cannot be rejected for the 1960 distribution while it can be rejected at 1% of significance level for the 2008 distribution. Henderson et al. (2008) find the same results but with a lager sample of countries (see their Table III). Hall and York (2001) discuss as Silverman's "test tends towards conservatism, even in large samples, in the sense that the actual level tends to be less than the nominal one". Their Table 1 shows that that in the case of test of multimodality a nominal level of 5% corresponds to an actual level of 1%, a nominal level of 10% to 3.2 %, a nominal level of 20% to 10.2 %, etc.. In our tables we follow the common (in literature) conservative choice of reporting nominal levels.

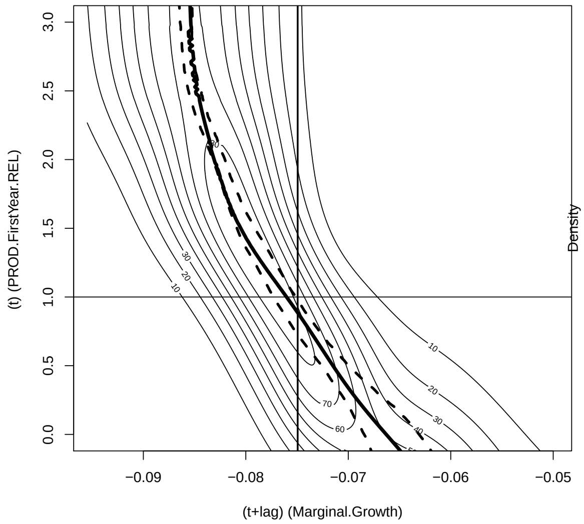

Figure 7: MGE conditioned on the initial level of productivity, estimated mean of MGE conditioned on initial level of productivity (thick solid line), its confidence bands (dashed lines), and unconditional mean (thin solid vertical line). Counterfactual variable: Initial Productivity.

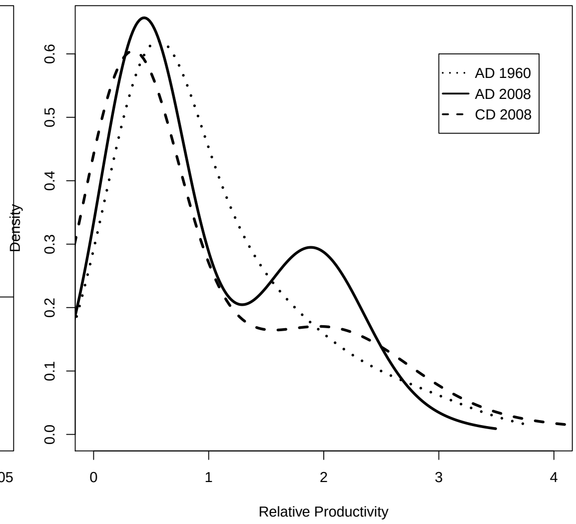

Figure 8: AD in 1960 (dotted line), AD in 2008 (solid line), and CD in 2008 (dashed line). Counterfactual variable: Initial Productivity.

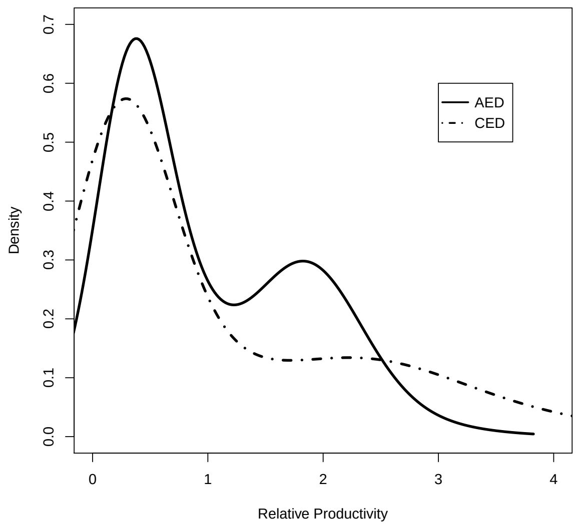

Figure 9: AED (thick line) and CED (thin line). Counterfactual variable: Initial Productivity.

	AD 1960	AD 2008	CD 2008
Gini	0.42	0.42	0.52
s.e.	(0.029)	(0.026)	(0.030)
$BIPOL_{NI}$	NA	0.69	0.62
s.e.	(-)	(0.087)	(0.329)
Unimodality Test			
p-value	0.271	0.009	0.092
		AED	CED
Gini		0.46	0.57
s.e.		(0.335)	(0.457)
$BIPOL_{NI}$		0.70	0.69
s.e.		(0.102)	(0.560)
Unimodality Test			
p-value		0.006	0.092

Table 2: Gini and $BIPOL_{NI}$ indexes, their standard errors, and tests of unimodality of AD, CD, AED and CED. Counterfactual variable: Initial Productivity.

for EAD versus CED).[22] On the contrary polarization, as measured by the $BIPOL_{NI}$ index proposed by Anderson et al. (2012), would have been lower, even though this difference is not statistically significant according to the reported standard errors (higher values of $BIPOL_{NI}$ index means higher polarization). The decreasing level of polarization is the result of a lower density at the modes (the higher within-cluster dispersion implies less polarization), contrasted by the higher distance between the two modes (the higher between-cluster distance implies more polarization).[23] Polarization in AED and CED is about of the same magnitude, but as the result of an increased within-cluster dispersion contrasted by an increased between-cluster distance.

Overall, although initial productivity contributes to a reduction in inequality, a tendency towards polarization still remains from AD 1960 to CD 2008. In fact, the inverse relationship between initial productivity and the growth rate holds on average, as shown also by the negative and significant coefficient of **y0** in Table 1; but, as Figure 7 shows, *this effect is not constant* across different initial productivity ranges. This result is in line with the remark of Bernard and Durlauf (1996) on the misleading implications of a negative coefficient of initial productivity in growth regressions, and is consistent with the presence of multiple equilibria.

III.D.ii. Employment Growth and Investment Ratio

Beaudry et al. (2005) find that changes in the patterns of accumulation of factors of production, labour and capital, play a very important role in the formation of two peaks in the distribution of productivity. In particular, they find that the change in their

[22] The difference between indexes of the AD and CD is statistically significant.

[23] Heuristically, the $BIPOL_{NI}$ index measures the degree of polarization of a bimodal distribution as the product of two components: i) the sum of the density of two modes (which should measure the so-called degree of "identification" within each of two clusters inversely related to the within-cluster dispersion); and ii) the distance between the two modes (which should measure the so called degree of "alienation" between the two clusters directly related to the between-cluster distance). When the estimated distribution appears to be unimodal, as e. g. in the AD in 2008, $BIPOL_{NI}$ index is not calculable and we report NA in the table. We refer to Appendix J for more technical details on the $BIPOL_{NI}$ index.

marginal effect is important, while the change in the distribution of the variable is not. In this paper we have shown that the employment growth and investment ratio have, respectively, a nonlinear negative effect and a nonlinear positive effect on productivity growth (see Table 1 and Figures 1-2), rejecting the hypothesis that there is a change in the marginal effect of two different linear specifications.

In this section we analyze the distributional effect of these variables on the observed polarization. The conditional mean of MGE of employment growth is statistically different from the unconditional mean for countries with an initial productivity higher than approximately 1.8 (see Figure 10). This is reflected in the CD in 2008 (see Figure 11): if all countries had had the same level of employment growth, the distribution in 2008 would have been unimodal and there would have been less inequality. This tendency also appears in the CED, shown in Figure 12 and Table 3. Hence, the employment growth acts as a force favouring divergence and polarization, in particular by pushing some high-productivity countries further above the mean, a result in contrast with Beaudry et al. (2005) .

As regards the investment ratio Figure 13 shows that its MGE is never statistically significant. However, its estimated partial effect is strongly nonlinear (see Figure 2). As discussed in Section II.C.ii. even when the marginal effect of the *k-th* variable is independent of the initial productivity level (i.e. when the condition in Eq. (1) is fulfilled) the nonlinear impact of the variable implies possible differences between the actual and counterfactual stochastic kernels. Indeed, the distribution in 2008 has been remarkably affected by the investment ratio: if all the countries had had the same level of investment ratio in 2008 there would have been less mass in the high-productivity peak (see Figure 14). This tendency also characterizes the long run (see Figure 15). However, the polarization measured by the $BIPOL_{NI}$ index would have been about the same, as the result of lower mass around the two modes, but higher distance between them.

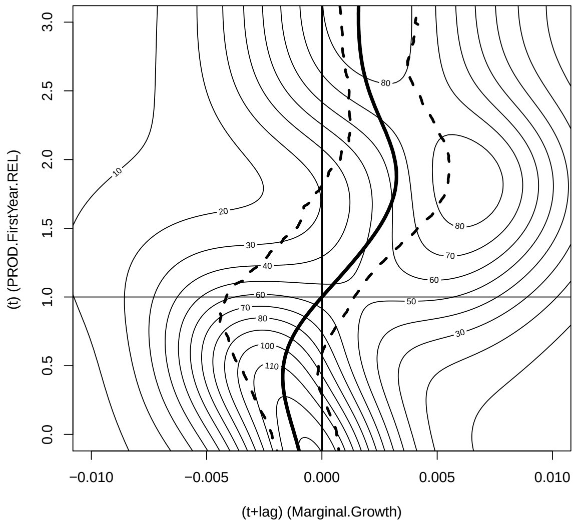

Figure 10: MGE conditioned on the initial level of productivity, estimated mean of MGE conditioned on initial level of productivity (thick solid line), its confidence bands (dashed lines), and unconditional mean (thin solid vertical line). Counterfactual variable: Employment Growth.

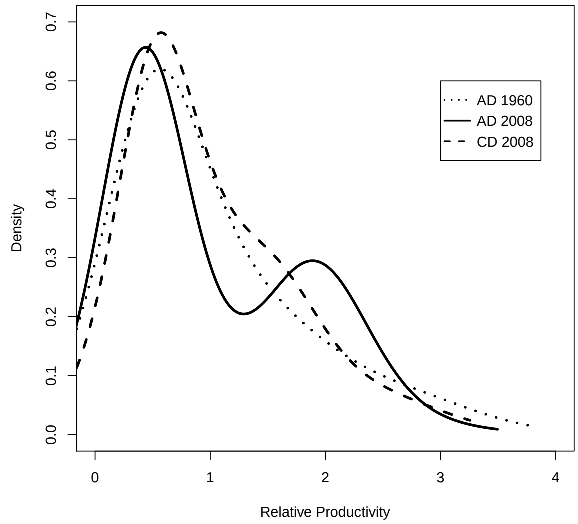

Figure 11: AD in 1960 (dotted line), AD in 2008 (solid line), and CD in 2008 (dashed line). Counterfactual variable: Employment Growth.

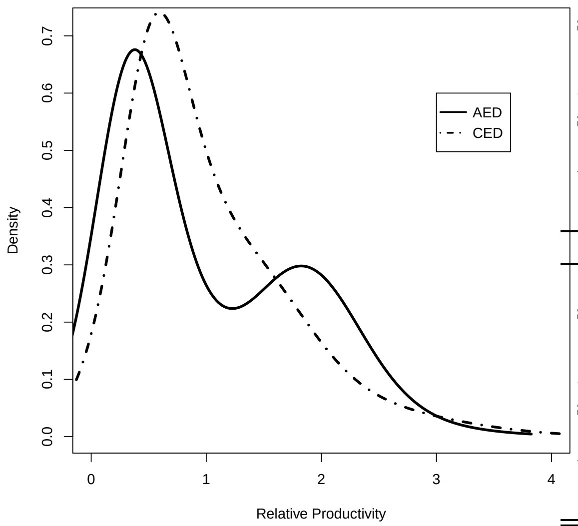

Figure 12: AED (thick line) and CED (thin line). Counterfactual variable: Employment Growth.

	AD 1960	AD 2008	CD 2008
Gini	0.42	0.42	0.38
s.e.	(0.029)	(0.026)	(0.028)
$BIPOL_{NI}$	NA	0.69	NA
s.e.	(-)	(0.087)	(-)
Unimodality Test			
p-value	0.271	0.009	0.310
		AED	CED
Gini		0.46	0.39
s.e.		(0.335)	(0.255)
$BIPOL_{NI}$		0.70	NA
s.e.		(0.102)	(-)
Unimodality Test			
p-value		0.006	0.56

Table 3: Gini and $BIPOL_{NI}$ indexes, their standard errors, and tests of unimodality of AD, CD, AED and CED. Counterfactual variable: Employment Growth.

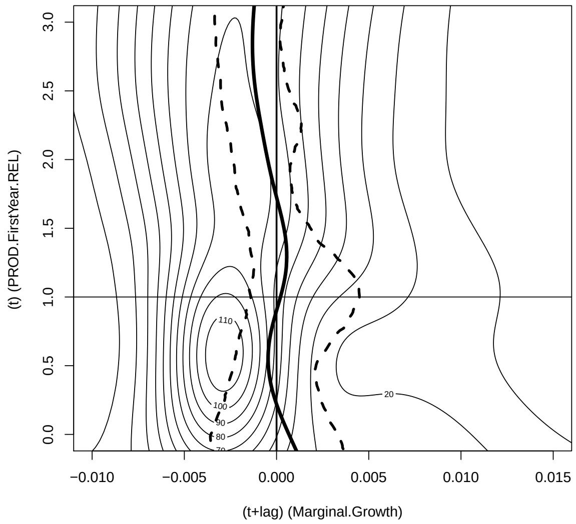

Figure 13: MGE conditioned on the initial level of productivity, estimated mean of MGE conditioned on initial level of productivity (thick solid line), its confidence bands (dashed lines), and unconditional mean (thin solid vertical line). Counterfactual variable: Investment Ratio.

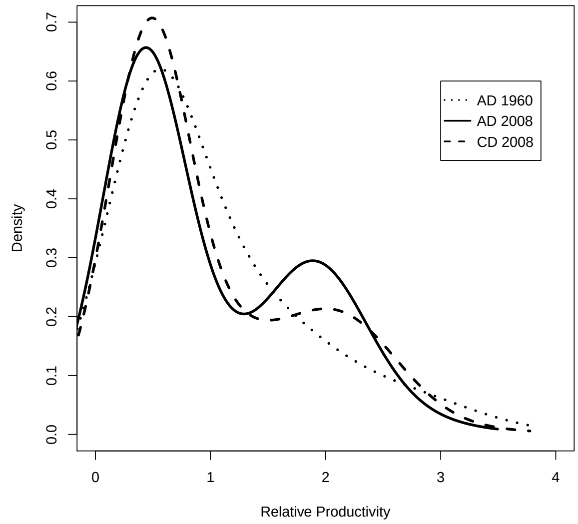

Figure 14: AD in 1960 (dotted line), AD in 2008 (solid line), and CD in 2008 (dashed line). Counterfactual variable: Investment Ratio.

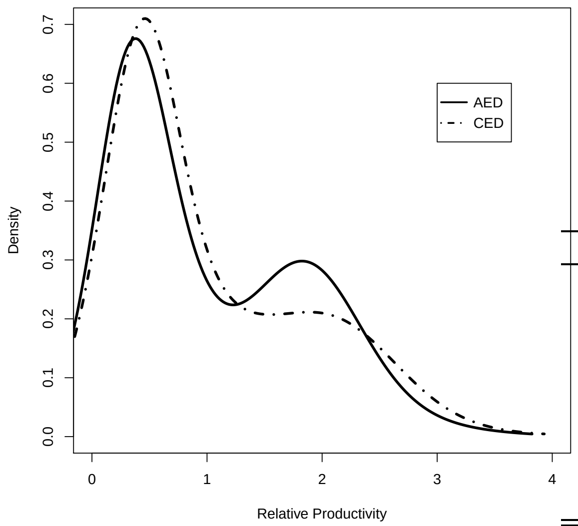

Figure 15: AED (thick line) and CED (thin line). Counterfactual variable: Investment Ratio.

	AD 1960	AD 2008	CD 2008
Gini	0.42	0.42	0.42
s.e.	(0.029)	(0.026)	(0.024)
$BIPOL_{NI}$	NA	0.69	0.69
s.e.	(-)	(0.087)	(0.192)
Unimodality Test			
p-value	0.271	0.009	0.029
		AED	CED
Gini		0.46	0.46
s.e.		(0.335)	(0.374)
$BIPOL_{NI}$		0.69	0.66
s.e.		(0.102)	(0.369)
Unimodality Test			
p-value		0.006	0.054

Table 4: Gini and $BIPOL_{NI}$ indexes, their standard errors, and tests of unimodality of AD, CD, AED and CED. Counterfactual variable: Investment Ratio.

III.D.iii. Education

The conditional mean of MGE of education is statistically different from the unconditional mean (see Figure 16) and, in particular, it is above the unconditional mean for countries with initial productivity above the average, while the opposite holds for countries with below-average initial productivity. Thus, education acts like a force of divergence. This is reflected in the CD in 2008 (see Figure 17): if all countries had had the same level of education, the distribution in 2008 would have been less dispersed and there would have been less inequality (about 6 base points in the Gini index). Moreover, the polarization would have been lower (although the difference does not appear statistically different). The polarization-enhancing role of education also emerges in the comparison between the AED and the CED shown in Figure 18 and Table 5.

These results are in stark contrast with Beaudry et al. (2005) who find that education does not explain the observed polarization.

IV. Concluding Remarks

In this paper we proposed a new method to analyze the factors driving convergence and divergence in the growth dynamics. The proposed methodology combines the growth regression approach, albeit allowing for a semiparametric specification, with the distribution dynamics approach. We applied our methodology to a sample of countries, and showed its potential to shed light on the identified tendency for polarization, by the analysis of the distributional effect of initial conditions, the accumulation of factors, labor and capital and education. In all cases it was possible to obtain information otherwise missed by existing methods of investigation of the determinants of distribution dynamics. The methodology can be further generalized by considering alternative growth models, which include, e.g., endogeneity, thresholds, and interactions terms, as well as panel growth models (see Durlauf et al., 2005).

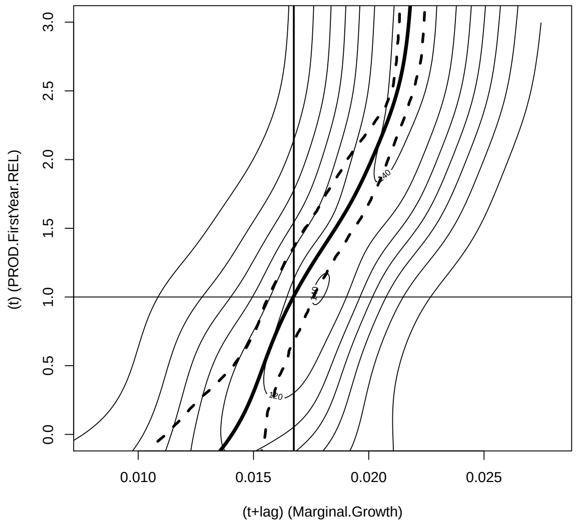

Figure 16: MGE conditioned on the initial level of productivity, estimated mean of MGE conditioned on initial level of productivity (thick solid line), its confidence bands (dashed lines), and unconditional mean (thin solid vertical line). Counterfactual variable: Education.

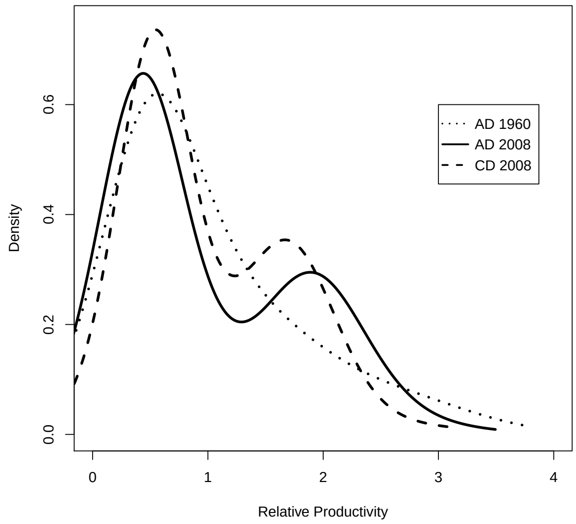

Figure 17: AD in 1960 (dotted line), AD in 2008 (solid line), and CD in 2008 (dashed line). Counterfactual variable: Education.

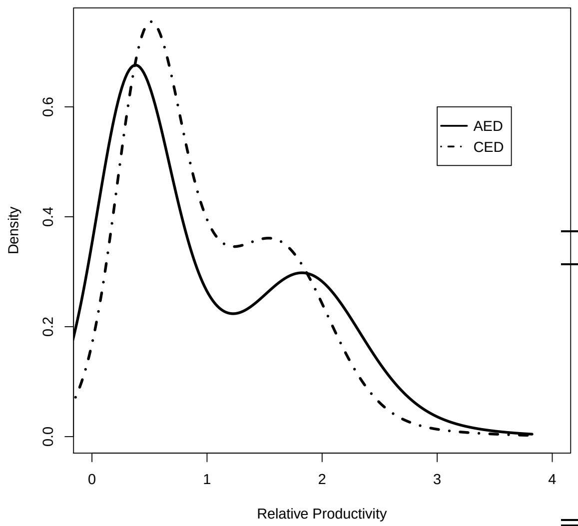

Figure 18: AED (thick line) and CED (thin line). Counterfactual variable: Education.

	AD 1960	AD 2008	CD 2008
Gini	0.42	0.42	0.36
s.e.	(0.029)	(0.026)	(0.023)
$BIPOL_{NI}$	NA	0.69	0.61
s.e.	(-)	(0.087)	(0.078)
Unimodality Test			
p-value	0.271	0.009	0.043
		AED	CED
Gini		0.46	0.38
s.e.		(0.335)	(0.235)
$BIPOL_{NI}$		0.69	0.57
s.e.		(0.102)	(0.139)
Unimodality Test			
p-value		0.006	0.016

Table 5: Gini and $BIPOL_{NI}$ indexes, their standard errors, and tests of unimodality of AD, CD, AED and CED. Counterfactual variable: Education.

Acknowledgements

We wish to thank Daniel Henderson, Thanasis Stengos, Roberto Basile, Stefano Magrini, and Julie Le Gallo for many useful comments on earlier drafts of this paper. We also thanks seminar participants at University of Pisa, Firenze, Palermo, SMYE 2009, ERSA 2010, ICEEE 2011, SNDE 2013, RCEA Workshop 2013. Any remaining error is our own responsibility. D.F. and A.P. has been supported by the Italian Ministry of University and Research as a part of the PRIN 2009 program (grant protocol number 2009H8WPX5).

Affiliation and Address of the Authors

Davide Fiaschi (corresponding author), Dipartimento Economia e Management, University of Pisa, Via Ridolfi 10, 56124 Pisa, Italy. Tel +39 050 2216208, Fax. +39 050 2216384, e-mail: dfiaschi@ec.unipi.it

Andrea Mario Lavezzi, Dipartimento di Scienze Giuridiche, della Società e dello Sport, University of Palermo, e-mail: mario.lavezzi@unipa.it.

Angela Parenti, Dipartimento Economia e Management, University of Pisa, e-mail: aparenti@ec.unipi.it.

A List of Countries in the Sample

Our sample of 73 countries consists of:[24]
Argentina, Australia, Austria, Belgium, Bangladesh, Bolivia, Brazil, Barbados*, Botswana*, Canada, Switzerland, Chile, China, Colombia, Costa Rica, Cyprus, Denmark, Dominican Republic, Ecuador, Egypt, Spain, Finland, Fiji, France, United Kingdom, Greece, Guatemala, Hong Kong*, Honduras, Indonesia, India, Ireland, Iran, Iceland*, Israel*, Italy, Jamaica, Jordan, Japan, Republic of Korea, Sri Lanka*, Lesotho*, Luxembourg*, Morocco, Mexico, Mozambique, Malaysia, Namibia*, Nicaragua, Netherlands, Norway, Nepal, New Zealand, Pakistan*, Panama, Peru, Philippines, Papua New Guinea*, Portugal, Paraguay, Romania, Singapore, El Salvador, Sweden, Syria,

[24]With respect to the Beaudry et al. (2005)'s sample, we have no data for Guyana and Tunisia. For countries marked with * we have no data on education.

Thailand, Trinidad&Tobago, Turkey, Taiwan*, Uruguay, United States, Venezuela, South Africa.

B Description of the Variables and Descriptive Statistics

The variables used in the estimation of the growth model in Eq. (3) are the following:[25]

- **y0** is the (log) of the real GDP chain per worker (*rgdpwok* in PWT 7.1).
- g is the corresponding annualized average growth rate of **y0**.
- **n** is the (log) growth rate of employment, where *workers* are computed as the population from 15 to 64 obtained from:

  $$workers = rgdpch/rgdpwok * pop;$$

  where *rgdpch* is the real GDP chain per capita and *pop* is the population in PWT 7.1.
- **i/y** is the (log) investment ratio at constant price and corresponds to the variable *ki* in PWT 7.1 divided by 100.
- **Edu** is the (log) average years of schooling and corresponds to the *TY15* in Cohen and Soto (2007) ("Years of schooling of population 15 and over, whether studying or not").

	g	**y0**	**n**	**i/y**	**Edu**
Mean	0.02	9.17	-2.66	-1.48	1.82
Stand. Dev.	0.01	1.01	0.15	0.27	0.47

Table 6: Mean and Standard Deviation of variables used in the growth regressions

[25] The choice follows Beaudry et al. (2005, p. 21)

	g	**y0**	**n**	**i/y**	**Edu**
g	1.00				
y0	-0.33	1.00			
n	-0.38	0.28	1.00		
i/y	0.45	0.16	-0.08	1.00	
Edu	0.07	0.76	-0.44	0.35	1.00

Table 7: Correlations among the variables used in the growth regressions

It can be observed that there is a high correlation between the education variable and initial income.

C GAM estimation

The estimation of Eq. (3) is obtained by penalized likelihood maximization (see Wood, 2011, for details). The model is fitted by minimizing:

$$||\mathbf{y} - \mathbf{X}\beta||^2 + \sum_{k=1}^{K} \lambda_k \int_0^1 \left[\mu_k''(x)\right]^2 dx, \tag{15}$$

where $\mathbf{y}$ is the vector of observations (g_i in our case), $\mathbf{X}$ is the matrix of explanatory variables, β is a vector of parameters to be estimated, λ_k, $k = (1, ..., K)$, are smoothing parameters, and the penalty, which controls the smoothness of the estimate, is represented by the integrated square of second derivatives of the smooth terms. The vector of parameters β originates from expressing every smooth term in Eq. (3), $\mu_j(.)$, as:

$$\mu_j(X_{i,j}) = \sum_{l=1}^{q} b_l \left(X_{i,j}\right) \beta_l \tag{16}$$

where $b_l\,(x)$ are *basis functions* and q is their number.

Parameters β_{l} are chosen to minimize the function in Eq. (15) for given values of the smoothing parameters λ_k (it is possible to show that the penalty can also be expressed as a function of β). Smoothing parameters are in turn chosen by the minimization of the restricted maximum likelihood (REML) score. Estimation proceeds by penalized iteratively re-weighted least squares (P-IRLS), until convergence in the estimates is reached.

For each estimation we report: 1) the estimated degrees of freedom for each term (EDF). The EDFs reflect the flexibility of the model: when the EDFs of a term are equal to one, the smooth term can be substituted by a linear function; 2) the value of AICc, that is the corrected Akaike information criterion; 3) the proportion of deviance explained, a generalization of R^2; 4) the REML score, which provides the fundamental information on the specification of the model; 5) the scale parameter, corresponding to the residual variance of the estimation; and, 6) the number of observations.

D Bootstrap Procedure to Test the Stability of Semiparametric Regressions

The bootstrap procedure is as follows:

1. Pool the sample of the two subperiods.

2. Extract from this pooled sample two samples of the same size of the two original samples.

3. Run a semiparametric regression for each sample and take the differences in the estimated partial effect of each variable.

4. Repeat B=1000 times points 1-3.

5. Report for each variable the confidence band at 95% confidence level of the calculated differences and check that the observed differences belong to these confidence bands.

Figures 19-21 report the results of this procedure.

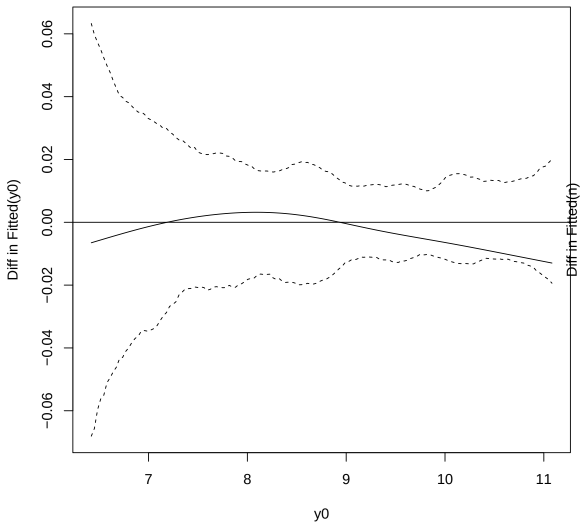

Figure 19: Stability test for **y0**. Thick line: observed difference; dotted lines: 95% confidence bands.

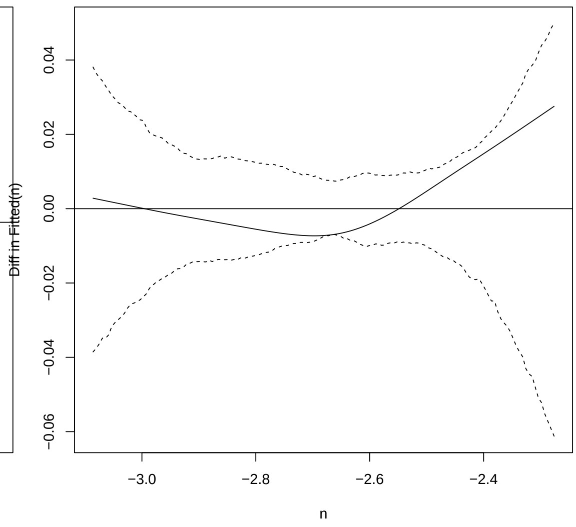

Figure 20: Stability test for **n**. Thick line: observed difference; dotted lines: 95% confidence bands.

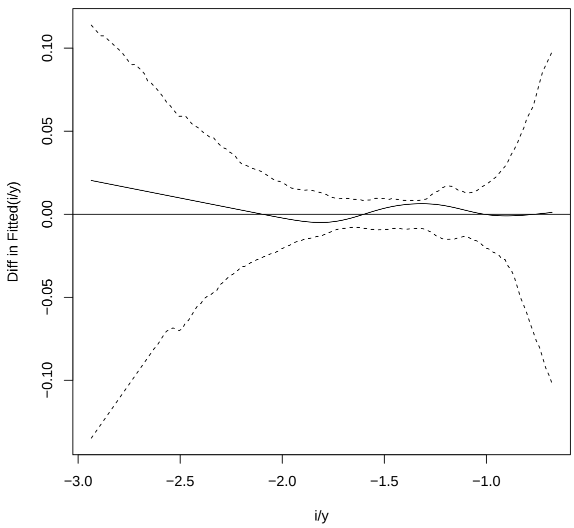

Figure 21: Stability test for **i/y**. Thick line: observed difference; dotted lines: 95% confidence bands.

E Endogeneity Test

The Control Function method (CFM) is used to perform the endogeneity test (see, e.g. Ng and Pinkse, 1995; Blundell and Powell, 2003). The CFM treats endogeneity as an omitted variable problem, where the inclusion of estimated first-stage residuals as a covariate corrects the inconsistency of the regression of the dependent variable on the endogenous explanatory variable. This method provides consistent estimation of the underlying regression coefficients. Therefore, according to CFM we use a two-stage procedure: i) first we run a semiparametric regression of each endogenous variable on the exogenous variables and the instruments; then, ii) we insert the first-stage residuals in the original semiparametric regression.

Since the second-stage regression contains generated regressors (i.e. the first-stage residuals), to obtain the appropriate standard errors we use the following bootstrap procedure. Given a sample of observations $(\mathbf{y}, \mathbf{X}, \mathbf{Z})$, where $\mathbf{y}$ is the vector of dimension N of dependent variable, $\mathbf{X}$ is the $N \times K$ matrix of explanatory variables (including the endogenous variables), and $\mathbf{Z}$ is the $N \times K$ matrix of instruments:

1. Select a bootstrap sample $(\mathbf{y}^*_\mathbf{b}, \mathbf{X}^*_\mathbf{b}, \mathbf{Z}^*_\mathbf{b})$ drawn with replacement from $(\mathbf{y}, \mathbf{X}, \mathbf{Z})$.

2. Run a semiparametric regression of each endogenous variable on the exogenous variables and the instruments.

3. Insert the first-stage residuals in the original semiparametric regression.

4. Repeat $B = 1000$ times points 1-3.

5. For each estimated parametric coefficients compute the corresponding equal-tail bootstrap p-value (see Davidson and MacKinnon (2007)):

$$P^*(\hat{\beta}) = 2 * \min \left(\frac{1}{B} \sum_{b=1}^{B} \#\{\hat{\beta}^*_b \leq 0\}, \frac{1}{B} \sum_{b=1}^{B} \#\{\hat{\beta}^*_b > 0\} \right) \quad (17)$$

6. For each estimated non-parametric coefficients compute the average partial effect ad the 95% confidence bands.

We use the following instruments:

- for **n**: the augmented growth rate of employment in 1960 (**n.1960**);
- for **i/y**: the investment ratio in 1960 (**i/y.1960**);
- for **Edu**: the number of years of schooling in 1960 (**Edu.1960**).

Results of the first-stage regressions reported in Table 8 show that almost all the instruments are significant. In the second-stage regression (Model "Including v" in Table 9) all the coefficients of the first-stage residuals are not statistically significant. Figures 22-23 graph the estimated additive component functions derived from the semiparametric estimation with (Including v) and without (Excluding v) controlling for endogeneity. We conclude that all variables are exogenous because no significant effect emerges from correcting for endogeneity.

Dep. Var:	**n**	**i/y**	**Edu**
	GAM **1978-2008**	GAM **1960-2008**	GAM **1960-2008**
Parametric coefficients:	Estimate	Estimate	Estimate
const	-2.9706***	-1.4808***	1.6499***
Non parametric coefficients:	EDF	EDF	EDF
y0	(0.0334**)	3.376*	(0.0343)
n.1960	5.468***	2.651**	(0.0521)
i/y.1960	1.285	1.055***	2.035
Edu.1960	3.485***	1.734	1.909***
AICc	-144.92	-17.33	-76.96
Dev. Expl.	0.85	0.61	0.94
REML score	-57.85	-1.18	0.0145
Scale est.	0.0038	0.0338	0.0138
Obs.	61	61	61

Table 8: First-stage regressions of potentially endogenous variables. Significance codes: 0.01"***" 0.05"**" 0.1"*". Terms in parenthesis enter linearly in the specification.

	Including v	**Excluding v**
Dep. Var: g	**GAM with CF 1960-2008**	**GAM 1960-2008**
Parametric coefficients:	Estimate	Estimate
const	0.0794	0.0796***
	(0.000)	
y0	-0.0084	-0.0083***
	(0.002)	
Edu	0.0092	0.0091**
	(0.038)	
Non parametric coefficients:	EDF	EDF
n	3.069	2.565***
i/y	4.711	4.016***
First-stage residuals coefficients:	Estimate	Estimate
n_res	-0.0028	
	(0.918)	
i/y_res	0.0097	
	(0.472)	
Edu_res	0.0097	
	(0.580)	

Table 9: Second-stage regression (Including v) and regression without control for endogeneity (Excluding v). Significance codes: 0.01"***" 0.05"**" 0.1"*". P-values form bootstrap in parenthesis.

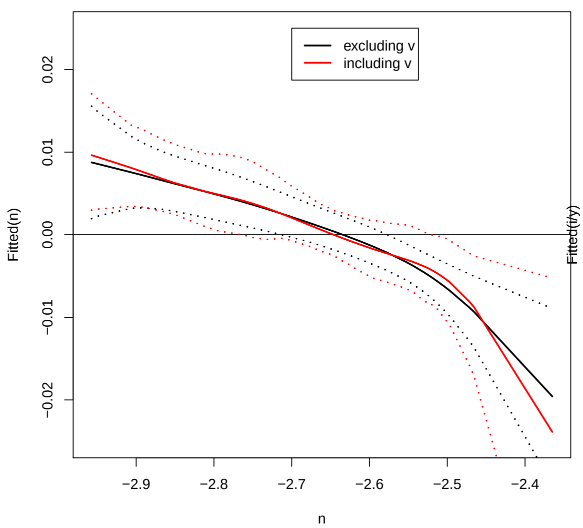

Figure 22: Second-stage regression for **n**; dotted lines: 95% confidence bands.

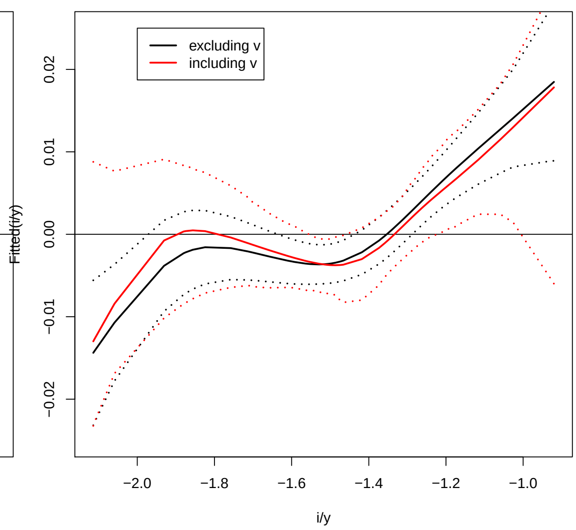

Figure 23: Second-stage regression for **i/y**; dotted lines: 95% confidence bands.

F Adaptive Kernel Estimation

When observations vary in sparseness over the support of the distribution, the adaptive kernel estimation is a two-stage procedure which mitigates the drawbacks of a fixed bandwidth in density estimation (see Silverman, 1986, p. 101). In general, given a multivariate data set $\boldsymbol{X} = \{\boldsymbol{X}_1, ..., \boldsymbol{X}_N\}$ and a vector of sample weights $\boldsymbol{W} = \{\omega_1, ..., \omega_N\}$, where $\boldsymbol{X}_i$ is a vector of dimension d and $\sum_{i=1}^{N} \omega_i = 1$, we first run the pilot estimate:

$$\tilde{f}(\boldsymbol{x}) = \frac{1}{N \det(\boldsymbol{H})} \sum_{i=1}^{N} \omega_i k \left\{ \boldsymbol{H}^{-1} (\boldsymbol{x} - \boldsymbol{X}_i) \right\}, \tag{18}$$

where $k(\boldsymbol{u}) = (2\pi)^{-1} exp\left(-\frac{1}{2}\boldsymbol{u}^2\right)$ is a Gaussian kernel and *bandwidth matrix* $\boldsymbol{H}$ is a diagonal matrix $(d \times d)$ with diagonal elements $(h_1, ..., h_d)$ given by the optimal normal bandwidths, i.e. $h_i = [4/(d+2)]^{1/(d+4)} \hat{\sigma}_i N^{-1/(d+4)}$; $\hat{\sigma}_i$ is the estimated standard error of the distribution of $\boldsymbol{X_i}$. The use of a diagonal bandwidth matrix instead of a full covariance matrix follows the suggestions in Wand and Jones (1993). In the case of $d = 1$ we have $\boldsymbol{H} = \det(\boldsymbol{H}) = (4/3)^{1/5} N^{-1/5} \hat{\sigma}$. We then define local bandwidth factors λ_i by:

$$\lambda_i = \left[\tilde{f}(\boldsymbol{X}_i) / g \right]^{-\alpha}, \tag{19}$$

where $\log(g) = \sum_{i=1}^{N} \omega_i \log\left(\tilde{f}(\boldsymbol{X}_i)\right)$ and $\alpha \in [0, 1]$ is a sensitivity parameter. We set $\alpha = 1/2$ as suggested by Silverman (1986, p. 103). Finally the adaptive kernel estimate $\hat{f}(x)$ is defined as:

$$\hat{f}(\boldsymbol{x}) = \frac{1}{N \det(\boldsymbol{H})} \sum_{i=1}^{N} \lambda_i^{-d} \omega_i k \left\{ \lambda_i^{-1} \boldsymbol{H}^{-1} (\boldsymbol{x} - \boldsymbol{X}_i) \right\}. \tag{20}$$

The Gaussian kernel guarantees that the number of modes is a decreasing function of the bandwidth; this property is at the basis of the test for unimodality (see Silverman, 1986, p. 139).

G Multimodality Test

The multimodality test for all estimate distributions follows the bootstrap procedure described in Silverman (1986, p. 146). Given a data set $\boldsymbol{X} = \{x_1, \ldots, x_N\}$ and a vector of sample weights $\boldsymbol{W} = \{\omega_1, \ldots, \omega_N\}$, we calculate the smallest value of bandwidth, $\hat{h}_0$, for which the estimated distribution is unimodal and the corresponding local bandwidth factors $\boldsymbol{\Lambda} = \lambda_1, \ldots, \lambda_N$. We then perform a *smoothed bootstrap* from the estimated density of observed data set. Since we use the Gaussian kernel, it amounts to: i) draw (with replacement) a vector $I = \{i_1, \ldots, i_N\}$ of size N from $\{1, \ldots, N\}$, given the sample weights $\boldsymbol{W}$; ii) define $Y = \{x_{i_1}, \ldots, x_{i_N}\}$ and $W^* = \{\omega_{i_1}, \ldots, \omega_{i_N}\}$, calculate

$$x_j^* = \bar{\boldsymbol{Y}} + \left(1 + \left(\hat{h}_0 \lambda_{i_j}\right)^2 / \hat{\sigma}_{\boldsymbol{Y}}^2\right)^{-\frac{1}{2}} \left(y_j - \bar{\boldsymbol{Y}} + \hat{h}_0 \lambda_{i_j} \epsilon_j\right), \quad j = 1, \ldots, N; \tag{21}$$

where $\bar{\boldsymbol{Y}}$ and $\hat{\sigma}_{\boldsymbol{Y}}^2$ are respectively the mean and the estimate variance of sample $\boldsymbol{Y}$ and ϵ_j are standard normal random variables; iii) find the minimum value of bandwidth, $\hat{h}_1^*$, for which the estimated density of $\boldsymbol{X}^*$ is unimodal; iv) repeat point i)-iii) B times in order to obtain a vector of critical values of bandwidth $\left\{\hat{h}_1^*, \ldots, \hat{h}_B^*\right\}$. Finally, p-value of null-hypothesis of unimodality is given by $\# \left\{\hat{h}_b^* \geq \hat{h}_0\right\} / B$. To test bimodality, point iii) has to be modified accordingly. We set $B = 1000$.

The multimodality test for the ergodic distribution follows the same logic, but at point i) vector I is drawn from $\{1, \ldots, n_{TR}\}$, where n_{TR} is the total number of observed transitions from t to $t+\tau$ (i.e.$n_{TR} = Nx\,(T - \tau)$); ii) x_j^* smoothed bootstrap transitions are calculated via Eq. (21) (note that all the variables are vectors of two elements); iii) $\hat{h}_1^*$ is the minimum value of bandwidth for which the ergodic distribution calculated from the estimated stochastic kernel relative to x_j^* is unimodal. In order to facilitate the calculation we consider the same bandwidth for actual and future observations (their difference is generally negligible).

H The Estimate of Ergodic Distribution

The ergodic distribution solves:

$$f_\infty(x) = \int_0^\infty g_\tau(x|z) f_\infty(z)\, dz, \tag{22}$$

where x and z are two levels of the variable, $g_\tau(x|z)$ is the density of x, given z, τ periods ahead, under the constraint

$$\int_0^\infty f_\infty(x)\, dx = 1. \tag{23}$$

Since in our estimates all variables are normalized with respect to their average, the ergodic distribution, moreover, must respect the additional constraint:

$$\int_0^\infty f_\infty(x)\, x dx = 1. \tag{24}$$

Following the methodology proposed by Johnson (2005) we first estimate the distribution $\tilde{f}_\infty(x)$, which satisfies Constraints 22 and 23, but not Constraint 24. We then calculate $f_\infty(x) = \tilde{\mu}_x \tilde{f}_\infty(x)$, where $\tilde{\mu}_x = \int_0^\infty \tilde{f}_\infty(x)\, x dx$, which will satisfy all Constraints 22, 23 and 24. In particular, Theorems 11 and 13 in (Mood et al., 1974, pp. 200 and 205) prove that if $\tilde{f}_\infty(x)$ satisfies Constraints 22 and 23 then $f_\infty(x)$ satisfies Constraints 22, 23 and 24. In fact, $g_\tau(z|x) = f_{z,x}(z,x)/f_x(x)$ and $f_{y,q}(y,q) = \mu_z \mu_x f_{z,x}(z,x)$, where $y = z/\mu_z$ and $q = x/\mu_x$. In all computations we set $\tau = 49$.

I Bootstrap Procedure to Compute Confidence Intervals

The bootstrap procedure used to calculate the confidence bands for the estimated median of the stochastic kernels and ergodic distributions is respectively based on the procedure in Bowman and Azzalini (1997, p. 44) and in Fiaschi and Romanelli (2009).

Given a sample of observations $\mathbf{Y} = \{\mathbf{Y}_1, ..., \mathbf{Y}_m\}$ where $\mathbf{Y}_i$ is a vector of dimension N, the bootstrap algorithm consists of three steps.

1. Estimate from sample $\mathbf{Y}$ the stochastic kernel, the median of the stochastic kernel and the corresponding ergodic distribution $\hat{\psi}$.

2. Select B independent bootstrap samples $\{\mathbf{Y}_1^*, ..., \mathbf{Y}_B^*\}$, each consisting of N data values drawn with replacement from $\mathbf{Y}$.

3. Estimate the the stochastic kernel, the median of stochastic kernel and the corresponding ergodic distribution $\hat{\psi}_b^*$ corresponding to each bootstrap sample $b = 1, ..., B$.

4. Use the distribution of $\hat{\psi}_b^*$ about $\hat{\psi}$ to mimic the distribution of $\hat{\psi}$ about ψ.

We set B=500 and in each bootstrap the bandwidth is set equal to the one calculated for the estimation of the density of the observed sample $\mathbf{Y}$.

J $BIPOL_{NI}$ *Index of Polarization*

We measure the polarization of a distributions by $BIPOL_{NI}$ index proposed by Anderson et al. (2012).

In particular, under the assumption that the observed bimodality is the result of the mixture of two unknown sub-distributions (corresponding to two subgroups as, e.g., rich and poor countries) *and* the observed estimated distribution is bimodal, the estimate of $BIPOL_{NI}$ is given by (see Eq. (6) in Anderson et al., 2012):

$$\widehat{BIPOL}_{NI} = \frac{1}{2}\left[\hat{f}\left(\hat{x}_{m,p}\right) + \hat{f}\left(\hat{x}_{m,r}\right)\right]|\hat{x}_{m,p} - \hat{x}_{m,r}|; \qquad (25)$$

where $\hat{x}_{m,p}$ and $\hat{x}_{m,r}$ are the estimated modes, and $\hat{f}(\hat{x}_{m,p})$ and $\hat{f}(\hat{x}_{m,r})$ are the values of the estimated density at the modal points. When $x_{m,r} \neq x_{m,p}$ and $\hat{f}(.)$ is estimated by a Gaussian kernel, it is possible to show that:[26]

$$\frac{\widehat{BIPOL}_{NI} - BIPOL_{NI}}{(nh^3)^{-1/2}} \rightarrow^D \mathcal{N}\left(Bias,\right.$$
$$\left.\frac{1}{4}\left[f\left(x_{m,p}\right)+f\left(x_{m,r}\right)\right]^2\left\{\frac{f\left(x_{m,p}\right)}{\left[f''\left(x_{m,p}\right)\right]^2}+\frac{f\left(x_{m,r}\right)}{\left[f''\left(x_{m,r}\right)\right]^2}\right\}\|K'\|_2^2\right), \quad (26)$$

where N is the number of observations, h the bandwidth used in the kernel, $|K'\|_2^2$ the L^2 norm of the first derivative of the Gaussian Kernel (equal to $\frac{1}{4\sqrt{\pi}}$), and $Bias$ reflects the bias in the estimate of $f(x_i)$ deriving from the use of kernel.

Instead, inference on $BIPOL_{NI}$ index runs into troubles when the modes of the two unknown distributions are so close to each other that the mixture of the two distributions appears to be unimodal; indeed this does not necessarily imply that polarization is absent, i.e. $BIPOL_{NI} = 0$, but only that it is not possible to calculate the index from the estimated observed distribution.

Therefore in the empirical application for the cases where the estimated distribution appears to be bimodal we report the value of $\widehat{BIPOL}_{NI}$ index and of its standard error calculated according to Eqq. (25) and (26); while for the case where the estimated distribution appears to be unimodal the $\widehat{BIPOL}_{NI}$ index is not calculated (we report NA).

References

G. Anderson, O. Linton, and T.W. Leo. A polarization-cohesion perspective on cross-country convergence. *Journal of Economic Growth*, 17:49–69, 2012.

R. J. Barro and X. Sala-i Martin. *Economic Growth*. Cambridge: MIT press, 2004.

R.J. Barro. Eonomic Growth in a Cross-section of Country. *Quarterly Journal of Economics*, 106(2):407–443, 1991.

[26] See Appendix A.1 in Anderson et al. (2012).

R.J. Barro and J-W. Lee. International comparisons of educational attainment. *Journal of Monetary Economics*, 32(3):363–394, 1993.

P. Beaudry, F. Collard, and D.A. Green. Changes in the world distribution of output per worker, 1960-1998: how a standard decomposition tells an unorthodox story. *Review of Economics and Statistics*, 87(4):741–753, 2005.

A.B. Bernard and S. N. Durlauf. Interpreting tests of the convergence hypothesis. *Journal of Econometrics*, 71(1-2):161–173, 1996.

R. Blundell and L. J. Powell. *Advance in Economics and Econometrics*, chapter Endogeneity in nonparametric and Semiparametric Regression Models, pages 312–357. Cambridge University Press, 2003.

A.W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations.* Oxford University Press, Oxford, 1997.

P. Cheshire and S. Magrini. Analizing growth distribibition dynamics: isolating divergence factors. *Paper preseted at* $45^t h$ *European Regional Science Association Congress*, Cambridge(Amsterdam), 2005.

D. Cohen and M. Soto. Growth and human capital: good data, good results. *Journal of Economic Growth*, 12(1):51–76, 2007.

R. Davidson and J. G. MacKinnon. Improving the reliability of bootstrap tests with the fast double bootstrap. *Computational Statistics & Data Analysis*, 51(7):3259–3281, 2007.

A. De La Fluente. Convergence equations and income dynamics: The sources of oecd convergence, 1970-1995. *Economica*, 70(280): 655–671, 2003.

J. DiNardo, N. Fortin, and T. Lemieux. Labor market institutions and the distribution of wages, 19731992: a semi parametric approach. *Econometrica*, 64:10011044, 1996.

S.N. Durlauf, A. Kourtellos, and A. Minkin. The local solow growth model. *European Economic Review*, 45:928–940, 2001.

S.N. Durlauf, P. Johnson, and J. Temple. Growth econometrics. In S. N. Durlauf and P. Aghion, editors, *Handbook of Economic Growth*. Elsevier, 2005.

James D Feyrer. Convergence by parts. *The BE Journal of Macroeconomics*, 8(1), 2008.

D. Fiaschi and A. M. Lavezzi. Productivity polarization and sectoral dynamics in european regions. *Journal of Macroeconomics*, 29: 612–637, 2007.

D. Fiaschi and M. Romanelli. Nonlinear dynamics in welfare and the evolution of world inequality. *Discussion Papers del Dipartimento di Scienze Economiche - Università di Pisa*, (n. 81), 2009. http://www-dse.ec.unipi.it/pubblicazioni/discussion_papers.asp.

O. Galor. Multiple growth regimes - insights from unified growth theory. *Journal of Macroeconomics*, 29(3):470–475, 2007.

P. Hall and M. York. On the calibration of silverman's test for multimodality. *Statistica Sinica*, 11:515–536, 2001.

W. Härdle, M. Müller, S. Sperlich, and A. Werwatz. *Nonparametric and Semiparametric Models*. Springer, Berlin, 2004.

D. J. Henderson and R. R. Russell. Human capital and convergence: A production-frontier approach. *International Economic Review*, 46(4):1167–1205, 2005.

D. J. Henderson, C. Parmeter, and R. R. Russell. Modes, weighted modes, and calibrared modes, evidence of clustering using modality tests. *Journal of Applied Econometrics*, 23:607–638, 2008.

P.A. Johnson. A continuous state space approach to convergence by parts. *Economic Letters*, 86:317–321, 2005.

Z. Liu and T. Stengos. Non-linearities in cross-country growth regressions: a semiparametric approach. *Journal of Applied Econometrics*, 14(5):527–538, 1999.

J. Machado and J. Mata. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, 20:445–465, 2005.

N.G. Mankiw, D. Romer, and D. Weil. A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, 107: 407–437, 1992.

A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, New York, 1974.

S. Ng and J. Pinkse. *Nonparametric Two-step Estimation of Unknown Regression Functions when the Regressors and the Regression Error are Not Independent*. Universit de Montral, Dp. de sciences conomiques, Montral, 1995.

D. Quah. Empirical cross-section dynamics in economic growth. *European Economic Review*, 37(2-3):426–434, 1993.

D. Quah. Twin Peaks: Growth and Convergence in Models of Distribution Dynamics. *Economic Journal*, 106(437):1045–55, 1996a.

D. Quah. Convergence Empirics Across Economies with (Some) Capital Mobility. *Journal of Economic Growth*, 1(1):95–124, 1996b.

D. Quah. Empirics for Growth and Distribution: Stratification, Polarization, and Convergence Clubs. *Journal of Economic Growth*, 2(1):27–59, 1997.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. URL `http://www.R-project.org/`. ISBN 3-900051-07-0.

B.W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London, 1986.

M.P. Wand and M.C. Jones. Comparison of smoothing parametrizations in bivariate kernel density estimation. *Journal of the American Statistical Association*, 88:520–528, 1993.

S. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, 1(73):3–36, 2011.

S. N. Wood. *Generalized Additive Models. An Introduction with R.* Chapman and Hall, London, 2006.

J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data.* MIT press, 2002.