On the Determinants of Distribution Dynamics: a New Method and an Application to a Cross-Section of Countries

(Preliminary version. Please do not quote without permission)

Davide Fiaschi, Andrea Mario Lavezzi and Angela Parenti*

December 11, 2012

Abstract

In this paper we provide a novel approach to identify the effect of growth determinants on the distribution dynamics, that integrates the counterfactual analysis of, e.g., Beaudry et al. (2005) with the estimation of conditioned stochastic kernels of Quah (1996b) and Quah (1997). The counterfactuals are constructed from a nonparametric growth regression in which the cross-section heterogeneity in the variable of interest is removed. The methodology also allows for measuring the marginal effect of individual variables on the distribution of per capita income (labor productivity), and to test for the possible presence of distributional effects in the residuals of growth regression, as a means to assess the goodness of fit of the initial growth regression. The methodology is applied to the analysis of productivity dynamics across a sample of countries. We show its capacity to highlight aspects of the identified tendency for polarization, otherwise missed by existing methods.

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

JEL: C14; C21; O40; O50

*Davide Fiaschi, Dipartimento Economia e Management, University of Pisa, e-mail: dfiaschi@ec.unipi.it; Andrea Mario Lavezzi, Dipartimento di Studi su Politica, Diritto e Società, University of Palermo, e-mail: mario.lavezzi@unipa.it; Angela Parenti, Dipartimento Economia e Management, University of Pisa, e-mail: aparenti@ec.unipi.it.
1 Introduction

The world income distribution has been largely studied over the last two decades and a new stylized fact appears: the distribution of per capita income has moved from a unimodal shape in the 1960s to a twin-peaks 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), and Basile (2009)). However, it is still unclear whether these twin-peaks are a persistent phenomenon (see Galor (2007)) and which factors drive the formation of the two peaks.

This paper aims at analysing the factors driving convergence and divergence processes in the growth dynamics. To this purpose, we propose a methodology to measure the distributional effect of individual growth determinants, which combines a semiparametric growth regression approach with the approach based on estimation of stochastic kernels, i.e., of the operators that map current distributions into future distributions of income or productivity. Specifically, we exploit the idea of performing a counterfactual analysis (see, e.g., Beaudry et al. (2005)) to evaluate the distributional impact of a given variable. That is, we estimate and compare actual and counterfactual distributions to estimate short-run effects, and the implied actual and counterfactual ergodic distributions to identify long-run tendencies.

In addition, our methodology allows: i) to measure the marginal effect of a variable of interest on the distribution, which provides information on the direction of the effect of the variable in different ranges of per capita income (or labour productivity) distribution, and ii) to test for the possible presence of distributional effects of the residuals of the growth regression, which we will utilize to assess the specification of the regression model. The advantage of studying the distributional effect of a growth determinant resides in the possibility to identify whether the same factor favors convergence in some range of the per capita income (productivity) distribution and divergence in another, an aspect which cannot be captured by the estimation of a single parameter as in standard growth regressions.

The paper is organized as follows: Section 2 describes the methodology for the empirical analysis and clarifies the relations with existing approaches; Section 3 presents an empirical application to European countries; Section 4 concludes. The appendices contain details on data and on the methodology.

2 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

\footnote{See, e.g., Quah (1997) for details.}
a semiparametric growth regression, and then utilize the results to estimate counterfactual distributions with respect to individual variables of interest; this allows to identify their contribution to convergence or divergence, that we denote as marginal growth effect.

2.1 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 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-sectional distribution evolves over time.

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)), which generally found 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.

De La Fluente (2003), in the spirit of the present paper, extends the GRA approach by decomposing the measures of $\sigma$ and $\beta$-convergence (Barro and Sala-i Martin, 2004) into sums of partial $\sigma$ and $\beta$-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 not being able to capture phenomena such as mobility, stratification and polarization in the world income distribution. On the contrary, operators such stochastic kernels (or transition matrices) may reveal information on these aspects of the growth process. A further step, aiming at evaluating the effects of individual explanatory variables on the distribution dynamics is taken by Quah (1996b) and Quah (1997), by introducing 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) introduces 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, p. 47), where weights are calculated with respect to a variable suspected to affect the income dynamics.

Another strand of literature proposes counterfactual analysis as an alternative methodology to

---

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

3 In addition to these types of criticism, Bernard and Durlauf (1996) show that a negative sign of the coefficient of initial income in a growth regression does not necessarily imply absolute or conditional convergence, as the data-generating process may be characterized by multiple, locally stable, equilibria.
identify the impact of individual explanatory variables on distributions (see e.g. DiNardo et al. (1996) and Machado and Mata (2005)). In particular, 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 over two periods, 1960-1978 and 1978-1998, by estimating linear growth regressions. They build counterfactual distributions for the second period by assuming that the variable 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.

Our methodology is close to that of 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 countries 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 the “simulated” stochastic kernel (computed on the basis of alternative values of the explanatory variables in the growth regression).

However, we differ from the current literature in two crucial aspects: i) the estimate of “conditioned” stochastic kernels (denoted counterfactual stochastic kernels), and ii) of counterfactual distributions. A further difference is that the preliminary step will be based on a semiparametric growth regression, to take into account the presence of nonlinearities which, as many recent studies show, strongly characterize economic growth (see Durlauf et al. (2005), for discussion and references).

In the following we detail our methodology, which is based on six steps: i) estimation of a semiparametric growth regression model (Section 2.2); ii) calculation of counterfactual productivity (Section 2.3.1); iv) estimation of counterfactual stochastic kernels (Section 2.3.1); v) estimation of counterfactual ergodic distributions (Section 2.3.1); vi) evaluation of the distributional effects of a variable and estimation of its marginal growth effect (Section 2.3.2); ii) test on the distributional effects of growth residuals (Section 2.4).

2.2 Modeling Productivity Growth

Assume there exist N countries, and define by \( y_i(t) \) labour productivity of region \( i \) at time \( t \). Labour productivity of region \( i \) at time \( T > 0 \), therefore, can be expressed as:

\[
y_i(T) = y_i(0)e^{g_iT},
\]

where \( g_i \) is the annual rate of growth of productivity in region \( i \), between periods 0 and \( T \).

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

\[
g_i = \varphi(X_i, v_i).
\]
Differently from other approaches to counterfactual analysis, we model the growth rate $g_i$ by a semi-parametric model, that is:

$$g_i = m(X_i) + v_i = \alpha + \sum_{j=1}^{K} \mu_j(X_{i,j}) + v_i$$

where $\alpha$ is a constant term, $\mu_j(\cdot)$ are one-dimensional nonparametric functions operating on each of the $K$ elements of $X_i$, and $v_i$ is an error term with the properties: $E(v_i|X_i) = 0$, $\text{var}(v_i|X_i) = \sigma^2(X_i)$ (i.e. the model allows for heteroskedasticity).

### 2.3 Distributional Effects of Individual Variables

Denote by $X_{i,k}$ the vector of all explanatory variables but $X_{i,k}$ for region $i$, i.e.:

$$X_{i,k} = (X_{i,1}, \ldots, X_{i,(k-1)}, X_{i,(k+1)}, \ldots, X_{i,K})$$

Eq. (3) can be rewritten as:

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

(4)

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

$$y_i(T) = y_i(0)e^{[\alpha + \sum_{j \neq k} \mu_j(X_{i,j}) + v_i]T} = y_i(0)e^{[\alpha + \sum_{j \neq k} \mu_j(X_{i,j})]T} e^{\mu_k(X_{i,k})T} e^{v_iT},$$

(5)

where $y_{i,k}(T) = y_i(0)e^{[\alpha + \sum_{j \neq k} \mu_j(X_{i,j})]T}$ is the level of productivity in period $T$ obtained by “factoring out” the effect of $X_{i,k}$; $g_{i,k}^M = \mu_k(X_{i,k})$ 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, $g_{i,k}^R = v_i$ is the annual “residual growth”, not explained by the variables in $X_i$. The modelling of growth in Eq. (5) will be the basis for the identification of the distributional effects of the $k$-th variable.

### 2.3.1 Counterfactual Stochastic Kernels and Ergodic Distributions

We define the counterfactual productivity $y_{i,k}^{CF}(T)$, 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

4Notation refers to H"ardle et al. (2004).
5Durlauf 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.
are collected in the N-dimensional vector \( \mathbf{X}_k \). That is, \( y_{i,k}^{CF}(T) \) 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 region the cross-section average value of the variable.

Hence, the counterfactual growth rate of region \( i \) with respect to the \( k \)-th variable, \( g_{i,k}^{CF} \), is defined as:

\[
g_{i,k}^{CF} = \alpha + \sum_{j \neq k} \mu_j(X_{i,j}) + \mu_k(\bar{X}_k) + v_i,
\]

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

\[
y_{i,k}^{CF}(T) = y_i(0)e^{g_{i,k}^{CF}T} = y_i(0)e^{\alpha + \sum_{j \neq k} \mu_j(X_{i,j}) + \mu_k(\bar{X}_k) + v_i}T.
\]

Counterfactual productivities are the bases to compute counterfactual stochastic kernels. Specifically, the actual and counterfactual stochastic kernels are respectively defined as \( \phi(y(T)|y(0)) \) and \( \phi^{CF}(y_k^{CF}(T)|y(0)) \), where \( y(0), y(T) \) and \( y_k^{CF}(T) \) are the vectors collecting regional productivities at times 0 and \( T \).

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 highlights, for every initial productivity level, the probability distribution over productivity levels at time \( T \) if the cross-region heterogeneity in the variable \( k \) is suppressed. 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 regions.

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).\(^8\) The ergodic distribution highlights whether the estimated distribution dynamics isolate the effects of the distribution of a variable by imposing on the second period the values of the variable in the first period.

\(^7\)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 \( t + \tau \). That is, 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 two levels 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) increases when the density of observations decreases.

\(^8\)Specifically, the ergodic distribution solves \( f_\infty(z) = \int_0^\infty g_\tau(z|x) f_\infty(x) \, dx \).
over the period of interest has completely exhausted its effects or, otherwise, significant distributional changes are expected in the future.

### 2.3.2 The Distributional Effect of Individual Variables and the Marginal Growth Effect

To evaluate the distributional effect of individual variables, we consider two aspects: i) we assess the capacity of an individual variable to make actual and counterfactual stochastic kernels differ; 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 parts of the productivity distribution it affects.

To analyze possible 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, $y_{i,k}^{CF}(T)$:

$$\log(y_i(T)) = \log(y_{i,k}^{CF}(T)) + [\mu_k(X_{i,k}) - \mu_k(X_k)]T + v_i T. \quad (8)$$

The expected value of (the log of) actual productivity of region $i$ 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 with $\tau = T$. In particular, its relation with the expected value from the counterfactual kernel can be expressed as:

$$E[\log(y_i(T))|y_i(0)] = E[\log(y_{i,k}^{CF}(T))|y_i(0)] + E[\mu_k(X_{i,k}) - \mu_k(X_k)]y_i(0)T. \quad (9)$$

From Eq. (9), 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.:

$$E[\log(y_i(T))|y_i(0)] = E[\log(y_{i,k}^{CF}(T))|y_i(0)] \quad (10)$$

if:

$$E[\mu_k(X_{i,k})|y_i(0)] = \mu_k(X_k). \quad (11)$$

The result in Eq. (11) 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.

Therefore, if Conditions $[1]$ and $[2]$ hold, we obtain the condition in Eq. (11), i.e.:

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

Eq. (12) 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.
Notice that, given our choice to base counterfactual analysis on cross-section averages, the use of a semiparametric specification that allows for nonlinearities, is necessary to identify possible differences between the actual and counterfactual stochastic kernels, 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).

As a second step to evaluate the impact of the individual variables on the 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 the marginal growth effect of the $k$-th variable in Eqns. (3)-(5), i.e. $g^M_{i,k} = \mu_k(X_{i,k})$. 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 the marginal growth $g^M_k$ conditioned on the initial level of productivity, i.e. by estimating $\phi^M(g^M_k | y(0))$. If the estimate of the marginal effect does not result statistically different from its unconditional mean, i.e. $\phi^M(g^M_k | y(0)) = E[g^M_k] \forall y(0)$, then the $k$-th variable has no distributional effects. On the contrary, if $\phi^M(g^M_k | y(0))$ is statistically different from its unconditional mean and, in particular, it is an increasing (decreasing) function of $y(0)$, 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:

\[ E_{X_k} [\mu_k(X_k)] = 0 \]  

(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.

Having detailed our methodology, let us finally remark that, with respect to the mentioned existing methods of estimating the effect of individual variables on the distribution dynamics, especially through the estimation of “conditioned” stochastic kernels: i) our method is based on a multivariate analysis to identify the effect of a specific variable, while the method proposed by Quah (1997) is based on the consideration of one variable at the time. Here, by excluding the variables of interest one by one, we are 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), on the contrary, 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 these variables; ii) in particular with respect to Cheshire and Magrini (2005), we use a semiparametric method, instead of a linear regression, for the baseline estimation.

### 2.4 Test of Distributional Effects of Residual Growth

As a final step, we propose a test for the goodness of fit. In particular, we elaborate a measure of goodness of fit of the growth regression conditional on the initial level of productivity, i.e. of the
presence of possible misspecifications of the model for some ranges of initial productivity.

Eq. (5) suggests to consider $\hat{g}^R$, defined as $\hat{g}^R \equiv \log \left( \frac{y(T)}{y(T-1)} \right)$, to test that:

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

(13)

If $y(0)$ is included in the set of regressors, the condition in Eq. (13) ensures that there is no omitted variable inconsistency related to $y(0)$ (see Wooldridge, 2002, pp. 61-63). This condition in Eq. (13) will be used as a test of misspecification of the growth model.

3 An Empirical Application to a Cross-Section of Countries

To demonstrate the practical use of our methodology, we provide an empirical application to the labour productivity of a sample of countries already studied in Beaudry et al. (2005). In particular, we consider 73 (61 with education) out of 75 (68 with education) countries used in Beaudry et al. (2005). Differently from Beaudry et al. (2005) we consider data over the period 1960-2008 from PWT 7.1, while data on education come from Choen and Soto (2007).

Specifically, in Section 3.1 we estimate the growth model of Eq. (3); in Section 3.2 we test for the presence of distribution effects in residual growth; in Section 3.3 we study the unconditional distribution dynamics of labour productivity that we will use as benchmark; finally, in Section 3.4 we present the distributional impact of regressors.

3.1 The Estimation of a Growth Model for Countries

Following 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) its (log) initial level of GDP per worker ($y_0$); ii) its (log) average annual employment growth rate ($n$); iii) its (log) average annual investment ratio at constant price ($i/y$); and, iv) its (log) average years of schooling ($Edu$).

Results of the estimated models are reported in Table 1. All regressors initially enter as nonparametric terms. However, they are substituted by linear terms if their effect results to be linear.

Following Beaudry et al. (2005), in Model I we check the stability of the growth regression in the period 1960-1978 versus 1978-1998, by focusing on the model with only $y_0$, $n$ and $i/y$. The estimates of
### 3.1 Estimation of Growth Model

**Model I**  
Dep. Var: $g$  
1960-1978  
GAM  

**Model II**  
Dep. Var: $g$  
1978-1998  
GAM  

**Model III**  
Dep. Var: $g$  
1960-1998  
GAM  

**Model IV**  
Dep. Var: $g$  
1960-2008  
GAM  
1960-2008

<table>
<thead>
<tr>
<th>Parametric coefficients:</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>0.0325***</td>
<td>0.0105</td>
<td>0.0210***</td>
<td>0.0209***</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non parametric coefficients:</th>
<th>Model I</th>
<th>Model II</th>
<th>Model III</th>
<th>Model IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>y0</td>
<td>2.4800***</td>
<td>1.0000**</td>
<td>7.5460***</td>
<td>7.1690***</td>
</tr>
<tr>
<td>n</td>
<td>1.9560***</td>
<td>2.2410***</td>
<td>3.5740***</td>
<td>3.2380***</td>
</tr>
<tr>
<td>i/y</td>
<td>1.0000***</td>
<td>3.5700**</td>
<td>5.7840***</td>
<td>4.7090***</td>
</tr>
<tr>
<td>Edu</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.45</td>
<td>0.32</td>
<td>0.61</td>
<td>0.69</td>
</tr>
<tr>
<td>GCV score (*10^5)</td>
<td>20.54</td>
<td>29.04</td>
<td>11.20</td>
<td>7.31</td>
</tr>
<tr>
<td>Scale est.(*10^5)</td>
<td>18.72</td>
<td>26.01</td>
<td>8.46</td>
<td>5.70</td>
</tr>
<tr>
<td>Obs.</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
</tr>
</tbody>
</table>

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 best specification.
the two subperiods turn to be not statistically different at 95% significance level for all the variables, according to a bootstrap test reported in Appendix C. Therefore, in contrast with Beaudry et al. (2005) the growth model appears stable over the two subperiods. The comparison of the estimate of the Model II and III, related to the periods 1960-1998 and 1960-2008 respectively, shows that we can further extend the period of the analysis.

Finally, we introduce education as further explanatory variable in Model IV. In its best specification the initial level of productivity and the education enter linearly. In particular, \( y_0 \) has a negative and significant effect, while \( \text{Edu} \) has a positive and significant effect. The estimated impacts of \( n \) and \( i/y \) are nonlinear as reported in Figures 1 and 2. In Appendix D we check that these results are robust to endogeneity.

Figure 1: Estimated partial effect of \( n \). Thick line: estimates; dotted lines: 95% confidence bands.

Figure 2: Estimated partial effect of \( i/y \). Thick line: estimates; dotted lines: 95% confidence bands.

### 3.2 Test of Residual Growth

Figure 3 reports the estimated density of the annual residual growth \( \hat{g}^R \) conditioned on the initial level of productivity. We also report the conditional mean (thick line) with the corresponding confidence bands, 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

\[ \text{Beaudry et al. (2005) also find a negative coefficient for } y_0 \text{ but no significant effect of } \text{Edu}. \]
appears correctly specified, at least conditioning on the initial level of productivity (see Eq. 13).

\begin{figure}[h]
\centering
\includegraphics[width=\linewidth]{figure3.png}
\caption{Conditional distribution of residual growth, the conditional mean (thick line), its confidence bands (dotted lines) and the unconditional mean (thin vertical line).}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\linewidth]{figure4.png}
\caption{Conditional distribution of residual growth with bias, the conditional mean (thick line), its confidence bands (dotted lines) and the unconditional mean (thin vertical line).}
\end{figure}

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.

### 3.3 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 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\(^\text{16}\) and the 45° line.

\begin{figure}[h]
\centering
\includegraphics[width=\linewidth]{figure5.png}
\caption{The actual stochastic kernel of productivity.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\linewidth]{figure6.png}
\caption{The actual distributions (AD) of productivity in 1960 and 2008, along with the actual ergodic distribution (AED).}
\end{figure}

\(^{16}\)The procedure is illustrated in Appendix E.
3.4 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.

3.4.1 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 unconstrained growth path.

Tests of multimodality 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. Tests of multimodality follow the bootstrap procedure described in Silverman (1986, p. 146), and are performed using 1000 bootstraps.
Conditional mean for every country with an initial productivity below the average, while the opposite holds for countries with above-average initial productivity.

**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.

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 fall of inequality from 1960 to 2008 in about 10 base points.¹⁸

¹⁸The difference between indexes related to AD and CD is statistically significant.
Overall, although initial productivity contributes to reduce inequality, still remains a tendency towards polarization 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 $y_0$ 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.

### 3.4.2 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. The distributional effect of 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).

The conditional mean of MGE of employment growth is statistically different from the unconditional mean for countries with an initial productivity higher than 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.
3.4 Conditional Distribution Dynamics

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.

Table 3: Gini Indexes and their standard errors of AD, CD, AED and CED. Counterfactual variable: Employment Growth.

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 2.3.2, 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 very 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).

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.
3.4 Conditional Distribution Dynamics

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

#### 3.4.3 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 every country with an initial productivity above the average, while the opposite holds for countries with below-average initial productivity. Thus, the 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 disperse and there would have been less inequality (about 6 base points in the Gini index). This tendency also appears in the CED, shown in Figure [18] and Table 4.

<table>
<thead>
<tr>
<th></th>
<th>AD 1960</th>
<th>AD 2008</th>
<th>CD 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gini</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>s.e.</td>
<td>(0.029)</td>
<td>(0.026)</td>
<td>(0.024)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unimodality Test</th>
<th>AED</th>
<th>CED</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.271</td>
<td>0.009</td>
</tr>
<tr>
<td>s.e.</td>
<td>(0.335)</td>
<td>(0.374)</td>
</tr>
</tbody>
</table>

Table 4: Gini Indexes and their standard errors of AD, CD, AED and CED. Counterfactual variable: Investment Ratio.
3.4 Conditional Distribution Dynamics

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.

Table 5: Gini Indexes and their standard errors of AD, CD, AED and CED. Counterfactual variable: Education.


4 Concluding Remarks

In this paper we proposed a method to analyze the factors driving convergence and divergence processes in the growth dynamics. The proposed methodology combines the growth regression approach, albeit allowing for a semiparametric specification, with the distribution dynamics approach. In particular, the potential distributional impact of a given variable is evaluated by the comparison of actual, counterfactual and ergodic distributions (and the related actual and counterfactual stochastic kernels), where counterfactuals are calculated by removing cross-section heterogeneity in the variable of interest through the imputation of sample averages to all the units of the cross-section. The methodology also allows for testing for the possible presence of distributional effects in the residuals of growth regression, as a means to assess the goodness of fit of the estimated model.

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 the education. In all cases it was possible to obtain information otherwise missed by existing methods of investigation of the determinants of distribution dynamics.

A List of Countries in the Sample

Our sample of 73 countries consists of: 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, 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:

- $y_0$ is the (log) of the real GDP chain per worker ($rgdpwok$ in PWT 7.1).
- $g$ is the corresponding annualized average growth rate of $y_0$.

19With respect to the Beaudry et al. (2005)’s sample, we lack Guyana and Tunisia. Countries marked with * do not have data on education.
C  Bootstrap Procedure to Test 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 points 1-3.

---

- **n** is the (log) growth rate of employment, where workers are computed as the population from 15 to 64 obtained from:

  \[ workers = \frac{rgdpch}{rgdpwok} \times pop; \]

  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 Cho and Soto (2007) (“Years of schooling of population 15 and over, whether studying or not”).

---

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

<table>
<thead>
<tr>
<th>g</th>
<th>y0</th>
<th>n</th>
<th>i/y</th>
<th>Edu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.02</td>
<td>9.17</td>
<td>-2.66</td>
<td>-1.48</td>
</tr>
<tr>
<td>Stand. Dev.</td>
<td>0.01</td>
<td>1.01</td>
<td>0.15</td>
<td>0.27</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>g</th>
<th>y0</th>
<th>n</th>
<th>i/y</th>
<th>Edu</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>1.00</td>
<td>-0.33</td>
<td>-0.38</td>
<td>0.45</td>
</tr>
<tr>
<td>y0</td>
<td>-0.33</td>
<td>1.00</td>
<td>-0.28</td>
<td>0.16</td>
</tr>
<tr>
<td>n</td>
<td>-0.38</td>
<td>0.28</td>
<td>1.00</td>
<td>-0.08</td>
</tr>
<tr>
<td>i/y</td>
<td>0.45</td>
<td>0.16</td>
<td>-0.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Edu</td>
<td>0.07</td>
<td>0.76</td>
<td>-0.44</td>
<td>0.35</td>
</tr>
</tbody>
</table>
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 $y_0$. 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.
D 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.

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);

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. Figure 22 graphs 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.
Table 8: First-stage regressions of potentially endogenous variables. Significance codes: 0.01"***" 0.05"**" 0.1"*".

<table>
<thead>
<tr>
<th>Dep. Var:</th>
<th>n</th>
<th>i/y</th>
<th>Edu</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parametric coefficients:</strong></td>
<td>Estimate</td>
<td>Estimate</td>
<td>Estimate</td>
</tr>
<tr>
<td><strong>const</strong></td>
<td>-2.6649***</td>
<td>-1.4808***</td>
<td>1.82420***</td>
</tr>
<tr>
<td><strong>Non parametric coefficients:</strong></td>
<td>EDF</td>
<td>EDF</td>
<td>EDF</td>
</tr>
<tr>
<td><strong>y0</strong></td>
<td>1.0000**</td>
<td>4.486*</td>
<td>3.690*</td>
</tr>
<tr>
<td><strong>n.1960</strong></td>
<td>5.827***</td>
<td>2.482**</td>
<td>1.000</td>
</tr>
<tr>
<td><strong>i/y.1960</strong></td>
<td>1.087</td>
<td>1.000***</td>
<td>5.505*</td>
</tr>
<tr>
<td><strong>Edu.1960</strong></td>
<td>3.907***</td>
<td>2.000</td>
<td>3.120***</td>
</tr>
<tr>
<td><strong>R^2</strong></td>
<td>0.82</td>
<td>0.55</td>
<td>0.61</td>
</tr>
<tr>
<td><strong>GCV score</strong></td>
<td>0.0048</td>
<td>0.0397</td>
<td>0.0145</td>
</tr>
<tr>
<td><strong>Scale est.</strong></td>
<td>0.0038</td>
<td>0.0326</td>
<td>0.0111</td>
</tr>
<tr>
<td><strong>Obs.</strong></td>
<td>61</td>
<td>61</td>
<td>61</td>
</tr>
</tbody>
</table>

D  ENDOGENEITY TEST
<table>
<thead>
<tr>
<th></th>
<th>Including v</th>
<th>Excluding v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dep. Var: $g$</td>
<td>GAM with CF</td>
<td>GAM</td>
</tr>
<tr>
<td></td>
<td>1960-2008</td>
<td>1960-2008</td>
</tr>
<tr>
<td><strong>Parametric coefficients:</strong></td>
<td>Estimate</td>
<td>Estimate</td>
</tr>
<tr>
<td>const</td>
<td>0.0776***</td>
<td>0.0779***</td>
</tr>
<tr>
<td>$y_0$</td>
<td>-0.0083***</td>
<td>-0.0082***</td>
</tr>
<tr>
<td>$Edu$</td>
<td>0.0097**</td>
<td>0.0092**</td>
</tr>
<tr>
<td><strong>Non parametric coefficients:</strong></td>
<td>EDF</td>
<td>EDF</td>
</tr>
<tr>
<td>$n$</td>
<td>4.026***</td>
<td>3.734***</td>
</tr>
<tr>
<td>$i/y$</td>
<td>4.162***</td>
<td>4.346***</td>
</tr>
<tr>
<td><strong>First-stage residuals coefficients:</strong></td>
<td>Estimate</td>
<td>Estimate</td>
</tr>
<tr>
<td>$n_{res}$</td>
<td>-0.0141</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5140)</td>
<td></td>
</tr>
<tr>
<td>$i/y_{res}$</td>
<td>0.0105</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2540)</td>
<td></td>
</tr>
<tr>
<td>$Edu_{res}$</td>
<td>0.0009</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.9470)</td>
<td></td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>0.69</td>
<td>0.70</td>
</tr>
<tr>
<td>GCV score ($\times 10^5$)</td>
<td>6.24</td>
<td>5.74</td>
</tr>
<tr>
<td>Scale est.($\times 10^5$)</td>
<td>4.79</td>
<td>4.69</td>
</tr>
<tr>
<td>Obs.</td>
<td>61</td>
<td>61</td>
</tr>
</tbody>
</table>

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 for first-stage residuals in parenthesis.
E  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 $Y = \{Y_1, ..., Y_m\}$ where $Y_i$ is a vector of dimension $n$, the bootstrap algorithm consists of three steps.

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

2. Select $B$ independent bootstrap samples $\{Y_1^*, ..., Y_B^*\}$, each consisting of $n$ data values drawn with replacement from $Y$.

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 $\psi$.

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 $Y$. 

Figure 22: Second-stage regression.
References


