Deep and Proximate Determinants of the World Income Distribution

Davide Fiaschi, Andrea Mario Lavezzi and Angela Parenti*

Abstract

This paper studies the deep and proximate determinants of the evolution of the cross-country distribution of GDP per worker in the period 1960-2008 by a novel method based on information criterion. We find that countries of our sample follow three distinctive growth regimes identified by two deep determinants, namely life expectancy at birth in 1960 and the share of Catholics in 1965, and that each regime is characterized by nonlinearities. Growth regimes appear to be the main cause of the increased inequality and polarization, while technological catch-up, proxied by the initial level of GDP per worker, acts in the opposite direction. Finally, human capital marginally reduces polarization, while investment rates and employment growth have no distributional effect.

Keywords: polarization, growth regimes, parameter heterogeneity, nonlinearities, model uncertainty.

JEL: *C14; O47*

*Davide Fiaschi (corresponding author), Dipartimento di Economia e Management, University of Pisa, Via Ridolfi 10, 56124 Pisa, Italy. Tel +39 050 2216208, Fax. +39 050 2216384, e-mail: davide.fiaschi@unipi.it ; Andrea Mario Lavezzi, Dipartimento di Giurisprudenza, University of Palermo, and CICSE, Piazza Bologni 8, 90134 Palermo, Italy, e-mail: mario.lavezzi@unipa.it; Angela Parenti, Dipartimento di Economia e Management, University of Pisa, Via Ridolfi 10, 56124 Pisa, Italy, e-mail: angela.parenti@unipi.it.

1 Introduction

The literature on growth empirics has not reached a consensus on the determinants of world income inequality (Johnson and Papageorgious, 2017). We believe that this failure is mainly due to the lack of consideration of a hierarchy among the set of candidate determinants which was, on the contrary, a key characteristic of the seminal paper by Durlauf and Johnson (1995). Therefore, in this paper we propose a new method based on information criterion which allows us to identify deep and proximate determinants in the spirit of Rodrik (2003), and to deal with other critical issues discussed in the literature such as model uncertainty, nonlinearities and endogeneity. We then apply this method to investigating the determinants of inequality and polarization in the world distribution of income, measured by GDP per worker, in the period 1960-2008.

The most significant stylized fact on the evolution of the cross-country income distribution is the shift from unimodality in the 1960s to bimodality in the 1990s (see Durlauf et al., 2005, for an exhaustive discussion and references). Several, potentially complementary, explanations have been advanced. A first explanation relies on the assumption that different countries obey different growth processes, i.e. they belong to different *growth regimes* according to their initial conditions proxied, for example, by GDP per capita, human capital, or life expectancy (see, e.g., Durlauf and Johnson, 1995, Durlauf et al., 2001, and Kourtellos, 2011). Another explanation is based on the effect of nonlinearities in the growth process (see, e.g., Liu and Stengos, 1999), while a third distinguishes *deep* (or fundamental) from *proximate* growth determinants, assuming that the former determine the latter and, ultimately, long-run outcomes (see Rodrik, 2003).[1] The deep determinants proposed in the literature include: institutions (Acemoglu et al., 2005); culture, in particular in the form of social capital (Knack and Keefer, 1997, Temple and Johnson, 1998) and religion (Durlauf et al., 2012); geography (Bloom et al., 2003); and ethnolinguistic fractionalization (Easterly and Levine, 1997, Tan, 2009). The proximate determinants are those typically appearing in the production function, i.e. factors of production and technology (see, e.g., Rodrik, 2003 and Weil, 2012). Embracing one or the other explanation implies profoundly different growth-enhancing policies (Rodrik, 2003).

Our empirical strategy integrates the insights of these different lines of research. Specifically, our method identifies growth regimes by a set of candidate deep determinants (we label "deep" determinant any variable used to identify growth regimes), and *simultaneously* estimates a nonlinear growth model within each growth regime, which includes the proximate determinants suggested by Mankiw et al. (1992): initial income per worker (as a proxy for technological catch-up), investment rate, employment growth and human capital.[2] Applying this method to a sample of 84 countries over the period 1960-

[1]Weil (2012, p. 53) classifies growth determinants into "proximate" and "ultimate".

[2]Brock (2001) proposes a taxonomy of growth determinants based on their time scales, where "deep" determinants are moving on a slower time scale than "proximate" determinants, while Tan (2009) distinguishes between "development clubs", identified on the basis of the sole deep determinants, and "growth regimes",

2008, we identify as relevant deep determinants initial health conditions, proxied by life expectancy at birth in 1960, and culture, proxied by the share of Catholics in 1965. In particular, we identify three regimes: "high life expectancy regime", "low life expectancy/high share of Catholics regime", and "low life expectancy/low share of Catholics regime". Furthermore, we show that nonlinearities within the regimes are a pervasive phenomenon. Using a counterfactual analysis we demonstrate that growth regimes are the main source of polarization and inequality. Among the proximate determinants, initial income has the opposite distributional effect, human capital marginally reduces polarization, while the investment rate and employment growth rate have no significant effects.

Our results contrast with several existing findings, and contain some novelties. Contributions on the relative importance of competing deep determinants of growth such as institutions and geography conclude that institutions prevail (Rodrik et al., 2004, Tan, 2009, Owen et al., 2009 and Flachaire et al., 2014). In identifying growth regimes we consider the largest set of candidates with respect to the existing literature, among which are institutions and geography: the latter are however both dominated by life expectancy at birth in 1960 and the share of Catholics in 1965. Our results therefore confirm the importance of culture, as proxied by the share of Catholics in 1965, in development, as thoroughly discussed in Guiso et al. (2006), but as regime identifier and not as a covariate in a growth regression.[3] Finally, with the partial exception of Kourtellos (2011), no previous work found life expectancy as an identifier of growth regimes.[4]

We find significant nonlinearities within regimes, suggesting that previous works based on linear specifications may suffer from misspecification bias. Moreover, proximate determinants are generally significant in all regimes with an important difference: human capital has a positive effect on growth only in the "low life expectancy/low share of Catholics" regime. This evidence can help to explain why Durlauf et al. (2012) do not find any effect of human capital, given their choice to consider our deep determinants as additional covariates in a regression, and to estimate linear models. We do not find any significant distributional effect of cross-country heterogeneity in investment and employment growth rates but, differently from Beaudry et al. (2005), a significant effect of the initial level of GDP per worker.

Lastly, we contribute to the debate on whether: "the transition to the long-run steady-state [can be] associated with non-monotonic evolution of the distribution of income across countries. Thus, convergence may be preceded by polarization and clustering, and club convergence will be generated by these models in the medium run" (Galor, 1996, p. 96). In particular, Lucas (2000) and Galor (2007) claim that polarization is a transitory phenomenon caused by the different timing of countries' take-off. Lucas (2000) argues that countries randomly start their growth process and subsequently

identified by both the deep and proximate determinants.

[3]The role of religion as a covariate has been convincingly challenged by Durlauf et al. (2012).

[4]However, Kourtellos (2011) does not deal with the issue of model uncertainty.

adopt the technology of the leading countries. Differently, Galor and Weil (2000) and Galor (2007) propose the Unified Growth Theory (UGT), according to which a country transits from a Malthusian Regime to a Post-Malthusian Regime, and finally reaches the Modern Growth Regime. Although our "high-life expectancy" regime has the characteristics of the Modern Growth Regime (but the other two regimes differ from those hypothesized by the UGT), our counterfactual analysis suggests that in the period 1960-2008 the predicted regime transitions did not take place for a large number of countries. Thus our analysis suggests that "club convergence" is a persistent phenomenon. An important (and obvious) caveat to this claim is that while Lucas (2000) and Galor (2007) consider a very long-run horizon, our analysis is limited to 48 years (which is however a long period compared to other studies on distribution dynamics).

The paper is organized as follows: Section 2 discusses the related literature; Section 3 describes the method; Section 4 presents the empirical analysis; Section 5 concludes. The appendices contain some technical details on the method and on data.

2 Related Literature

The importance of classifying growth determinants into deep and proximate is discussed, among others, by Rodrik (2003) and Weil (2012). An investigation of the impact of deep determinants for long-term development is proposed by Spolaore and Wacziarg (2013), who also offer an exhaustive review of the existing literature. The main thrust of the argument is that, while the proximate determinants *directly* affect growth, they are themselves determined by other, *deeper*, determinants such as geography, institutions and culture.[5] Weil (2012) highlights the various links between deep and proximate determinants. For example, geographic location can favor trade and technological spillovers; institutions can encourage savings and the accumulation of factors; culture can imply openness or closure to new ideas and technologies, a positive attitude towards hard work, favoring efficiency, or to thriftiness, favoring accumulation.[6] Understanding economic growth and comparative development, therefore, requires that the relevant deep determinants be identified. The novelty of our approach is the *joint* identification of the relevant deep determinants and of the growth models within each of the identified growth regimes.

In the literature, different methods have been utilized to identify growth regimes. Durlauf and Johnson (1995) and Tan (2009) use clustering algorithms (denoted by CART and GUIDE respectively) that sequentially partition countries into regimes on the basis of some deep determinants; Desdoigts

[5]Among the deep determinants, Weil (2012) also considers inequality, while Rodrik (2003) includes trade openness.

[6]For a detailed account see Weil (2012). Rodrik (2003) contains some remarks on the exogeneity, or (partial) endogeneity, of the deep determinants and on their interrelations.

(1999) utilizes a projection pursuit approach based on proximate determinants, and indirectly identifies the relevant deep determinants; Owen et al. (2009), Flachaire et al. (2014) and Anderson et al. (2016) use finite mixture models, while Bos et al. (2010) split a sample of countries by a multinomial logit model, and then estimate a stochastic frontier model within each growth regime. The main difference with respect to these works is that we allow for nonlinearities within each regime and study the effect of growth regimes on the evolution of income distribution.[7] Moreover, our procedure allows for model selection under uncertainty that if "ignored, [would imply that] precision is often overestimated, achieved confidence interval coverage is below the nominal level, and predictions are less accurate than expected" (Burnham and Anderson, 2003, p. 3).

Our work is also related to the studies on the determinants of distribution dynamics. Specifically, Quah (1996) introduces the concept of *conditioned* stochastic kernel,[8] based on residuals from a regression of GDP per worker on proximate determinants, while Quah (1997) proposes a conditioned stochastic kernel based on GDP per capita normalized with respect to a weighted sample average, where weights are defined by geographical proximity or intensity of trade with other countries. By considering the residuals of a regression, however, Quah (1996) can only obtain an estimate of the *joint* distributional effect of the determinants included in the regression, while we are able to identify the effect of individual variables. Quah (1997), instead, considers one variable at the time, but does not control for the effect of other determinants. Johnson (2005) and Feyrer (2008), differently, explain the income distribution dynamics by a comparison with the distributions of proximate growth determinants, such as human capital, physical capital and total factor productivity, assuming a common worldwide Cobb-Douglas production function. By allowing for the presence of growth regimes, we do not assume the existence of a common production function.

Finally, the use of counterfactual analysis to study the determinants of distribution dynamics was previously proposed by Beaudry et al. (2005), Cheshire and Magrini (2005) and Henderson and Russell (2005). In particular, Beaudry et al. (2005) analyze the distributional effect of proximate determinants comparing the periods 1960-1978 and 1978-1998, characterized by the emergence of polarization. Their strategy consists in estimating counterfactual distributions for the second period assuming that a factor of interest (e.g. the estimated coefficient of a growth regression or the distribution of a growth determinant) maintains in the second period the same value as in the first. Differently, Cheshire and Magrini (2005) estimate a growth regression, and then compute counterfactual distributions by

[7] Partial exceptions are Desdoigts (1999), who does not specify any growth model, and Bos et al. (2010), who estimate a stochastic frontier model assuming a translog production function. In addition, Maasoumi et al. (2007) consider a nonlinear growth model *assuming* the existence of two regimes, i.e. OECD and non-OECD countries, but focus on growth rate distribution.

[8] The stochastic kernel is an operator mapping the current distribution into the future distribution. See on-line Appendix .8 for details.

comparing 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 determinants in the growth regression), while Henderson and Russell (2005) propose a counterfactual analysis based on the production-frontier approach. None of these works, however, allow for growth regimes and nonlinearities.

3 The Methodology for the Empirical Investigation

As an introduction to our method, Figure 1 reports the estimated distributions of relative (with respect to sample mean) GDP per worker in 1960 and 2008, along with the estimated long-run equilibrium distribution, denoted as *ergodic* distribution for a sample of 84 countries.[9] In the following we will denote these three types of distribution as *actual* distributions.

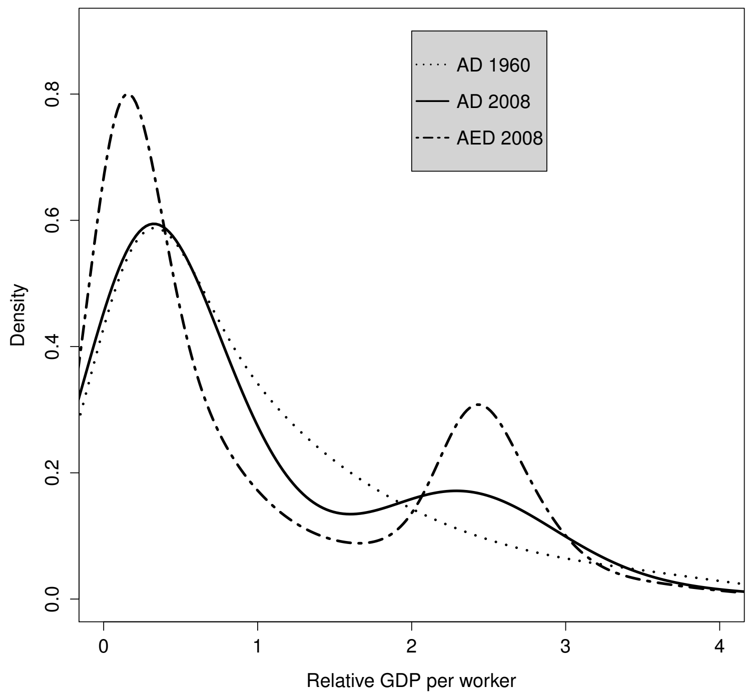


Figure 1: Actual distribution (AD) in 1960 (dotted line), in 2008 (solid line) and actual ergodic distribution (AED) (dashed line) of GDP per worker for a sample of 84 countries.

Figure 1 confirms the stylized fact emerging from the literature: the distribution is initially unimodal, but subsequently becomes twin-peaked (see, e.g., Quah, 1997). Moreover, the shape of the

[9]See on-line Appendix .1 for data sources, and on-line Appendix 3.1.1 for the country list. Technical details on the estimation can be found in the online on-line Appendix. Dataset and codes are available at authors' website.

ergodic distribution suggests that the tendency of polarization is doomed to persist in the long run.[10] In terms of the BIPOL bipolarization index proposed by Anderson et al. (2012), polarization increases from 0.75 in 2008 to 1.26 in the ergodic distribution. Inequality, measured by the Theil index, also increases over the period: the index rose from 0.54 in 1960 to 0.68 in 2008. Our aim is to identify the determinants of these changes in inequality and polarization through a novel method. In particular, in Section 4 we will investigate the role of deep and proximate determinants of growth by a method including six steps: i) identification of growth regimes in the presence of nonlinearities (Section 3.1); ii) specification and estimation of a nonlinear, regime-specific growth regression (Section 3.2); iii) decomposition of a country's GDP per worker (Section 3.3); iv) computation of counterfactual final (i.e. end-of-period) distributions (Section 3.4); v) estimation of counterfactual ergodic distributions (Section 3.5); vi) evaluation of the distributional effect of proximate determinants by their *marginal growth effect* (Section 3.6).

3.1 Identification of Growth Regimes

In this section we describe the procedure to identify growth regimes based on information theory which has fundamental advantages with respect to existing methods (CART, GUIDE, threshold regressions, finite mixture approach, etc.). The use of the Akaike information criterion (AIC) allows model selection with *non-nested*, *non-linear* models in the presence of endogeneity and, at the same time, *model selection uncertainty* to be tackled by ranking the candidate models in terms of their probability of being the *best approximating model* of the true model.[11]

The approach based on AIC also has advantages with respect to Bayesian methods: it does not depend on the choice of prior probabilities and it is computationally less demanding when the number of models under consideration is high. Even though model selection based on AIC implicitly assumes that the "true" model is in the set of candidate models, which is likely to be false, Takeuchi (1976) derives a generalized AIC robust to the absence of the "true" model in the candidate set, and concludes that AIC represents a "parsimonious approach to bias correction" due to the absence of the "true" model in the candidate set (see Anderson, 2007, p. 70). Finally, the use of Bayesian methods does not allow the use of information theory, which underpins our approach to account for model selection uncertainty (see Section 3.1.2 below). In particular, the Bayesian information criterion (BIC), the best known alternative in Bayesian literature to AIC, which appears very similar to AIC to the casual

[10] Silverman's bootstrap tests for multimodality show that the null hypothesis of unimodality cannot be rejected at the usual significance levels for the 1960 distribution, while it can be rejected at 1% of significance for the 2008 distribution and for the ergodic distribution (Silverman, 1986). Henderson et al. (2008) find the same results with a larger sample of countries (see their Table III).

[11] See Anderson (2007) for a general introduction to this approach and Claeskens and Hjort (2008) for a technical exposition of model selection based on AIC.

eye, “has [unfortunately] nothing linking it to information theory, [it is] a misnomer” (see Anderson, 2007, p. 160).

3.1.1 The Procedure to Explore all Potential Growth Regimes

The procedure to explore all potential growth regimes is structured in five steps:

1. Define the set of deep determinants $\mathbf{Z}$ and consider a subset $\mathbf{Z}_q$. For each deep determinant, a threshold value will be used to partition the sample.

2. On the basis of $\mathbf{Z}_q$ identify all possible P_q partitions of countries by sequentially splitting the sample. For each subset q and partition p a maximum number of regimes $M_{q,p}$ can be identified. Assign each country to a growth regime and gather them in the set $\mathbf{GR}_{q,p} = \left\{R_{q,p,1}, ..., R_{q,p,M_{q,p}}\right\}$, where $R_{q,p}$ are the possible regimes. Collect all these partitions of countries in the set $\mathbf{GR}_q = \left\{\mathbf{GR}_{q,1}, ..., \mathbf{GR}_{q,P_q}\right\}$.

3. For each growth regime estimate a semiparametric growth regression controlling for endogeneity using the Control Function Method, and obtain the residuals:
$$\hat{\upsilon}_{q,p,i} = g_i - \hat{\alpha}_t\left(m_{q,p}\right) - \sum_{j=1}^{K} \hat{\mu}_j\left(X_{i,j}, m_{q,p}\right),$$
where $m_{q,p}$ is the growth regime of country i, given Z_q and P_p.

4. Compute the total log-likelihood (up to a constant) of the model:
$$\log\left(\mathcal{L}\right)_{q,p} = -\left(\frac{N}{2}\right)\log\left(\frac{\sum_{m=1}^{M_{q,p}}\sum_{i\in R_{q,p,m}}\hat{\upsilon}^2_{q,p,i}}{N}\right),$$
and the related $AICc$:
$$AICc_{q,p} = -2\log\left(\mathcal{L}\right)_{q,p} + 2F\left(\frac{N}{N-F-1}\right), \tag{1}$$
where F is the total number of estimated parameters in the model.[12]

5. The minimum $AICc_{q,p}$ in q and p, $AICc_{\min}$, *jointly* identifies i) the best partition of countries into different growth regimes and ii) the best estimation of the semiparametric growth model for each growth regime: this represents the *best model* for our sample.

The number of deep determinants that can be used in the identification of growth regimes is limited by the number of countries N and sub-periods S, given that in each partition a minimum number of observations is needed for the semiparametric estimation of the growth model. For example,

[12]The use of semiparametric techniques increases the number of estimated parameters proportionally to the identified nonlinearities (see Section 4.2).

using two deep determinants, $\mathbf{Z}_q = (\mathbf{Z}_{q,1}, \mathbf{Z}_{q,2})$, and one threshold for each of them ($Z_{q,1}^{TRESH}$ and $Z_{q,2}^{TRESH}$) means searching for the existence of four growth regimes in the $(\mathbf{Z}_{q,1}, \mathbf{Z}_{q,2})$-space. On average, $(N \times S)/4$ observations will be available for the estimation of the growth model within each regime. In particular, in Figure 2 it is assumed that $\mathbf{Z}_{q,1}$ is the first partitioning variable, and each partition cannot be populated by less than N^{MIN} countries. The resulting partition can include at least one, and at most four growth regimes. Moreover, the total number of possible partitions P_q depends on the number of deep determinants considered in the analysis, and on the different values they display. For example, considering the same two deep determinants $\mathbf{Z}_q = (\mathbf{Z}_{q,1}, \mathbf{Z}_{q,2})$, each taking on N different values (i.e. the maximum, equal to the number of countries in the sample), implies that $P_q = N \times N = N^2$. This means that the maximum number of partitions is equal to $N^2 \left(|\mathbf{Z}|^2 - |\mathbf{Z}|\right)$.

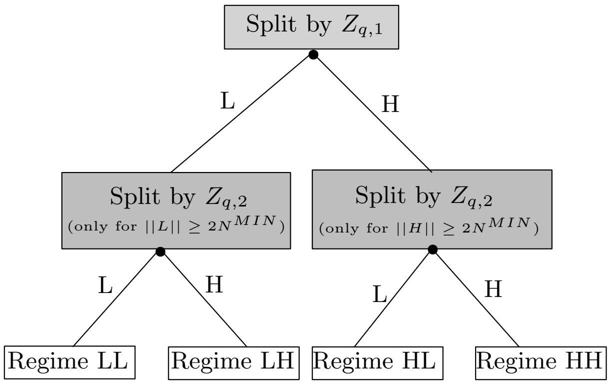


Figure 2: Sequential splitting procedure to identify all possible partitions.

3.1.2 Model selection uncertainty

To account for the *model selection uncertainty*, we compute for each model the *loss of information* as in Anderson (2007, pp. 84-86):

$$\Delta AICc_{q,p} = AICc_{q,p} - AICc_{\min}, \tag{2}$$

where $AICc_{q,p}$ is the AIC of Model (q,p) (corresponding to a partition p of countries and a subset q of deep determinants, see on-line Appendix 3.1.1) adjusted for the degrees of freedom (see Anderson, 2007), and $AICc_{\min}$ is the model with the minimum AIC among all the models considered in the procedure. Thus, $\Delta AICc_{q,p}$ ranks the candidate models: the larger the $\Delta AICc_{q,p}$, the less likely Model (q,p) is the best approximating model in the candidate set. The simple transformation:

$$\exp\left(-\frac{\Delta AICc_{q,p}}{2}\right) \tag{3}$$

provides the likelihood of Model (q,p), and the following normalization:

$$w_{q,p} = \frac{\exp\left(-\Delta AICc_{q,p}/2\right)}{\sum_q \sum_p \exp\left(-\Delta AICc_{q,p}/2\right)} \tag{4}$$

gives the probability of Model (q, p) to be the best model in the candidate set. It is worth remarking that $w_{q,p}$ can be interpreted as the: "Bayesian posterior model probabilit[y] under the assumption of savvy model priors" (see Anderson, 2007, p. 88), or as the probability to be the *least false* model under the plausible assumption that the true model is unknown or too complex to be modeled (Claeskens and Hjort, 2008). Our choice of the *best* model will be therefore based on the probabilities given by Eq. (4). These probabilities provide information similar to the tests on thresholds proposed by Hansen (2000), but with the advantage, in the presence of multiple thresholds, of being a joint test and not a sequential test on each threshold.

3.2 Modelling Growth of Output Per Worker

Consider a set of countries indexed by i, $i = 1, ..., N$, partitioned into growth regimes indexed by m, $m = 1, ..., M$. Denote the set of countries in regime m as R_m. Growth is observed over a period of T years, indexed by t. Output per worker of country i at time t, y_{it}, can be expressed as:

$$y_{it} = y_{i0} e^{g_i t}, \tag{5}$$

where y_{i0} is the initial level of output per worker and g_i is the *annual* rate of growth.

The growth rate of country i is modeled by a semiparametric specification to account for nonlinearities, that is:

$$g_i = \alpha(m) + \sum_{j=1}^{K} \mu_j (X_{i,j}, m) + \upsilon_i, \tag{6}$$

where $\mathbf{X}_i = (X_{i,1}, ..., X_{i,K})$ is a collection of K proximate determinants, $\alpha(m)$ is a constant term for countries in R_m, $\mu_j(\cdot, m)$ are one-dimensional nonparametric functions operating on each of the K elements of $\mathbf{X}_i$ for countries in R_m, and υ_i is an error term with the properties: $E(\upsilon_i|\mathbf{X}_i) = 0$, $\text{var}(\upsilon_i|\mathbf{X}_i) = \sigma^2(\mathbf{X}_i, m)$ (i.e. the model allows for heteroskedasticity). The semiparametric specification allows for a varying marginal effect of proximate determinants on growth. As we will show in Section 3.3 the semiparametric specification is crucial for the correct identification of the proximate determinants' distributional effect, i.e. their effect on inequality and polarization in the distribution dynamics..

3.3 Decomposition of the Growth Rate

The starting point for the identification of the distributional effect of the *k-th* proximate determinant is the decomposition of the growth rate. In particular Eq. (6) can be rewritten as:

$$g_i = \alpha(m) + \sum_{j=1, j\neq k}^{K} \mu_j (X_{i,j}, m) + \mu_k (X_{i,k}, m) + \upsilon_i. \tag{7}$$

i.e.:

$$g_i = g_i^{-k} + g_i^k + g_i^r, \tag{8}$$

where g_i^{-k} is the growth rate of output per worker obtained by "factoring out" the effect of $X_{i,k}$, i.e. $g_i^{-k} = \alpha(m) + \sum_{j=1, j\neq k}^{K} \mu_j(X_{i,j}, m)$; $g_i^k = \mu_k(X_{i,k}, m)$ is the part of the annual growth rate explained by $X_{i,k}$, capturing the "marginal" effect of $X_{i,k}$ on g_i, that we denote as *marginal growth effect*; $g_i^r = \upsilon_i$ is the annual "residual growth", not explained by the determinants in $\mathbf{X}_i$.

3.4 Counterfactual Distribution

We will compute two types of counterfactual distribution to identify the distributional effect of, respectively, the proximate determinants and growth regimes. In particular, we model the distributional effect of a proximate determinant as determined by its sample distribution.

Let $\tilde{y}_{iT}^k$ denote the counterfactual output per worker for the *k-th* proximate determinant, i.e. the output per worker that country i would attain at T if there were no differences within the sample in the level of the *k-th* determinant. To identify this effect, we impose upon each country the *sample mean* of that determinant.[13]

Hence, the counterfactual growth rate of country i for the k-th proximate determinant, $\tilde{g}_i^k$, is defined as:

$$\tilde{g}_i^k \equiv \hat{\alpha}(m) + \sum_{j\neq k} \hat{\mu}_j(X_{i,j}, m) + \hat{\mu}_k\left(\bar{X}_k, m\right), \tag{9}$$

where $\bar{X}_k = N^{-1}\sum_{i=1}^{N} X_{i,k}$, and $\hat{\mu}_k(\cdot)$ is the estimated smooth function relative to the *k-th* determinant, obtained from the estimation of Eq. (6). Therefore, the counterfactual output per worker of country i at T is given by:

$$\tilde{y}_{iT}^k = y_{i0} e^{\tilde{g}_i^k T}. \tag{10}$$

The distribution of $\tilde{y}_{iT}^k$ is the counterfactual distribution with respect to the *k-th* determinant. Given the assumption on the existence of growth regimes, the effect of the *k-th* proximate determinant on the distribution dynamics is evaluated *within each regime*. The estimation of the counterfactual distribution for growth regimes is based instead on the idea of a random assignment of each country to one of the M regimes. Let $\tilde{y}_i^R$ denote the counterfactual output per worker for the growth regimes, i.e. the *expected value* of output per worker that country i would attain at T if, instead of belonging to a specific regime, it had a probability $1/M$ of belonging to one of the existing regimes.[14] In particular,

[13]If the determinant of interest is characterized by the presence of outliers, the median of the distribution could be preferable as a more robust measure. The use of the sample mean of the determinant aims to approximate its average effect on country growth. Other counterfactuals could be built using quantiles of the distribution. For example, Sirimaneetham and Temple (2009) compute counterfactual growth rates by imposing on each country of their sample the value of the determinant of interest (an index of macroeconomic stability) measured at the 95th percentile of the sample.

[14]An alternative counterfactual analysis corresponds to the case where regimes do not exist. When regimes exist, as we show below, its computation is however not feasible because such a case cannot be observed.

we compute the counterfactual growth rate of country i for growth regimes as:

$$\tilde{g}_i^R \equiv \frac{\sum_{m=1}^{M} \left[\hat{\alpha}(m) + \sum_j \hat{\mu}_j (X_{i,j}, m)\right]}{M}, \tag{11}$$

from which we obtain the counterfactual output per worker of country i:

$$\tilde{y}_{iT}^R = y_{i0} e^{\tilde{g}_i^R T}. \tag{12}$$

The distribution of $\tilde{y}_{iT}^R$ is the counterfactual distribution with respect to the growth regimes. In a pooled cross-section analysis, like the one we perform in Section 4, random assignment to regimes amounts to assuming random transitions across regimes in each sub-period considered.

3.5 Actual and Counterfactual Ergodic Distributions

The actual and counterfactual output per worker allow the actual and counterfactual ergodic distributions to be estimated, based on the actual and counterfactual stochastic kernels for each determinant and for growth regimes. In particular, the ergodic distribution shows 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.[15]

3.6 The Distributional Effect of Proximate Determinants

The distributional effect of a proximate determinant is evaluated by the *conditional* marginal growth effect, and by the differences between the actual and counterfactual distributions at time T and in the long run.

3.6.1 The Conditional Marginal Growth Effect

The effect of the *k-th* proximate determinant on the distribution dynamics is well captured by the relation between the *marginal growth effect* (MGE) of the *k-th* determinant in Eq. (8), g_i^k, and the initial level of output per worker y_{i0}, i.e. $g_i^k|y_{i0}$, that we denote as *conditional marginal growth effect* (CMGE) of the *k-th* determinant. It may be observed that the estimation of Eq. (6) must include all the explanatory variables in order to avoid omitted-variable problems and obtain an unbiased estimate of the marginal growth effect.

If $E[g_i^k|y_{i0}]$ is not statistically different from the expected value of the marginal growth effect, $E[g_i^k]$, i.e. if $E[g_i^k|y_{i0}] = E[g_i^k]\ \forall y_{i0}$, then the *k-th* determinant has no distributional effects. On the contrary, if $E[g_i^k|y_{i0}]$ differs statistically from $E[g_i^k]$ and, in particular, has everywhere an increasing (decreasing) relation with y_{i0}, then the *k-th* determinant is a source of divergence (convergence) within a regime.

[15] For details on the estimations see the on-line Appendix .8.

Figure 3 shows the case of $E[g_i^k|y_{i0}]$ decreasing in y_{i0}, which implies a more dispersed counterfactual distribution. Clearly, other types of CMGE can be observed. For example, if $E[g_i^k|y_{i0}]$ displays a

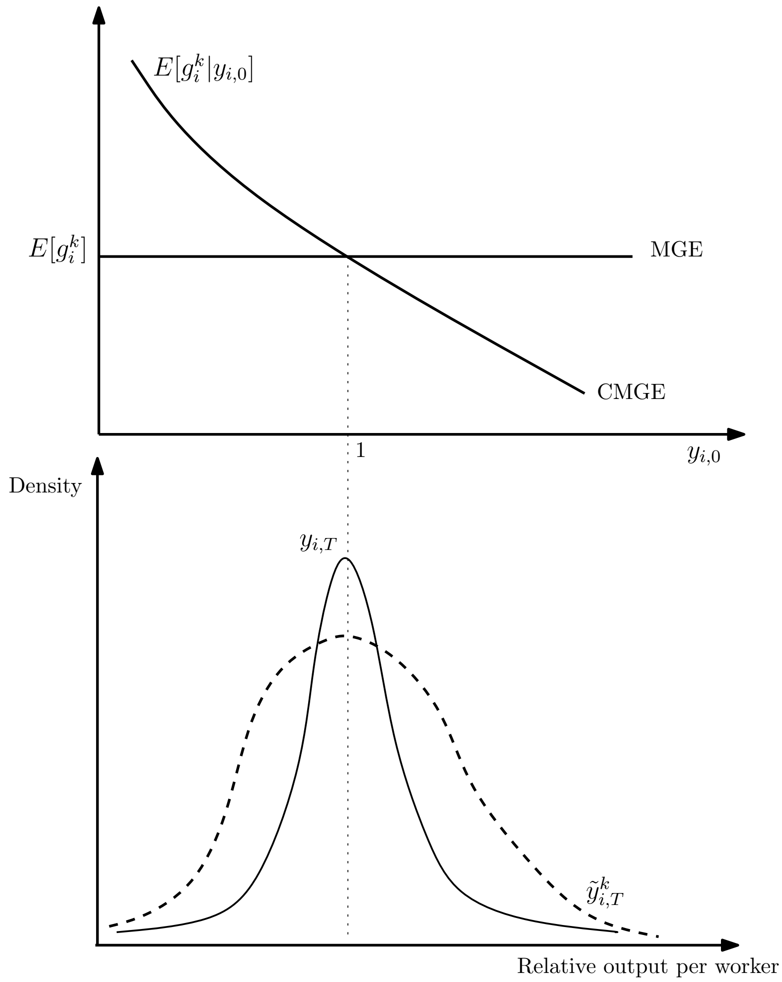


Figure 3: The case of the *k-th* determinant with a distributional effect in favor of a less dispersed distribution: CMGE is a decreasing function of initial level of GDP per worker.

"$\backsim$"-shaped form, the determinant is a potential source of polarization within the regime.

3.6.2 Nonlinearities and Differences between Actual and Counterfactual Distributions

In the presence of nonlinearities in the growth model, the *k-th* determinant can have an effect on the distribution dynamics even if the expected value of the CMGE is not statistically different from the expected value of the MGE, i.e. if $E[g_i^k|y_{i0}] = E[g_i^k]\ \forall y_{i0}$. In particular:

$$E\left[\log\left(y_{iT}\right)|y_{i0}\right] = E\left[\log\left(\tilde{y}_{iT}^k\right)|y_{i0}\right], \tag{13}$$

if:[16]

$$\sum_{m=1}^{M} E\left[\mu_k(X_{i,k}, m)|y_{i0}\right] = \sum_{m=1}^{M} \mu_k(\bar{X}_k, m). \tag{14}$$

[16]See on-line Appendix .2 for the derivation of the condition in Eq. (14).

The condition in Eq. (14) holds under the following two (sufficient) conditions:

1. $E\left[\mu_k(X_{i,k}, m)|y_{i0}\right] = E\left[\mu_k(X_{i,k}, m)\right]$, i.e. $\mu_k(X_{i,k}, m)$ and $y_{i,0}$ are mean-independent, i.e. the effect of the *k-th* determinant on output per worker in country i has to be independent of the initial output per worker in each regime m.

2. $E\left[\mu_k(X_{i,k}, m)\right] = \mu_k\left[E(X_{i,k}, m)\right] = \mu_k(\bar{X}_k, m)$, i.e. $\mu_k(\cdot, m) = \beta_k^m X_{i,k}^m$; the *marginal* effect of the *k-th* determinant has to be constant in each regime m, i.e. the term $X_{i,k}$ in growth regime m has a linear effect on growth.

Therefore, even if the CMGE of the *k-th* determinant is not statistically different from the MGE (i.e. Condition 1 holds),[17] such a determinant can have a distributional effect if it has a nonlinear effect on growth (i.e. Condition 2 fails). In growth empirics violations of Conditions 1 and 2 are common. For example Durlauf et al. (2001) find violations of Condition 1, while Liu and Stengos (1999) find violations of Condition 2. In Section 4 we show that also in our sample violations of these two conditions generally occur.

4 Explaining the Evolution of World Income Distribution

In this section we apply the method described in Section 3. In particular, in Section 4.1 we describe the dataset, in Section 4.2 we report the estimate of the best model, in Section 4.3 we use counterfactual analysis to investigate the distributional effect of proximate determinants and growth regimes and, finally, in Section 4.4 we provide a summary and a general discussion of our findings.

4.1 Data

Our sample consists of 84 countries for the period 1960-2008 (see Table 9 in on-line Appendix .3 for the country list). The dependent variable in the growth regressions is the average annual growth rate of GDP per worker.

Drawing on the vast literature discussed in the introduction we consider as candidate determinants of growth regimes five main types of "deep" determinants: *initial conditions*, i.e. the values in 1960 of GDP per worker, human capital (in particular the share of the workforce with primary or secondary education), and life expectancy at birth;[18] *geography*, proxied by the absolute latitude, the malaria

[17] Note that by definition $g_i^k \equiv \mu_k(X_{i,k})$.

[18] Life expectancy at birth is a typical proxy for the health conditions of a country. Education and health are widely considered within a broader concept of human capital (see, e.g. Mushkin, 1962, Sachs and Warner, 1997 and Weil, 2007).

ecological index, the percentage of tropical area, the land area within 100 km from the coast or navigable rivers, the average number of frost-days, and the proportion of land with five or more frost-days per month (see, e.g., Tan, 2009, Rodrik, 2002); the *quality of institutions*, proxied by the initial level of democracy or of constraints on the executive; a measure of *ethnolinguistic fractionalization* (see, e.g., Easterly and Levine, 1997); and *culture*, proxied by different shares of the population in 1965 belonging to the following religious denominations: Protestant, Catholic, Islam, Animist (see, e.g., Rodrik, 2002, Guiso et al., 2006, Durlauf et al., 2008). Following Mankiw et al. (1992) we consider as proximate determinants the initial level of GDP per worker, the investment rate, the growth rate of employment, and human capital, in the form of average years of schooling.[19]

Three remarks are in order on the sources of data. First, we choose PWT 7.1 instead of the more recent PWT 9.0 to maximize the number of countries available in the sample.[20] Second, we use the most recent version (2.0) of Barro and Lee (2013)'s dataset on human capital, in which many shortcomings of the previous versions have been eliminated (see Cohen and Leker, 2014, for details).[21] Finally, the number of countries in the sample is reduced with respect to its potential largest value based solely on data from PWT for the inclusion of institutions (with a reduction from a potential sample size of 109 to 90 countries) and, secondly, of human capital (with a reduction to 97 countries). Overall, the inclusion of both variables restricts the sample to 85 countries. Finally, the inclusion of life expectancy entails a further reduction to 84 countries.

4.2 The Best Model

In the estimation of the semiparametric growth model in Eq. (6), we pool cross-section data on five sub-periods: 1961-1970, 1971-1980, 1981-1990, 1991-2000, and 2001-2008. The dependent variable, g_{it}, is the average annual growth rate of GDP per worker of each sub-period. The proximate determinants are: i) the (log of) the initial level of GDP per worker of the sub-period, whose effect proxies for technological catch-up and/or decreasing marginal productivity of capital (log.y0); ii) the (log of) average annual growth rate of employment augmented by the depreciation rate and the exogenous rate of technological progress (equal to 0.03 and 0.02 respectively, see Mankiw et al., 1992) (log.n); iii) the (log of) the average annual investment rate (log.i/y); and, iv) the (log of) average years of schooling (log.h) as a proxy for the stock of human capital. Averages are computed over each sub-

[19]See on-line Appendix .1 for the definition, source and descriptive statistics of the variables.

[20]The use of PWT 9.0 would limit the number of countries to 61; the reason for this marked reduction is the different use of the many rounds of the International Comparison Program, see: `http://www.rug.nl/research/ggdc/data/pwt/v80/comparing_pwt80_with_pwt71.pdf` for more details.

[21]The use of one of the most important alternative datasets, proposed by Cohen and Soto (2007), yields measures of human capital highly correlated with those used in the paper (never below 0.91), but it would reduce the sample to 79 countries.

period. The growth model includes time dummies to account for possible changes across sub-periods in the exogenous growth rate of technological progress.

Proximate determinants in growth regression are likely to be endogenous for several reasons, in particular for: simultaneity (when an explanatory variable is jointly determined with the dependent variable, typically because both variables depend on an omitted explanatory variable) and measurement error. The identification of valid and strong instruments is highly debated in the growth empirics literature. Durlauf et al. (2005, p. 638-639) point out that: "the belief that it is easy to identify valid instrumental variables in the growth context is deeply mistaken. We regard many applications of instrumental variable procedures in the empirical growth literature to be undermined by the failure to address properly the question of whether these instruments are valid [...] Since growth theories are mutually compatible, the validity of an instrument requires a positive argument that it cannot be a direct growth determinant or correlated with an omitted growth determinant". Bazzi and Clemens (2013) provide evidence on ways instruments that are valid in some studies can be invalid in others, and show the ways in which plausibly valid instruments can mask important weak instrument biases.

In the estimation of growth regressions within each regime, we control for the presence of endogeneity in all proximate determinants (except for the initial level of GDP per worker) using as instruments their value in 1960. Although we expect such instruments to be relevant and strong, some concerns about their validity are present. Growth has been extensively related to initial conditions and initial stocks of human capital (e.g. Cohen, 1996 and Goetz and Hu, 1996). Moreover, the initial levels of investment rate and employment growth could easily be correlated with omitted growth determinants, and therefore our instrumental variables could be correlated with omitted growth determinants. However, in our analysis some of these potentially omitted determinants are likely included in the candidate set of deep determinants, and the use of a semiparametric specification reduces the possibility that model misspecification would lead to endogeneity. Moreover, we provide a test of omitted-variable bias due to initial conditions proxied by the initial level of GDP per worker (see on-line Appendix .7 for details). Given the presence of semiparametric additive components, we used the Control Function Method (CFM) instead of two-stage least squares (see on-line Appendix .6).

4.2.1 Growth Regimes

The best (approximating) model among all those that were fitted, identified following the procedure described in Section 3.1, contains three growth regimes based on life expectancy at birth in 1960 and the share of the Catholic population in 1965: a "high life expectancy regime" (*Regime H*) comprising 20 countries; a "low life expectancy/high share of Catholics" regime (*Regime LH*) comprising 47 countries, and, "low life expectancy/low share of Catholics" regime (*Regime LL*), comprising 17 countries (see Figure 4). In particular, the threshold for life expectancy is 68.35 years, while the threshold for the share of Catholics is 0.03 for the "low life expectancy" countries. Table 9 in on-line Appendix .3

contains the list of countries in the three regimes. The best model has a 99% probability of being the best (approximating) model among all those that were fitted, as shown in Table 8 in on-line Appendix .3. Moreover, taking life expectancy and the share of Catholics as partitioning variables, in order to check the robustness of the thresholds we calculate the (*conditioned*) probabilities of the models estimated for all possible partitions of countries being the least false. Our best model has about a 30% probability of being the least false, and only six alternative partitions have more than 5% probability but involve only marginal changes in the thresholds (see Figure .9 in on-line Appendix .3).

The importance of taking into account the possible existence of growth regimes can be appreciated by comparing the value of AICc of our best model, equal to 2230.49, with that of the pooled regression without regimes equal to 2081.81 (the implied probability of the pooled regression being the least false model is about zero).[22] Moreover, if we consider the deep determinants as proximate determinants (i.e. they are included as covariates), as is commonly done in the literature, AICc surges to 2100.59 when only life expectancy and the share of Catholics are considered, and to 2102.62 with the additional inclusion of one variable for each type of deep determinant, i.e. latitude, democracy, and ethnolinguistic fractionalization (their implied probabilities of being the least false model are both about zero). Overall, this evidence suggests that growth regimes are strongly informative on country dynamics, and our deep determinants, namely life expectancy and share of Catholics, contain (almost) all the information of the set of candidate deep determinants.

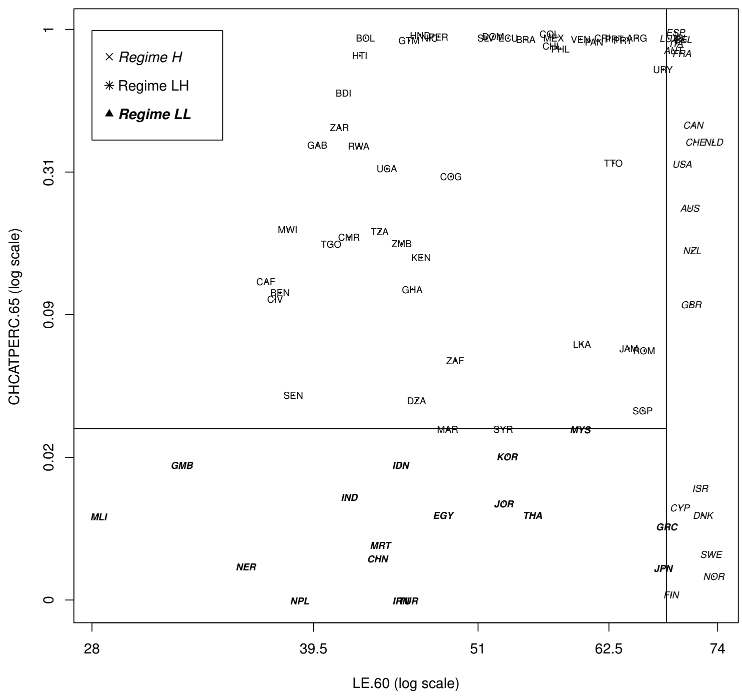


Figure 4: Partition of countries into three growth regimes.

[22]Results are available upon request.

Regime H mainly includes Western countries and countries from the Western offshoots; Regime LH comprises two European countries (Portugal and Romania), some Arab countries, all Central and South American countries, many Sub-Saharan countries and South Africa, and Sri Lanka, the only Asian country; Regime LL mainly contains Asian countries, especially from the Middle East and South-East Asia, two Sub-Saharan countries, and Greece. The three growth regimes can be ordered in terms of their average relative GDP per worker in 1960: 2.35 (H), 0.63 (LH), and 0.44 (LL). Their average growth rate of GDP per worker over the period was respectively 2.1%, 1.1%, and 2.6%, suggesting that convergence only occurred between Regimes LL and H. Indeed, in 2008 their average relative GDP per worker became, with respect to the sample mean, 2.42 (H), 0.51 (LH), and 0.70 (LL), showing that, on average, countries in LL overtook countries in LH.

The identified regimes are not strictly related to long-run outcomes, but they correspond to different growth models, e.g. having a high life expectancy at birth in 1960 or a certain share of Catholics in 1965, or otherwise, is not unambiguously related to experiencing a high or low growth rate or convergence to a certain GDP level. Only in the case of high-life expectancy at birth is there clear-cut identification of highly-developed countries. In this case culture does not partition countries.[23] Life expectancy, in particular, prevails over the quality of institutions for these countries, perhaps not surprisingly in the light of the evidence discussed in Weil (2014) on the primacy of health for the development of countries. Regime LH highlights the emergence of a similarity based on the share of Catholics for countries from Africa and South America, suggesting that the widespread use of continental dummies in growth empirics might not be fully appropriate.

Our results are in contrast with Owen et al. (2009), Tan (2009), and Flachaire et al. (2014) who, adopting different methods, find a primacy of institutions on the initial level of human capital and geography in the identification of (only two) growth regimes. Moreover, unlike Tan (2009), we do not find that ethnic fractionalization identifies growth regimes. However, none of these works included religion among the possible regime identifiers. In this respect, in their seminal work on growth regimes Durlauf and Johnson (1995, p. 378) point out that some anomalies of their partition into four growth regimes may be explained by omitted initial conditions, such as social capital, that should proxy for: "cultural norms and values ... which may range from attitudes towards work to respect of property rights". Surprisingly, subsequent work on growth regimes mainly ignored this remark and, to the best of our knowledge, the importance of culture as a possible deep growth determinant, has been overlooked.

Figure 5 presents a re-examination of the tendency of polarization reported in Figure 1 as the result of the distribution dynamics between and within the three growth regimes. Table 1 shows that

[23]This result is in contrast with Desdoigts (1999) who finds a cluster of developed countries in which a partition into Protestant and Catholic groups emerges.

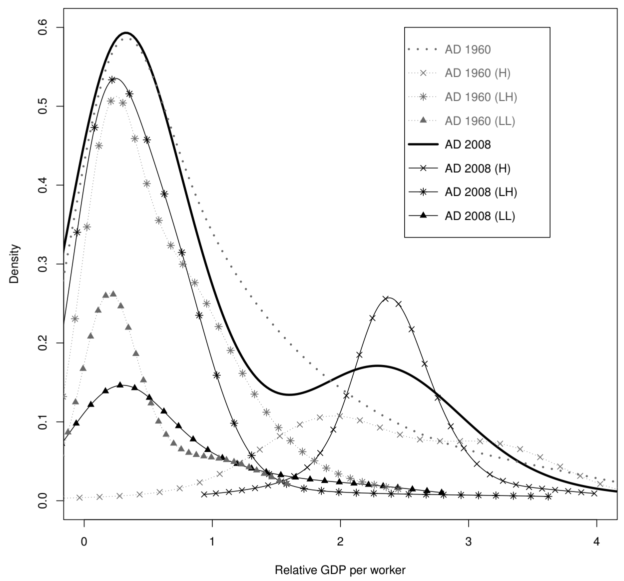


Figure 5: Estimated distributions of GDP in 1960 and 2008 for the whole sample of countries (black), and for each growth regime (blue, orange, and red for Regimes H, LH, and LL respectively).

Theil	Total	Between-group	Within-group
AD 1960	0.42 (0.04)	0.24 (0.05)	0.18 (0.03)
AD 2008	0.47 (0.05)	0.27 (0.05)	0.20 (0.04)
BIPOL			
AD 1960	NA		
AD 2008	0.76 (0.17)		

Table 1: Variable: GDP per worker. Theil index of total, between-group, and within-group inequality and BIPOL index of polarization in 1960 and 2008. Bootstrap standard errors are reported in parenthesis.

inequality, measured by the Theil index, increased by 5 percentage points between 1960 and 2008. In particular, both the between- and within-group components show a moderate increase, while the between-group component accounts for the largest share of inequality in both years. Polarization is a phenomenon emerging only at the end of the period: the BIPOL index is in fact not computable for the distribution in 1960 which is clearly unimodal.

Figure 5 shows that the emergent polarization is the result of a strong tendency for convergence within Regime H, and a substantial immobility of the distribution of countries in Regime LH, characterized only by a slight tendency to within-group convergence. Countries in Regime LL, although starting from a lower average initial GDP per worker than those of Regime LH, display a tendency to spread out on a larger GDP range, especially due to the presence of some fast-growing economies such as China and South Korea. We interpret this evidence as supporting the presence of *club convergence*, where club membership is determined by the share of Catholics in 1965 and life expectancy at birth in 1960. Anderson et al. (2016) find a similar dynamic among three income groups identified by a finite mixture model. In particular, their analysis of transitions in 1970-2010 reveals that convergence only occurred between the low-income and middle-income groups.

To further characterize the regimes, Table 2 reports the mean and standard deviation of the distribution of proximate determinants for the whole sample and within each regime (see also Figure .10 in on-line Appendix .4).[24] Initial income is on average higher and less dispersed in Regime H, while no significant differences exist in the average levels of Regimes LH and LL, but the distribution is different as two peaks characterize Regime LH. The main aspect of the distributions of the employment growth rate is that it is on average lower in Regime H. No noteworthy differences appear in the investment rate across Regimes, while human capital appears clearly higher on average and less dispersed in Regime H.

Among the identified regimes, Regime H has the characteristics predicted by UGT of Galor and Weil (2000): high income, low employment growth (which may proxy for low population growth), and, overall, high human capital levels.[25] The striking feature of our results is that the variable that better identifies this regime is life expectancy at birth in 1960, supporting the idea that a sufficiently high level of health is a necessary condition for the accumulation of human capital. However, the characteristics of the other two regimes do not support the prediction of UGT, in particular we do not find any difference in their demographic patterns.

[24]Feyrer (2008) and Johnson (2005) proposed to explain the income distribution dynamics by the distribution of growth determinants. We will see that, without accounting for nonlinearities, this approach can be misleading.

[25]See also Kuznets and Murphy (1966) on the concept of *modern growth*.

		Whole Sample	Regime H	Regime LH	Regime LL
Initial GDP per worker	*Mean*	19535	46997	10760	11484
	SD	19990	17182	10087	13530
(Augmented) Employment growth	*Mean*	0.07	0.06	0.08	0.07
	SD	0.01	0.01	0.01	0.01
Investment rate	*Mean*	0.23	0.24	0.21	0.24
	SD	0.09	0.05	0.09	0.11
Human capital	*Mean*	5.66	9.07	4.67	4.38
	SD	3.07	1.93	2.36	2.98

Table 2: Mean and standard deviation of distribution of proximate determinants for the whole sample and within each regime.

4.2.2 Semiparametric Growth Regressions

Table 3 reports the estimation results of the semiparametric growth model in Eq. (6) within each regime. In each estimation exogeneity cannot be rejected at 5% significance level.[26] As expected, results of the first-stage regressions show that almost all the instruments are significant (Tables 10-**??** in on-line Appendix .6). However, as discussed in Section 4.2, our instruments could be invalid. Although this hypothesis cannot be formally tested, we find some evidence in favor of its validity. Ashley and Parmeter (2015) quantify the minimum degree of correlation between the possibly-endogenous variables and the model errors which is sufficient to overturn the inference on the regression parameters. By applying their method to the model of Mankiw et al. (1992) (as in our case), they find that quite substantial correlations are necessary to reverse the inference on the estimated parameters, concluding that in such a case the need for valid instruments is mitigated. Moreover, in a recent paper Guo et al. (2016) study the properties of the endogeneity test under invalid instruments and find that if some instruments (even a single one) are moderately (or strongly) invalid, then the endogeneity test will always reject the null hypothesis of exogeneity even if there is truly no endogeneity present. Accordingly, if our instruments were invalid we would have always rejected the null hypothesis of exogeneity. Finally, no omitted-variable bias seems to be present in the best model at 5% significance level (See on-line Appendix .7).

The goodness of fit, measured by generalized R^2, is fairly high in all the regimes, ranging from 0.43 in Regime LH to 0.74 in Regime H.[27] A comparison of the estimation of a growth model without regimes, reported in the first column of Table 3, and the models estimated within each regime,

[26] See on-line Appendix .6 and on-line Appendix .9 for the details on the estimations.

[27] The better fit of the augmented Solow model in the group of the most developed countries is found in other studies (see, e.g., Durlauf and Johnson, 1995, p. 375, Tan, 2009, p. 1119, Owen et al., 2009, p. 276) which, however, do not allow for nonlinearities within the regimes.

	No Regimes	Regimes		
		Regime H	**Regime LH**	**Regime LL**
Dep. Var: g	Pooled GAM **1960-2008**	Pooled GAM **1960-2008**	Pooled GAM **1960-2008**	Pooled GAM **1960-2008**
Parametric coefficients:	Estimate	Estimate	Estimate	Estimate
Intercept	0.034***	0.028***	0.030***	0.046***
	(0.002)	(0.003)	(0.004)	(0.005)
$D_{1970-1980}$	−0.014***	−0.013***	−0.011***	−0.017***
	(0.003)	(0.003)	(0.004)	(0.006)
$D_{1980-1990}$	−0.029***	−0.010***	−0.038***	−0.023***
	(0.003)	(0.003)	(0.005)	(0.007)
$D_{1990-2000}$	−0.025***	−0.003	−0.028***	−0.031***
	(0.004)	(0.004)	(0.005)	(0.007)
$D_{2000-2008}$	−0.018***	−0.011**	−0.016***	−0.027***
	(0.004)	(0.004)	(0.005)	(0.007)
Semi-parametric coefficients:	EDF	EDF	EDF	EDF
log.y0	2.6***	1.7***	1.0**	4.3***
	(9.88)	(10.67)	(4.96)	(10.6)
log.n	1.0***	1.0**	2.3***	2.2***
	(49.83)	(5.96)	(5.06)	(15.6)
log.i/y	1.0***	2.1**	2.1***	2.3
	(52.34)	(3.49)	(10.95)	(1.39)
log.h	2.1**	1.0	1.9	1.0***
	(2.64)	(2.52)	(1.53)	(0.009)
Endogeneity	NO	NO	NO	NO
Omitted-variable bias	NO	NO	NO	NO
Observations	420	100	235	85
Countries	84	20	47	17
Generalized R^2	0.40	0.74	0.43	0.69
Scale estimate (*10^{-5})	39.7	6.7	42.2	29.6
REML score	-1001.4	-291.0	-535.28	-179.94
AICc	2081.81	2230.49		

Table 3: Estimates of the semiparametric growth model in Eq. (6). Significant asymptotic levels: 1%"***" 5%"**" 10%"*". Standard errors and F-values are reported between brackets for parametric and semiparametric coefficients respectively. **GAM**: Generalized Additive Model. **EDF**: estimated degrees of freedom in the estimate of $\mu_j(\cdot)$. **Endogeneity**: test on the presence of endogeneity and endogeneity-robust estimation via Control Function (see on-line Appendix .6). **Omitted-variable bias**: test for omitted-variable bias with distributional effects (see on-line Appendix .7). **Generalized $\mathbf{R^2}$**: generalization of R^2 to be used in ML estimates (see Nagelkerke, 1991). **Log.likelihood**: the logarithm of model's likelihood. **Scale estimate**: scale parameter (corresponding to the residual variance of the estimation, see on-line Appendix .9). **REML score**: score of the restricted maximum likelihood estimation (it provides the fundamental information on the specification of the model, see on-line Appendix .9). **AICc**: Akaike Information Criterion calculated as in Eq.(1) in on-line Appendix 3.1.1.

shows that not accounting for growth regimes represents a serious misspecification of the model. The estimation of the regime-specific growth models reported in columns 2-4 of Table 3 highlights substantial parameter heterogeneity across regimes, both in terms of magnitude of non-explained growth (see the estimated values of the intercept and of the time dummy coefficients), and nonlinearities. In particular, the time average of non-explained growth, which reflects total factor productivity (TFP) growth, is equal to 2.1% in Regime H, to 1.1% in Regime LH and to 2.6% in Regime LL.[28] The values of the time dummies also show that countries in Regimes LH and LL seem to be more sensitive to shocks than countries in Regime H. All proximate determinants are statistically significant at the usual significance levels, with the exception of the investment rate in Regime LL and human capital in Regimes H and LH.

Figure 6 reports the effects of each proximate determinant on growth of GDP per worker, along with the estimated effect of each determinant in the model with no regimes (labeled as *pooled*). The relation between initial per worker GDP and the growth rate highlights the tendency to within-regime convergence. This tendency is clear in Regimes H and LH, although the estimated function is steeper in Regime H, indicating a higher speed of convergence. The estimated function is instead concave in Regime LL, indicating that a moderate but not-uniform tendency to within-regime catch-up characterizes this regime. Employment growth has the expected negative marginal effect on growth in all regimes, with nonlinearities in Regimes LL and LH. The effect of the investment rate on growth is non-significant in Regime LL, while it is nonlinear in Regimes LH and H. Given the large confidence bands, however, in the latter two regimes the effect is likely to be non-increasing in relevant ranges of the variable. Finally, human capital has a clear, positive marginal effect on growth in Regime LL alone.[29]

4.3 Distributional Effects

In this section we present the estimated distributional effect of each proximate determinant and of growth regimes, while we refer to Section 4.4 for a general discussion of our findings.

Table 4 shows that the counterfactual distribution of initial GDP per worker is characterized by a much higher value of the Theil index, implying that initial GDP per worker considerably reduces inequality: if all countries had had the same value of initial GDP per worker, inequality would have been much higher. This effect is mainly due to the between-group component, which would have

[28]The time-averaged non-explained growth in regime H is calculated as the weighted average of the following values: 2.8%, 1.5%(=2.8%-1.3%), 1.8%(=2.8%-1.0%), 2.8%(=2.8%-0%) and 1.7%(=2.8%-1.1%), with weights 10/48, 10/48, 10/48, 10/48 and 8/48 respectively. Non-explained growth in the other regimes is computed in the same way.

[29]For Regime H this may reflect the sorting of the countries, which have reduced the cross-country variation in human capital within this regime.

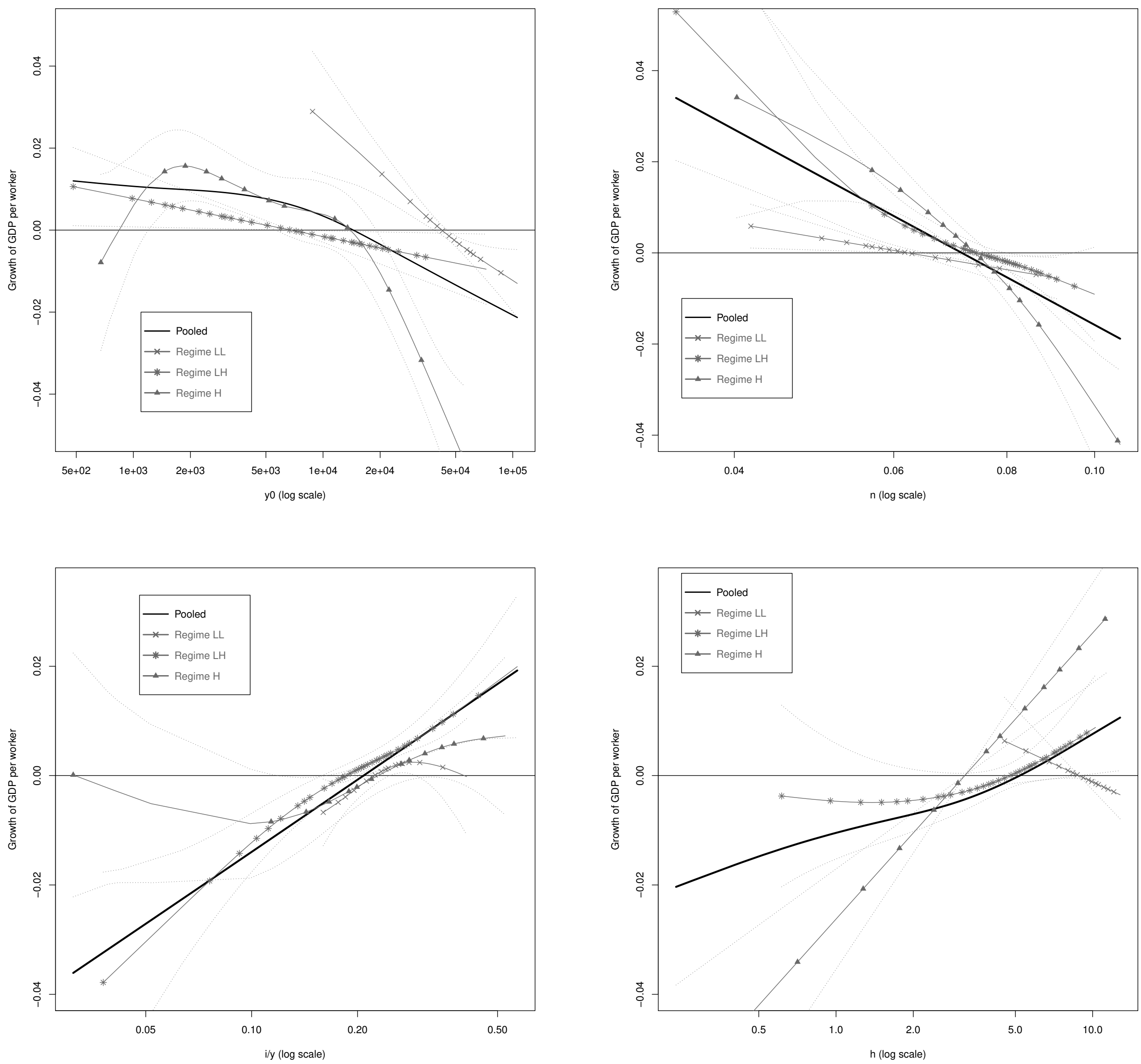


Figure 6: Estimated nonparametric function $\mu_j(\cdot)$ of Equation (6) for model without regimes (light grey line, corresponding to columns 1 in Table 3) and for the best model with three growth regimes (blue, orange and red lines, corresponding to columns 2-4 in Table 3 respectively). 95% confidence bands (dotted lines) are derived from the estimated standard errors based on the Bayesian posterior covariance matrix of the parameters (see Wood, 2011).

been three times higher. Figure 7a shows that, in the counterfactual distribution for their regime, countries in Regime H would have had a much higher level of GDP per worker. In particular, given that the conditional marginal growth effect in Regime H has a steep negative slope (see Figure .11a), "assigning" the sample average value of log.y0 to all countries in that regime would amount to assigning to these countries a much higher growth rate than what most of them actually experienced. The counterfactual distribution of Regime LH is not very different from the actual one, while for Regime HH the counterfactual distribution shows that some countries would have been even further away from the others with very high income levels. Moreover, initial GDP per worker strongly reduces polarization: Table 4 shows that in 2008 the polarization index is much higher in the counterfactual distribution. The same tendency is confirmed for the long run, as illustrated by a comparison of the BIPOL index for the actual ergodic distribution (AED) and that computed for the counterfactual ergodic distribution (CED), as well as their graphical representation in Figure 7d.

The growth rate of employment moderately increases inequality as the Theil index and its between-group component are slightly higher in the actual than in the counterfactual distribution (see Table 4). Overall the effect is small, as shown by the negligible differences between actual and counterfactual distributions in Figure 7b and the almost flat conditional marginal growth effects reported in Figure .11b. Also, employment growth moderately acts in favor of polarization, as shown by the values of the BIPOL index. The effect on polarization is more pronounced in the comparison between the actual and counterfactual ergodic distributions (see Figure 7e).

The investment rate appears to slightly increase inequality, as the Theil index is higher in the actual than in the counterfactual distribution. This mainly appears to depend on within-group inequality (see Table 4). However, the investment rate reduces polarization, as the BIPOL index is lower in the actual distributions (both in 2008 and in the long run, see also Figure 7f). Overall, the effect is modest (see Figures 7c and .11c).

Human capital tends to marginally decrease inequality and polarization (see Table 4). Examination of the Theil index reveals that human capital reduces between-group inequality, but increases within-group inequality. In fact, due in particular to the strong positive effect in Regime LL (Figure .11d), human capital contributed to the growth and catch-up of these countries, which would have been otherwise more dispersed (see Figure 8a). In the long run, human capital generated a less-polarized distribution than the one that would have obtained if all countries shared the same human capital value (Figure 8c).

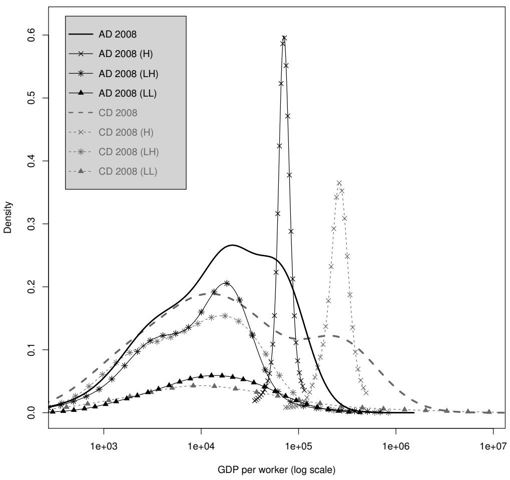
Growth regimes are a major source of inequality and, especially, of polarization. Table 4 shows that both components of the Theil index are lower in the counterfactual than in the actual distribution. In other words, if countries were allowed to randomly switch among regimes in each sub-period, the distribution of the expected value of their counterfactual GDP per worker would have displayed less inequality. Figure 8b highlights that countries in Regime LH would have displayed a much higher

	Variable: log.y0			**Variable: log.n**		
Theil	Total	Between-group	Within-group	Total	Between-group	Within-group
AD 2008	0.47	0.27	0.20	0.47	0.27	0.20
	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)
CD 2008	0.90	0.63	0.27	0.41	0.23	0.18
	(0.09)	(0.10)	(0.08)	(0.05)	(0.04)	(0.03)
BIPOL						
AD 2008	0.76			0.76		
	(0.17)			(0.17)		
CD 2008	0.93			0.65		
	(0.74)			(0.41)		
AED	1.26			1.26		
	(0.04)			(0.04)		
CED	6.05			2.06		
	(0.40)			(0.03)		

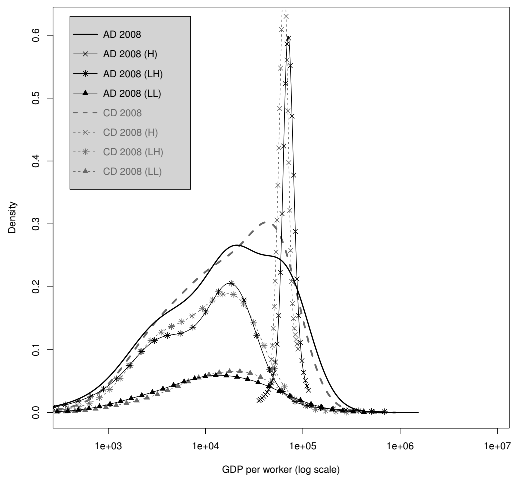
	Variable: log.i/y			**Variable: log.h**		
Theil	Total	Between-group	Within-group	Total	Between-group	Within-group
AD 2008	0.47	0.27	0.20	0.47	0.27	0.20
	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)	(0.04)
CD 2008	0.43	0.28	0.15	0.52	0.42	0.09
	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.02)
BIPOL						
AD 2008	0.76			0.76		
	(0.17)			(0.17)		
CD 2008	0.82			0.90		
	(0.23)			(0.30)		
AED	1.26			1.26		
	(0.04)			(0.04)		
CED	2.02			3.31		
	(0.02)			(0.04)		

	Variable: growth regimes		
Theil	Total	Between-group	Within-group
AD 2008	0.47	0.27	0.20
	(0.05)	(0.05)	(0.04)
CD 2008	0.24	0.10	0.14
	(0.03)	(0.02)	(0.02)
BIPOL			
AD 2008	0.76		
	(0.17)		
CD 2008	NA		
	(NA)		
AED	1.26		
	(0.04)		
CED	NA		
	(NA)		

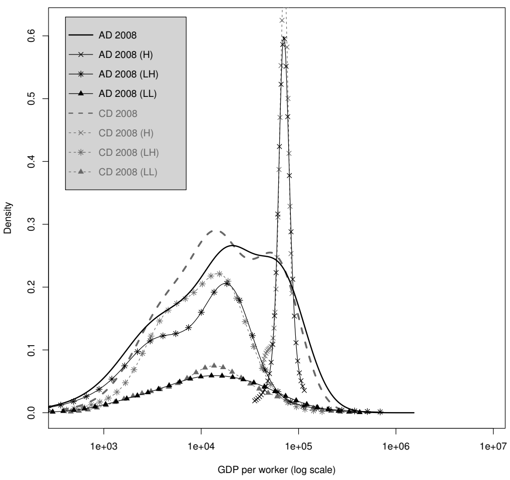
Table 4: Theil index of total, between-group, and within-group inequality and BIPOL polarization index in 1960 and 2008 for the actual, ergodic, and counterfactual distributions. Bootstrap standard errors in parenthesis.



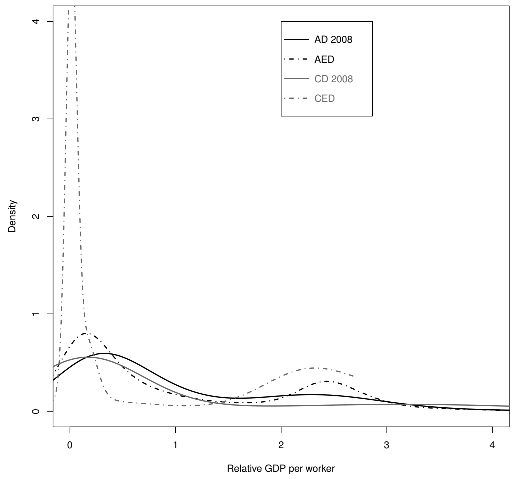
(a) Variable: log.y0. Actual and counterfactual distributions in 2008 (whole sample and growth regimes).



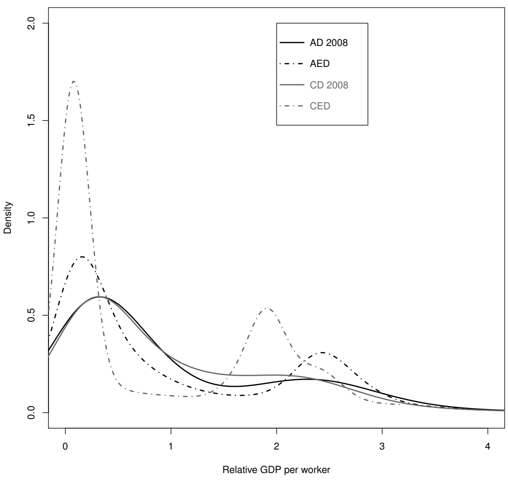
(b) Variable: log.n. Actual and counterfactual distributions in 2008 (whole sample and growth regimes)



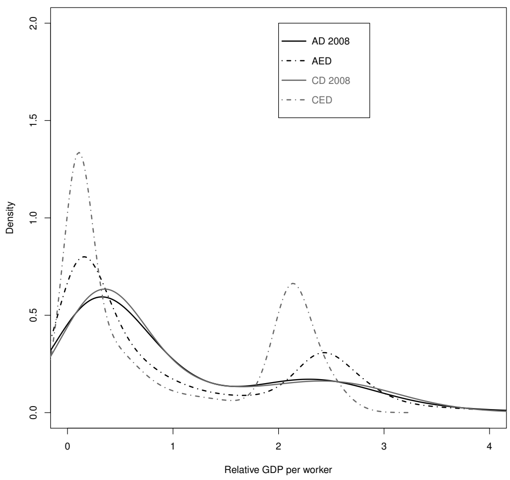
(c) Variable: log.i/y. Actual and counterfactual distributions in 2008 (whole sample and growth regimes)



(d) Variable: log.y0. Actual and counterfactual final and ergodic distributions.

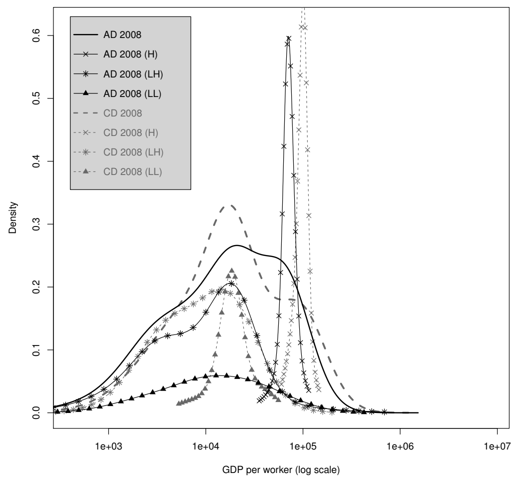


(e) Variable: log.n. Actual and counterfactual final and ergodic distributions.

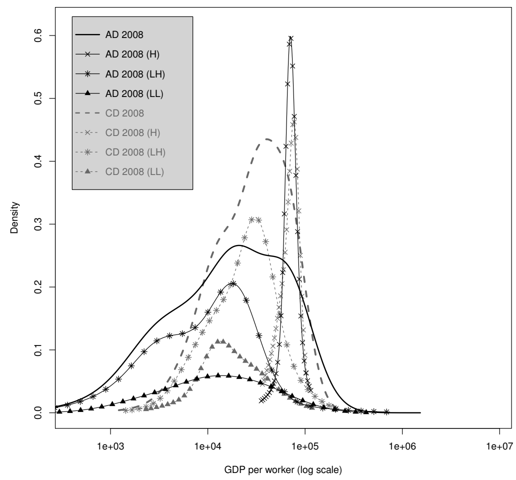


(f) Variable: log.i/y. Actual and counterfactual final and ergodic distributions.

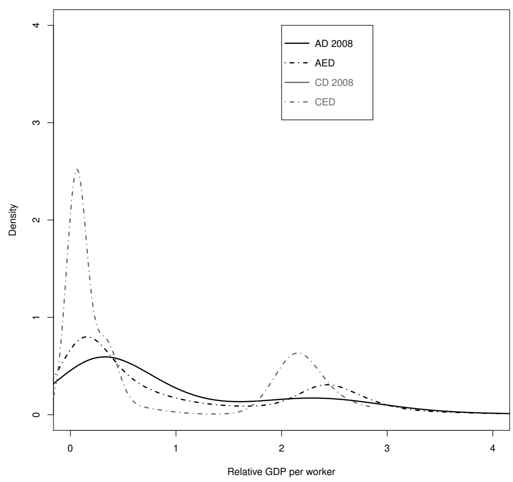
Figure 7: Actual and counterfactual distributions.



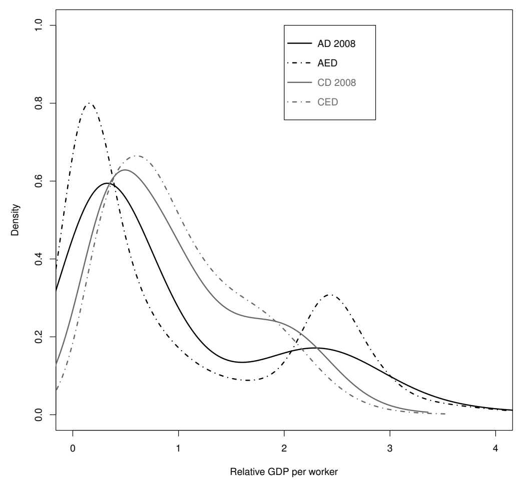
(a) Variable: log.h. Actual and counterfactual distributions in 2008 (whole sample and growth regimes)



(b) Growth regimes. Actual and counterfactual distributions in 2008 (whole sample and growth regimes)



(c) Variable: log.h. Actual and counterfactual final and ergodic distributions.



(d) Growth regimes. Actual and counterfactual final and ergodic distributions.

Figure 8: Actual and counterfactual distributions.

mobility onwards, while countries in Regime LL would have been much less dispersed. The most striking result, however, is that the counterfactual distribution of 2008 and the counterfactual ergodic distribution do not show evidence of polarization, as shown by the dynamics displayed in Figure 8d.

4.4 Discussion of Results

Our findings contribute to the debate on the evolution of the cross-country income distribution in many respects. First, the roots of the observed increase in inequality and polarization do not seem ascribable to the traditional Solovian growth determinants, i.e. the accumulation of physical capital and employment growth, but to the existence of growth regimes, i.e. of different growth processes followed by countries. This result is in contrast with, among others, Beaudry et al. (2005) and Feyrer (2008).

Three main differences among regimes emerged: i) the levels of TFP growth are remarkably heterogeneous. TFP growth is very similar in Regime H and Regime LL (equal to 2.1% and 2.6% respectively), while in regime LH it is approximately half the value of the other regimes (1.1%); ii) the conditional marginal growth effect of initial GDP, a proxy for technological catching-up, is decreasing in Regimes H and LH and nonlinear in Regime LL; iii) the marginal growth effect of human capital is significant and strongly increasing in GDP per worker only in Regime LL.

The result at point i) offers a novel view on religion, viewed by economists as one of the primary determinants of culture (see, e.g., Weil, 2012, p. 436 and Guiso et al., 2006), although its impact on growth is a controversial issue (Guiso et al., 2006): the influential study of Barro and McCleary (2003) for example finds that some measures of religious behavior significantly affect growth, while, e.g., Durlauf et al. (2012) subsequently played down the role of religion.

We find that religion is associated with substantially different levels of TFP growth only for countries with low life expectancy in 1960. This result is consistent with the claim of Guiso et al. (2006) according to which: "[the] dependence of [growth] on cultural variables weakens for more educated people, consistent with the idea that more educated individuals rely less on their inherited culture when they form their priors." In our case, the significant dimension of human capital is not education but health. Among the countries with low life expectancy, i.e. countries where inherited culture could crucially affect individual decisions, we find that Catholic religion appears associated to lower TFP, suggesting in our view a lower capacity of adopting foreign technology and/or developing new technology. These countries therefore appear endowed with low levels of *social capability*, a concept introduced by Abramovitz (1986), referring to the capacity of a country to introduce new ideas and to exploit existing ones, to capture economic opportunities, etc. This evidence complements the findings of Guiso et al. (2006) on the importance of culture, as proxied by trust, for economic development.

Our finding of a negative correlation between Catholic religion and TFP levels, especially in Regimes LH and LL, seems at odds with the historically high Catholic propensity to establish education institutions (see, e.g., Bader and Maussen, 2012 for Europe) as education, by increasing human capital accumulation, should favor technology development and/or adoption.[30] Becker and Woessmann (2009) make a similar point with respect to the diffusion of Protestantism. They argue that the spread of Protestantism implied the diffusion of education to promote literacy development, and that this, and not the spread of the "Protestant ethic", fostered economic growth. However, our results do not support the "human capital view" on the role of religion on growth. In fact, in our sample PRI.60 has a positive correlation both with PROT.65 and CAT.65, but only for the former is it high and significant (in the same respect see Figure II of Becker and Woessmann, 2009),[31] while SEC.60 has a positive correlation only with PROT.65.[32] However, PROT.65 is not found as a significant regime identifier.

The result at point ii) highlights the fact that technological catch-up occurs within all regimes, but at different speeds. In Regime H the speed of convergence is high and uniform for all countries (the slope of the estimated relationship is almost constant); in Regime LH the speed of convergence is uniform but almost nil; finally, in Regime LL, the speed of convergence is nil for very poor countries and very high for the richest. This evidence is consistent with the differences observed in TFP growth among regimes.

The result at point iii) instead supports the insight of Nelson and Phelps (1966) that the key role of human capital is to facilitate the adoption of technology and not to be a productive factor *per se*. In particular, we find a positive and significant marginal effect of human capital on growth in Regime LL only (see Figure 6), the only regime which seems to have enjoined significant technological spillovers from Regime H, with a very similar level of TFP growth. Human capital in the form of education, therefore, appears an important growth determinant for countries with low life expectancy at birth. In addition, if the major determinant of long-run growth is TFP (as argued, for example, by Hall and Jones, 1999), the presence of considerable differences in TFP across regimes casts doubt on the primacy of institutions as a fundamental driver of long-term development (Acemoglu et al., 2005), as we did not find institutions as a primary regime identifier. The higher relevance of culture and health with respect to geography and institutions is consistent with Spolaore and Wacziarg (2013, p. 341)'s claim that: "human traits are important to account for comparative development patterns, quite apart from the effects of geographic and institutional factors". An important caveat is that they refer to long-term development, while we focus on a more recent and shorter period.

[30] We thank an anonymous referee for pointing this out.

[31] Bivariate regressions of PRI.60 on PROT.65 and CAT.65 return coefficients of, respectively, 0.57 (s.e. 0.14) and 0.21 (s.e. 0.08).

[32] Bivariate regressions of SEC.60 on PROT.65 and CAT.65 return coefficients of, respectively, 0.27 (s.e. 0.08) and 0 (s.e. 0.05).

Finally, we document that in the period 1960-2008 inequality and polarization across countries increased, and such a tendency is expected to continue in the long run. Specifically, the counterfactual analysis suggests that the persistent nature of the twin-peaked distribution is to be attributed to the existence of regimes and to the persistence of countries within each regime: if transitions across regimes were allowed, the long-run distribution would have been single-peaked. In other words, the estimate of a long-run polarized distribution suggests that no significant transitions across regimes have been taking place in the period of analysis. This evidence challenges the idea that polarization is a transitory phenomenon, as pointed out by Lucas (2000) and Galor (2007). Our evidence of a persistent twin-peaked distribution, on the contrary, is in line with the much-discussed "middle-income trap", according to which many episodes of growth spurts by initially poor countries suddenly stop before the complete catch-up with the richest countries has been achieved (see, e.g., Eichengreen et al., 2012, World Bank, 2013, and Pritchett and Summers, 2014). With a different method Anderson et al. (2016) arrive at a similar conclusion: in the period 1970-2010 they find catch-up from the low- to the middle-income class, but not from the middle-income to the high-income class.

5 Concluding Remarks

In this paper we contributed to the literature on growth empirics by proposing a new method based on information theory which jointly identifies growth regimes and estimates a semiparametric growth model within each regime. We applied our method to a sample of countries in the period 1960-2008, which experienced an increase in inequality and polarization in the distribution of GDP per worker. We found three growth regimes, identified by life expectancy in 1960 and the share of Catholics in 1965. Countries in each regime follow specific nonlinear "augmented" Solow models. Our findings point to heterogeneity in TFP across regimes, technological catch-up, and, marginally, human capital, as the main determinants of the observed increase in inequality and polarization.

A general policy implication of our analysis is to adopt any action favouring transitions across regimes. In particular, we do not find evidence of poverty trap determined by thresholds in the level of GDP per worker in 1960, raising doubts on the utility of foreign aid (Easterly, 2006); on the contrary, a qualified foreign aid pointing to guarantee an adequate level of health could be very effective supporting regime transitions (see, in the same vein, Sachs and Warner, 1997 and Easterly, 2001). Stimulus to the accumulation of human capital, advocated by a large literature, instead, seems to be effective only in specific cultural environments (Benhabib and Spiegel, 2005).

Our findings also suggest some directions for further research. The first consists in integrating the studies on the evolution of income distribution and technological catch-up (see, e.g., Phillips and Sul, 2009 and Battisti et al., 2013) with those on the identification of growth regimes by spatial econometric techniques, where proximity among countries is explicitly taken into account. In this respect, a promising line of research is to consider growth models with technological spillovers modelled as spatial externalities (see, e.g., Ertur and Koch, 2007). The second direction is to develop a more sophisticated framework of model selection based on "multimodel inference" proposed by Anderson (2007), which represents an alternative approach to Bayesian Model Averaging. The third direction is to analyse transitions across regimes and, in this respect, to understand the reasons why countries do not make such transitions. Jerzmanowski (2006), Bos et al. (2010) and Anderson et al. (2016) represent interesting recent contributions in this line of research. Finally, there remains the key question as to why culture (religion) appears to be so important for TFP growth, i.e. why international technological spillovers can be mainly driven by culture.

Acknowledgements

Previous versions of this paper circulated with the title: "On the Determinants of Distribution Dynamics". We thank Roberto Basile, Michele Battisti, Daniel Henderson, Paul Johnson, Andros Kourtellos, Julie Le Gallo, Stefano Magrini, Thanasis Stengos, Joseph Zeira, and three anonymous referees for many useful comments on earlier drafts of this paper. We also thank seminar participants at the Universities of Pisa, Florence and Palermo, and at SMYE 2009, ERSA 2010, ICEEE 2011, RCEA Workshop 2013, and SNDE 2016. Remaining errors are our own responsibility.

References

Abramovitz, M., "Catching up, forging ahead, and falling behind," *Journal of Economic History*, 46, 385–406, 1986.

Acemoglu, D., S. Johnson and J. A. Robinson, "Institutions as a fundamental cause of long-run growth," *Handbook of Economic Growth*, 1, 385–472, 2005.

Anderson, D., *Model based inference in the life sciences: a primer on evidence.* Springer Science & Business Media, 2007.

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

Anderson, G., M. G. Pittau and R. Zelli, "Assessing the convergence and mobility of nations without artificially specified class boundaries," *Journal of Economic Growth*, 21, 283–304, 2016.

Ashley, R. A. and C. F. Parmeter, "When is it justifiable to ignore explanatory variable endogeneity in a regression model?" *Economics Letters*, 137, 70–74, 2015.

Bader, V. and M. Maussen, "Religious schools and tolerance," *Tolerance and Cultural Diversity in Schools. Comparative Report*, 87–107, 2012.

Barro, R. J. and J. W. Lee, "A new data set of educational attainment in the world, 1950–2010," *Journal of Development Economics*, 104, 184–198, 2013.

Barro, R. J. and R. M. McCleary, "Religion and economic growth across countries," *American Sociological Review*, 68, 760–781, 2003.

Battisti, M., G. Di Vaio and J. Zeira, "Global divergence in growth regressions," *CEPR Discussion Paper No. DP9687*, 2013.

Bazzi, S. and M. A. Clemens, "Blunt instruments: avoiding common pitfalls in identifying the causes of economic growth," *American Economic Journal: Macroeconomics*, 5, 152–186, 2013.

Beaudry, P., F. Collard and D. 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, 741–753, 2005.

Becker, S. O. and L. Woessmann, "Was Weber wrong? A human capital theory of protestant economic history," *Quarterly Journal of Economics*, 124, 531–596, 2009.

Benhabib, J. and M. M. Spiegel, "Human capital and technology diffusion," *Handbook of Economic Growth*, 1, 935–966, 2005.

Bloom, D., D. Canning and J. Sevilla, "Geography and poverty traps," *Journal of Economic Growth*, 8, 355–378, 2003.

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

Bos, J., C. Economidou, M. Koetter and J. Kolari, "Do all countries grow alike?" *Journal of Development Economics*, 91, 113–127, 2010.

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

Brock, W., *Cycles Growth and Structural Change: Theory and Empirical Evidence*, chapter Complexity-based models in cycles and growth: any potential value?, 302–338. Routledge, 2001.

Burnham, K. P. and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer Science & Business Media, 2003.

Cheshire, P. and S. Magrini, "Analyzing growth distribution dynamics: isolating divergence factors," *Paper presented at* $45^t h$ *European Regional Science Association Congress*, Cambridge, 2005.

Claeskens, G. and N. L. Hjort, *Model Selection and Model Averaging.* Cambridge University Press, 2008.

Cohen, D., "Tests of the "convergence hypothesis": some further results," *Journal of Economic Growth*, 1, 351–361, 1996.

Cohen, D. and L. Leker, "Health and education: another look with the proper data," *CEPR Discussion Paper No. DP9940*, 2014.

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

Desdoigts, A., "Patterns of economic development and the formation of clubs," *Journal of Economic Growth*, 4, 305–330, 1999.

Durlauf, S. and P. Johnson, "Multiple regimes and cross-country growth behaviour," *Journal of Applied Econometrics*, 10, 365–384, 1995.

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

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

Durlauf, S. N., A. Kourtellos and C. M. Tan, "Are any growth theories robust?" *Economic Journal*, 118, 329–346, 2008.

Durlauf, S. N., A. Kourtellos and C. M. Tan, "Is God in the details? A reexamination of the role of religion in economic growth," *Journal of Applied Econometrics*, 27, 1059–1075, 2012.

Easterly, W., *The Elusive Quest for Growth: Economists' Adventures and Misadventures in the Tropics*. MIT press, 2001.

Easterly, W., "Reliving the 1950s: the big push, poverty traps, and takeoffs in economic development," *Journal of Economic Growth*, 11, 289–318, 2006.

Easterly, W. and R. Levine, "Africa's growth tragedy: policies and ethnic divisions," *Quarterly Journal of Economics*, 112, 1203–1250, 1997.

Eichengreen, B., D. Park and K. Shin, "When fast-growing economies slow down: International evidence and implications for china," *Asian Economic Papers*, 11, 42–87, 2012.

Ertur, C. and W. Koch, "Growth, technological interdependence and spatial externalities: Theory and evidence," *Journal of Applied Econometrics*, 22, 10331062, 2007.

Feyrer, J., "Convergence by parts," *BE Journal of Macroeconomics*, 8, 1935–1690, 2008.

Flachaire, E., C. García-Peñalosa and M. Konte, "Political versus economic institutions in the growth process," *Journal of Comparative Economics*, 42, 212–229, 2014.

Gallup, J. L., J. D. Sachs and A. D. Mellinger, "Geography and economic development," *International Regional Science Review*, 22, 179–232, 1999.

Galor, O., "Convergence? Inferences from theoretical models," *Economic Journal*, 106, 1056–1069, 1996.

Galor, O., "Multiple growth regimes - Insights from unified growth theory," *Journal of Macroeconomics*, 29, 470–475, 2007.

Galor, O. and D. N. Weil, "Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond," *American Economic Review*, 90, 806–828, 2000.

Goetz, S. J. and D. Hu, "Economic growth and human capital accumulation: Simultaneity and expanded convergence tests," *Economics Letters*, 51, 355–362, 1996.

Guiso, L., P. Sapienza and L. Zingales, "Does culture affect economic outcomes?" *Journal of Economic Perspectives*, 20, 23–48, 2006.

Guo, Z., H. Kang, T. T. Cai and D. S. Small, "Testing endogeneity with possibly invalid instruments and high dimensional covariates," *arXiv:1609.06713*, 2016.

Hall, R. E. and C. Jones, ""Why do some countries produce so much more output per worker than others?"," *Quarterly Journal of Economics*, 114, 83–116, 1999.

Hansen, B. E., "Sample splitting and threshold estimation," *Econometrica*, 68, 575–603, 2000.

Hastie, T. and R. Tibshirani, *Generalized Additive Models.* CRC Press, 1990.

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

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

Heston, A., R. Summers and B. Aten, *Penn World Table Version 7.1.* Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, 2012.

Jerzmanowski, M., "Empirics of hills, plateaus, mountains and plains: A Markov-switching approach to growth," *Journal of Development Economics*, 81, 357–385, 2006.

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

Johnson, P. and C. Papageorgious, "What remains of cross-country convergence?" *Journal of Economic Literature*, forthcoming, 2017.

Knack, S. and P. Keefer., "Does social capital have an economic payoff? A cross-country investigation," *Quarterly Journal of Economics*, 112, 1251–1288, 1997.

Kourtellos, A, *Chapter 13 Modeling Parameter Heterogeneity in Cross-Country Regression Models*, 367–387. 2011.

Kuznets, S. and J. T. Murphy, *Modern Economic Growth: Rate, Structure, and Spread*, volume 2. Yale University Press New Haven, 1966.

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

Lucas, R. E., "Some macroeconomics for the 21st century," *Journal of Economic Perspectives*, 14, 159–168, 2000.

Maasoumi, E., J. Racine and T. Stengos, "Growth and convergence: A profile of distribution dynamics and mobility," *Journal of Econometrics*, 136, 483–508, 2007.

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

Maoz, Z. and E. A. Henderson, "The world religion dataset, 1945–2010: Logic, estimates, and trends," *International Interactions*, 39, 265–291, 2013.

Masters, W. A. and M. S. McMillan, "Climate and scale in economic growth," *Journal of Economic Growth*, 6, 167–186, 2001.

Mushkin, S. J., "Health as an investment," *Journal of Political Economy*, 70, 129–157, 1962.

Nagelkerke, N., "A note on a general definition of the coefficient of determination," *Biometrika*, 78, 691–692, 1991.

Nelson, R. R. and E. S. Phelps, "Investment in humans, technological diffusion, and economic growth," *American Economic Review*, 56, 69–75, 1966.

Ng, S. and J. Pinkse, *Nonparametric Two-step Estimation of Unknown Regression Functions when the Regressors and the Regression Error are Not Independent.* Université de Montréal, Départment de Sciences Economiques, Montréal, 1995.

Owen, A., J. Videras and L. Davis, "Do all countries follow the same growth process?" *Journal of Economic Growth*, 14, 265–286, 2009.

Phillips, P. C. B. and D. Sul, "Economic transition and growth," *Journal of Applied Econometrics*, 24, 1153–1185, 2009.

Pritchett, L. and L. H. Summers, "Asiaphoria meets regression to the mean," Technical report, National Bureau of Economic Research, 2014.

Quah, D., "Convergence empirics across economies with (some) capital mobility," *Journal of Economic Growth*, 1, 95–124, 1996.

Quah, D., "Empirics for growth and distribution: Stratification, polarization, and convergence clubs," *Journal of Economic Growth*, 2, 27–59, 1997.

R Development Core Team., *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0.

Rodrik, D., "Institutions, integration, and geography: In search of the deep determinants of economic growth," *Modern Economic Growth: Analytical Country Studies*, 2002.

Rodrik, D., *In Search of Prosperity: Analytic Narratives on Economic Growth.* Princeton University Press, 2003.

Rodrik, D., A. Subramanian and F. Trebbi, "Institutions rule: the primacy of institutions over geography and integration in economic development," *Journal of Economic Growth*, 9, 131–165, 2004.

Roeder, P. G., "Ethnolinguistic fractionalization (ELF) indices, 1961 and 1985," 2001.

Sachs, J. D., "Institutions don't rule: direct effects of geography on per capita income," Technical report, National Bureau of Economic Research, 2003.

Sachs, J. D. and A. M. Warner, "Fundamental sources of long-run growth," *American Economic Review*, 87, 184–188, 1997.

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

Sirimaneetham, V. and J. Temple, "Macroeconomic stability and the distribution of growth rates," *The World Bank Economic Review*, 23, 443–479, 2009.

Spolaore, E. and R. Wacziarg, "How deep are the roots of economic development?" *Journal of Economic Literature*, 51, 325–69, 2013.

Takeuchi, K., "Distribution of informational statistics and a criterion of model fitting," *Suri-Kagaku (Mathematical Sciences)*, 153, 12–18, 1976.

Tan, C., "No one true path: Uncovering the interplay between geography, institutions, and fractionalization in economic development," *Journal of Applied Econometrics*, 25, 1100–1127, 2009.

Temple, J. and P. A. Johnson, "Social capability and economic growth," *Quarterly Journal of Economics*, 113, 965–990, 1998.

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

Weil, D., "Accounting for the effect of health on economic growth*," *Quarterly Journal of Economics*, 122, 1265–1306, 2007.

Weil, D., *Economic Growth.* Prentice Hall, 2012.

Weil, D. N., "Health and economic growth," In "Handbook of Economic Growth," volume 2, 623–682. Elsevier, 2014.

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

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

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

World Bank., "China 2030. building a modern, harmonious, and creative society," *International Bank for Reconstruction and Development / The World Bank and the Development Research Center of the State Council, P. R. China*, 2013.

World Bank., *World Development Indicators 2015.* World Bank Publications. The World Bank, 2015.

.1 Sources and Descriptive Statistics of Determinants

The dependent variables used in our analysis is:

- **g** is the annualized average growth rate of the real GDP chain per worker.

The deep growth determinants used in our analysis are:

- **GDPpw.60** is the (log) of the real GDP chain per worker in the first year of the sample (*rgdpwok* in Heston et al. (2012) PWT 7.1)

- **PRI.1960** is the percentage of population aged 15 or above with at least primary education and corresponds to the sum of *lp*, *ls*, and *lh* in Barro and Lee (2013) ("percentage of primary (lp), secondary (ls), and tertiary (lh) attained in population"), in 1960.

- **SEC.60** is the percentage of population aged 15 or above with at least secondary education and corresponds to the sum of *ls* and *lh* in Barro and Lee (2013) ("percentage of secondary (ls) and tertiary (lh) attained in population"), in 1960.

- **LE.60** is the life expectancy at birth (years) in 1960 from World Bank (2015).

- **DEM.60.65** corresponds to the average of the *POLITY2* scores of the Polity IV dataset over the period 1960-1965.[33] In particular, the score is computed by subtracting the *AUTOC* score from the *DEMOC* score; the resulting unified polity scale ranges from +10 (strongly democratic) to -10 (strongly autocratic).[34]

[33] Due to the high number of missing values in 1960, we consider the average level of the variable over the period 1960-1965.

[34] In particular, *DEMOC* is derived from codings of the competitiveness of political participation (*XRCOMP*), the openness of political participation (*XROPEN*), the competitiveness of executive recruitment (*PARCOMP*), and constraints on the chief executive (*XCONST*).

- **CON.60-65** corresponds to the average of the *XCONST* score of the Polity IV dataset over the period 1960-1965. In particular, this variable refers to the extent of institutionalized constraints on the decision making powers of chief executives, whether individuals or collectivities. A seven-category scale is used: Unlimited Authority, Intermediate Category, Slight to Moderate Limitation on Executive Authority, Intermediate Category, Substantial Limitations on Executive Authority, Intermediate Category, Executive Parity or Subordination.[35]

- **ELF.61** is the ethnolinguistic fractionalization index of Roeder (2001) in 1961. In particular, we use the index constructed using the Taylor and Hudson formula that uses none of the groupings reported in the sources when data on sub-groups are available and treats racial distinctions within ethnolinguistic groups as separate ethnic groups.

- **ABS.LAT** is the absolute value of a country's latitude.

- **ME** is the Malaria Ecology index of Sachs (2003) predictive of malaria risk built upon climatological and vector conditions on a country-by-country basis.

- **TROP.AR** is the percentage of land in the geographical tropics as in Gallup et al. (1999).

- **LND100CR** measures the proportion of a country's total land area within 100 km of the ocean or an ocean-navigable river as in Gallup et al. (1999).

- **AR.FR** is the average number of frost-days within the country's borders as in Masters and McMillan (2001).

- **5.FR.DAYS** is the proportion of the country's land that receives five or more frost-days per month as in Masters and McMillan (2001).

- **PROT.65** is the percentage of Protestant adherents in the country in 1965 sourced from Maoz and Henderson (2013).

- **CAT.65** is the percentage of Roman Catholic adherents in the country in 1965 sourced from Maoz and Henderson (2013).

- **ISLAM.65** is the percentage of Islamists in the country in 1965 taken from Maoz and Henderson (2013).

- **ANI.65** is the percentage of Animists in the country in 1965 taken from Maoz and Henderson (2013).

- **log.y0** is the (log) of initial level of the real GDP chain per worker (*rgdpwok* in PWT 7.1).

[35] Notice that the *XCONST* also enters in the definition of the *POLITY2* indicators.

The proximate growth determinants used in our analysis are:

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

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

  where *rgdpch* is the real GDP chain per capita and *pop* is the population in PWT 7.1.

- **log.i/y** is the (log) investment rate at constant price and corresponds to the variable *ki* in PWT 7.1 divided by 100.

- **log.h** is the (log) average years of schooling attained and corresponds to the *yr_sch* in Barro and Lee (2013).

Tables 5 and 6 contain the descriptive statistics and correlations among the determinants.

.2 Derivation of the Differences between Actual and Counterfactual Distributions

In this appendix we derive an expression to evaluate the difference between the actual and counterfactual distributions. Consider the differences between the actual and the counterfactual distributions at period T, by expressing the value of (log) actual output per worker in period T, y_{iT}, in terms of the estimated counterfactual output per worker, $\tilde{y}_{iT}^{k}$:

$$\begin{aligned} \log\left(y_{iT}\right) &= \log\left(\tilde{y}_{iT}^{k}\right) + \sum_{m=1}^{M}\Bigg[\alpha\left(m\right) + \sum_{j=1, j\neq k}^{K}\mu_j(X_{i,j}, m) + \mu_k(X_{i,k}, m) + \upsilon_i \\ &- \hat{\alpha}\left(m\right) - \sum_{j=1, j\neq k}^{K}\hat{\mu}_j(X_{i,j}, m) - \hat{\mu}_k(\bar{X}_k, m)\Bigg], \end{aligned} \tag{15}$$

and taking the expected value of (the log of) actual output per worker of country i in period T conditional upon the actual output per worker in period 0, that is:

$$\begin{aligned} E\left[\log\left(y_{iT}\right)|y_{i0}\right] &= E\left[\log\left(\tilde{y}_{iT}^{k}\right)|y_{i0}\right] + \sum_{m=1}^{M}E\left[\alpha\left(m\right) - \hat{\alpha}\left(m\right)|y_{i0}\right] + \\ &+ \sum_{m=1}^{M}\sum_{j=1, j\neq k}^{K}E\left[\mu_j(X_{i,j}, m) - \hat{\mu}_j(X_{i,j}, m)|y_{i0}\right] + \\ &+ \sum_{m=1}^{M}\left(E\left[\mu_k(X_{i,k}, m)|y_{i0}\right] - E\left[\hat{\mu}_k(\bar{X}_k, m)|y_{i0}\right]\right) + \\ &+ E\left[\upsilon_i|y_{i0}\right]. \end{aligned} \tag{16}$$

Table 5: Mean and standard deviation of growth rate, deep and proximate determinants.

	g	GDPpw.1960	PRI.60	SEC.60	LE.60	DEM.60.65	CON.60.65	ELF.61	ABS.LAT	ME	TROP.AR
Mean	0.02	11759.63	0.55	0.15	54.83	1.02	0.32	0.47	24.56	3.72	0.54
SD	0.03	11214.42	0.32	0.16	12.48	7.3	12.48	0.28	17.99	6.76	0.48
	LND100CR	AR.FR	5.FR.DAYS	PROT.65	CAT.65	ISLAM.65	ANI.65	log.y0	log.n	log.i/y	log.h
Mean	0.47	7.12	0.4	0.11	0.39	0.17	0.09	9.24	-2.64	-1.57	1.52
SD	0.38	9.04	0.45	0.22	0.4	0.32	0.17	1.25	0.17	0.43	0.74

Table 6: Correlations among growth rate, deep and proximate determinants.

	g	PRI.60	SEC.60	LE.60	DEM.60.65	CON.60.65	ELF.61	ABS.LAT	ME	TROP.AR	LND100CR
g	1	0.170	0.130	0.200	0.110	0.060	-0.210	0.190	-0.200	-0.230	0.170
PRI.60	0.170	1	0.710	0.910	0.710	0.290	-0.490	0.630	-0.570	-0.500	0.530
SEC.60	0.130	0.710	1	0.690	0.600	0.270	-0.310	0.520	-0.390	-0.520	0.340
LE.60	0.200	0.910	0.690	1	0.670	0.240	-0.580	0.670	-0.610	-0.610	0.550
DEM.60.65	0.110	0.710	0.600	0.670	1	0.240	-0.240	0.410	-0.390	-0.300	0.510
CON.60.65	0.060	0.290	0.270	0.240	0.240	1	-0.120	0.220	-0.170	-0.170	0.090
ELF.61	-0.210	-0.490	-0.310	-0.580	-0.240	-0.120	1	-0.600	0.560	0.560	-0.460
ABS.LAT	0.190	0.630	0.520	0.670	0.410	0.220	-0.600	1	-0.460	-0.900	0.260
ME	-0.200	-0.570	-0.390	-0.610	-0.390	-0.170	0.560	-0.460	1	0.510	-0.370
TROP.AR	-0.230	-0.500	-0.520	-0.610	-0.300	-0.170	0.560	-0.900	0.510	1	-0.170
LND100CR	0.170	0.530	0.340	0.550	0.510	0.090	-0.460	0.260	-0.370	-0.170	1
AR.FR	0.220	0.540	0.530	0.560	0.380	0.230	-0.480	0.870	-0.410	-0.820	0.100
5.FR.DAYS	0.220	0.500	0.510	0.600	0.310	0.160	-0.530	0.900	-0.480	-0.940	0.210
PROT.65	-0.020	0.420	0.290	0.400	0.400	0.160	-0.220	0.480	-0.150	-0.260	0.040
CAT.65	-0.120	0.260	-0.030	0.210	0.070	0.010	-0.160	-0.060	-0.310	0.130	0.220
ISLAM.65	0.010	-0.510	-0.320	-0.420	-0.340	-0.120	0.150	-0.080	0.240	-0.030	-0.210
ANI.65	-0.150	-0.530	-0.360	-0.570	-0.350	-0.110	0.450	-0.430	0.680	0.390	-0.420
GDPpw.1960	0.010	0.710	0.680	0.760	0.620	0.300	-0.370	0.650	-0.420	-0.550	0.330
log.y0	0.010	0.740	0.620	0.810	0.570	0.270	-0.480	0.640	-0.510	-0.590	0.490
log.n	-0.350	-0.480	-0.330	-0.470	-0.340	-0.180	0.410	-0.570	0.260	0.490	-0.250
log.i/y	0.310	0.340	0.230	0.390	0.090	0.030	-0.190	0.240	-0.250	-0.290	0.140
log.h	0.050	0.760	0.620	0.760	0.540	0.250	-0.380	0.470	-0.550	-0.440	0.430
	AR.FR	5.FR.DAYS	PROT.65	CAT.65	ISLAM.65	ANI.65	GDPpw.1960	log.y0	log.n	log.i/y	log.h
g	0.220	0.220	-0.020	-0.120	0.010	-0.150	0.010	0.010	-0.350	0.310	0.050
PRI.60	0.540	0.500	0.420	0.260	-0.510	-0.530	0.710	0.740	-0.480	0.340	0.760
SEC.60	0.530	0.510	0.290	-0.030	-0.320	-0.360	0.680	0.620	-0.330	0.230	0.620
LE.60	0.560	0.600	0.400	0.210	-0.420	-0.570	0.760	0.810	-0.470	0.390	0.760
DEM.60.65	0.380	0.310	0.400	0.070	-0.340	-0.350	0.620	0.570	-0.340	0.090	0.540
CON.60.65	0.230	0.160	0.160	0.010	-0.120	-0.110	0.300	0.270	-0.180	0.030	0.250
ELF.61	-0.480	-0.530	-0.220	-0.160	0.150	0.450	-0.370	-0.480	0.410	-0.190	-0.380
ABS.LAT	0.870	0.900	0.480	-0.060	-0.080	-0.430	0.650	0.640	-0.570	0.240	0.470
ME	-0.410	-0.480	-0.150	-0.310	0.240	0.680	-0.420	-0.510	0.260	-0.250	-0.550
TROP.AR	-0.820	-0.940	-0.260	0.130	-0.030	0.390	-0.550	-0.590	0.490	-0.290	-0.440
LND100CR	0.100	0.210	0.040	0.220	-0.210	-0.420	0.330	0.490	-0.250	0.140	0.430
AR.FR	1	0.880	0.500	-0.150	-0.130	-0.270	0.580	0.530	-0.500	0.250	0.460
5.FR.DAYS	0.880	1	0.320	-0.130	-0.010	-0.350	0.580	0.600	-0.480	0.290	0.450
PROT.65	0.500	0.320	1	-0.260	-0.240	-0.110	0.490	0.330	-0.330	0.070	0.300
CAT.65	-0.150	-0.130	-0.260	1	-0.480	-0.310	0.160	0.240	0.040	-0.030	0.190
ISLAM.65	-0.130	-0.010	-0.240	-0.480	1	-0.010	-0.280	-0.230	0.220	-0.070	-0.400
ANI.65	-0.270	-0.350	-0.110	-0.310	-0.010	1	-0.430	-0.550	0.250	-0.290	-0.420
GDPpw.1960	0.580	0.580	0.490	0.160	-0.280	-0.430	1	0.820	-0.330	0.240	0.590
log.y0	0.530	0.600	0.330	0.240	-0.230	-0.550	0.820	1	-0.340	0.380	0.730
log.n	-0.500	-0.480	-0.330	0.040	0.220	0.250	-0.330	-0.340	1	-0.130	-0.240
log.i/y	0.250	0.290	0.070	-0.030	-0.070	-0.290	0.240	0.380	-0.130	1	0.390
log.h	0.460	0.450	0.300	0.190	-0.400	-0.420	0.590	0.730	-0.240	0.390	1

If $\hat{\alpha}$ and $\hat{\mu}_j$ ($j = 1, ..., K$), are conditional unbiased estimators of α and μ, and $E[\upsilon_i|y_{i0}] = 0$, Eq. (16) reduces to:[36]

$$E[\log(y_{iT})|y_{i0}] - E\left[\log\left(\tilde{y}_{iT}^k\right)|y_{i0}\right] = \sum_{m=1}^{M}\left\{E[\mu_k(X_{i,k},m)|y_{i0}] - \mu_k(\bar{X}_k,m)\right\}. \tag{17}$$

From Eq. (17), we can derive a condition for the equality of the expected values of the actual and counterfactual (log of) output per worker at time T conditional on the initial level $\mathbf{y}_0$. In particular, these values are equal, i.e.:

$$E[\log(y_{iT})|y_{i0}] = E\left[\log\left(\tilde{y}_{iT}^k\right)|y_{i0}\right], \tag{18}$$

if:

$$\sum_{m=1}^{M} E[\mu_k(X_{i,k},m)|y_{i0}] = \sum_{m=1}^{M}\mu_k(\bar{X}_k,m). \tag{19}$$

.3 Identification of Growth Regimes

The search for the best model is based on the exploration of all possible partitions according to the chosen set of deep determinants. Given the number of countries ($N = 84$) and periods ($S = 5$), the search is limited to partitions where each growth regime includes at least 11 countries, no more than two deep determinants are jointly considered, and for each determinant we consider at most one threshold (including the case in which the same determinant is used in the two steps). The sequence in which the determinants are used to split the sample could be crucial. Therefore, for each pair of deep determinants $(\mathbf{Z}_{q,1}, \mathbf{Z}_{q,2})$ the procedure is repeated switching the order of $\mathbf{Z}_{q,1}$ and $\mathbf{Z}_{q,2}$ in the splitting. In implementing this method the thresholds are defined by the values assumed by the two deep determinants in the sample. The results in terms of the minimum AICc for all pairs of candidate deep determinants is reported in Table 7, together with Bayesian posterior probability of the best model for each pair of deep determinants (see Eq. (4)).

The minimum level of AICc is found for the pair: life expectancy in 1960 and the Catholic share of the country's population .

Table 9 contains the country list for the three growth regimes corresponding to the best model.

.4 Distribution of proximate determinants within regimes

.5 Conditional marginal growth effects

.6 The Control Function Method and Endogeneity Test

The Control Function Method (CFM) treats endogeneity as an omitted variable problem, where inclusion of estimated first-stage residuals as a covariate corrects the inconsistency of the regression

[36]Note that $E\left[\hat{\mu}_k(\bar{X}_k,m)|y_{i0}\right] = E\left[\hat{\mu}_k(\bar{X}_k,m)\right] = \mu_k(\bar{X}_k,m)$ given that $\bar{X}_k$ is constant with respect to i and $\hat{\mu}_k$ is an unbiased estimator of μ_k.

Table 7: AICc of the best model for each possible pair of deep determinants referring to only the best models for each pair of deep determinants.

$1^{st}/2^{nd}$	None	GDPpw.60	PRI.60	SEC.60	LE.60	DEM.60.65	CON.60.65	ELF.61	ABS.LAT	ME	TROP.AR	LND100CR	AR.FR	5.FR.DAYS	PROT.65	CAT.65	ISLAM.65	ANI.65
GDPpw.60	-2104.28	-2109.40	-2187.54	-2172.28	-2210.70	-2137.10	-2113.81	-2147.36	-2179.52	-2157.81	-2120.22	-2134.45	-2180.14	-2170.48	-2160.17	-2158.22	-2129.14	-2119.46
PRI.60	-2172.00	-2182.22	-2183.28	-2183.78	-2204.35	-2182.48	-2182.74	-2172.48	-2180.25	-2177.56	-2180.87	-2181.46	-2188.38	-2188.52	-2188.21	-2208.98	-2185.44	-2180.17
SEC.60	-2145.59	-2190.45	-2183.04	-2166.20	-2203.97	-2161.63	-2159.50	-2165.16	-2178.53	-2153.58	-2159.81	-2158.41	-2166.88	-2166.90	-2173.22	-2178.41	-2156.85	-2155.55
LE.60	-2189.19	-2201.71	-2192.33	-2197.46	-2210.83	-2192.17	-2193.01	-2198.05	-2200.34	-2200.36	-2200.34	-2195.76	-2205.15	-2205.48	-2201.64	-2230.50	-2198.04	-2196.54
DEM.60.65	-2140.51	-2165.58	-2173.29	-2170.23	-2193.75	-2153.29	-2149.78	-2160.87	-2191.90	-2155.48	-2160.04	-2163.46	-2173.57	-2174.71	-2177.15	-2167.32	-2149.17	-2151.49
CON.60.65	-2109.06	-2117.40	-2182.74	-2159.50	-2200.95	-2149.78	-2119.12	-2153.75	-2175.53	-2146.63	-2144.02	-2125.23	-2168.22	-2160.98	-2151.82	-2147.36	-2119.13	-2120.49
ELF.61 45	-2133.86	-2151.65	-2170.31	-2162.97	-2196.17	-2161.43	-2148.03	-2130.25	-2156.23	-2146.01	-2136.06	-2148.91	-2157.13	-2142.10	-2184.01	-2152.28	-2144.15	-2146.05
ABS.LAT	-2150.10	-2186.77	-2180.61	-2170.46	-2212.17	-2183.60	-2164.32	-2170.37	-2161.62	-2159.61	-2163.55	-2163.04	-2172.48	-2173.07	-2198.74	-2184.38	-2160.68	-2146.09
ME	-2124.77	-2186.08	-2177.56	-2153.58	-2200.36	-2150.11	-2137.46	-2156.24	-2159.61	-2142.05	-2149.61	-2132.83	-2149.47	-2149.47	-2175.13	-2162.63	-2137.53	-2129.80
TROP.AR	-2119.33	-2186.77	-2183.79	-2169.59	-2212.17	-2183.60	-2150.67	-2171.59	-2152.19	-2152.76	-2119.33	-2140.26	-2134.10	-2129.18	-2165.25	-2157.53	-2141.54	-2125.19
LND100CR	-2099.70	-2131.22	-2168.56	-2149.06	-2184.26	-2151.14	-2137.66	-2148.91	-2160.79	-2132.83	-2140.64	-2134.41	-2158.96	-2136.04	-2157.58	-2147.15	-2119.51	-2116.60
AR.FR	-2125.11	-2190.42	-2196.09	-2172.44	-2219.16	-2190.76	-2156.95	-2175.45	-2159.20	-2166.33	-2128.95	-2148.01	-2146.36	-2139.02	-2171.52	-2159.73	-2143.87	-2141.33
5.FR.DAYS	-2125.11	-2190.42	-2190.39	-2175.33	-2216.98	-2190.76	-2156.95	-2179.35	-2158.49	-2166.88	-2129.18	-2148.01	-2140.18	-2130.77	-2171.52	-2162.92	-2158.24	-2133.74
PROT.65	-2161.39	-2175.44	-2186.78	-2173.79	-2193.22	-2173.97	-2163.78	-2173.74	-2175.26	-2163.74	-2173.76	-2174.14	-2178.93	-2178.94	-2182.20	-2197.74	-2175.15	-2168.22
CAT.65	-2120.41	-2136.94	-2218.46	-2187.61	-2212.88	-2179.11	-2140.22	-2171.70	-2189.82	-2174.68	-2154.27	-2132.33	-2162.48	-2170.95	-2185.29	-2139.66	-2166.07	-2128.66
ISLAM.65	-2101.56	-2155.84	-2185.44	-2156.55	-2198.04	-2147.90	-2131.46	-2144.13	-2188.00	-2152.03	-2144.24	-2117.48	-2167.13	-2178.62	-2175.15	-2161.23	-2117.23	-2111.26
ANI.65	-2097.91	-2111.74	-2184.55	-2174.81	-2210.99	-2151.49	-2114.23	-2149.27	-2146.09	-2146.32	-2120.07	-2111.33	-2145.24	-2125.82	-2168.22	-2136.43	-2117.90	-2100.51

Table 8: Bayesian model posterior probability of the best model for each possible pair of deep determinants referring to only the best models for each pair of deep determinants.

$\mathbf{1}^{st}/\mathbf{2}^{nd}$	None	GDPpw.60	PRI.60	SEC.60	LE.60	DEM.60.65	CON.60.65	ELF.61	ABS.LAT	ME	TROP.AR	LND100CR	AR.FROST	5.FR.DAYS	PROT.65	CAT.65	ISLAM.65	ANI.65
GDPpw.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PRI.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SEC.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LE.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	**0.992**	0.000	0.000
DEM.60.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CON.60.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ELF.61	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ABS.LAT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ME	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
TROP.AR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
LND100CR	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AR.FR	0.000	0.000	0.000	0.000	**0.003**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5.FR.DAYS	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
PROT.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CAT.65	0.000	0.000	**0.002**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ISLAM.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ANI.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Regime H	Regime LH	Regime LL
Australia	Algeria	China
Austria	Argentina	Egypt
Belgium	Benin	Gambia, The
Canada	Bolivia	Greece
Cyprus	Brazil	India
Denmark	Burundi	Indonesia
Finland	Cameroon	Iran
France	Central African Republic	Japan
Ireland	Chile	Jordan
Israel	Colombia	Korea, Republic of
Italy	Congo, Democratic Republic	Malaysia
Luxembourg	Congo, Republic of	Mali
Netherlands	Costa Rica	Mauritania
New Zealand	Cote d'Ivoire	Nepal
Norway	Dominican Republic	Niger
Spain	Ecuador	Thailand
Sweden	El Salvador	Turkey
Switzerland	Gabon	
United Kingdom	Ghana	
United States of America	Guatemala	
	Haiti	
	Honduras	
	Jamaica	
	Kenya	
	Malawi	
	Mexico	
	Morocco	
	Nicaragua	
	Panama	
	Paraguay	
	Peru	
	Philippines	
	Portugal	
	Romania	
	Rwanda	
	Senegal	
	Singapore	
	South Africa	
	Sri Lanka	
	Syria	
	Tanzania	
	Togo	
	Trinidad & Tobago	
	Uganda	
	Uruguay	
	Venezuela	
	Zambia	

Table 9: Lists of countries in the three growth regimes for the best model

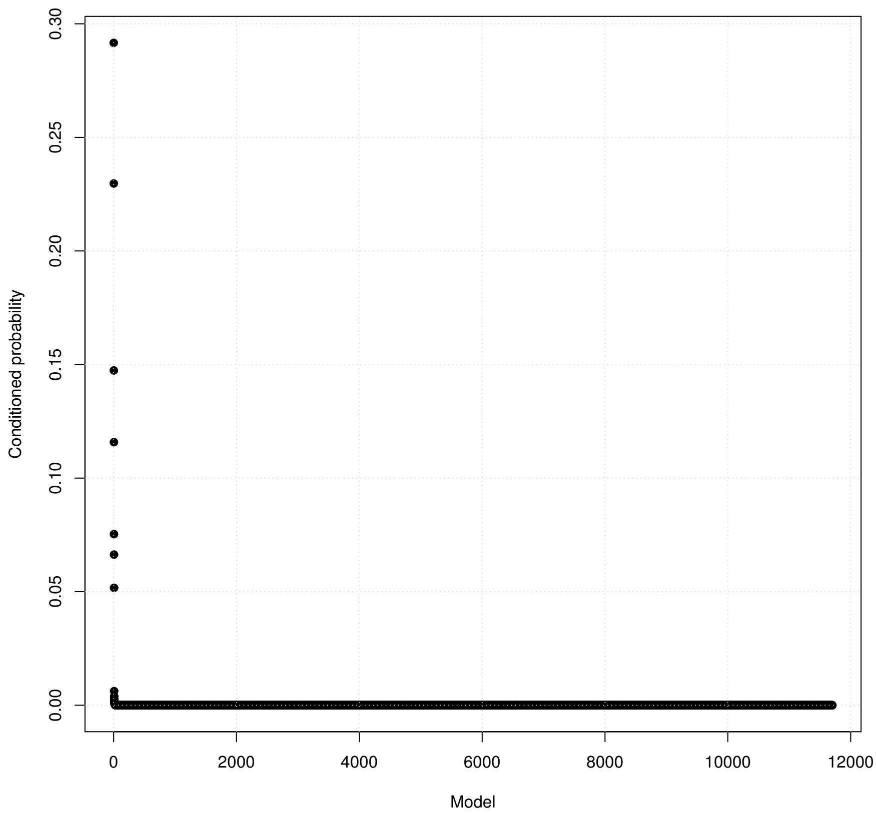
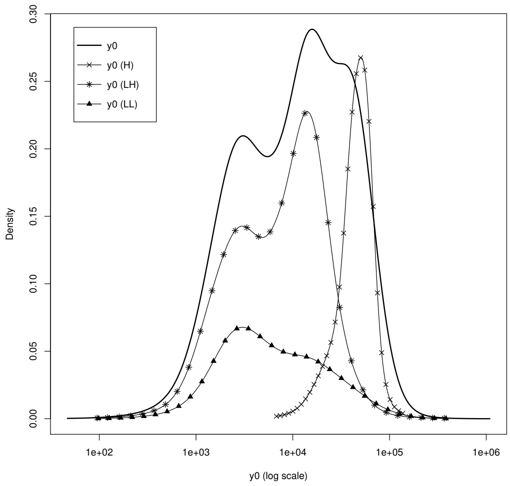
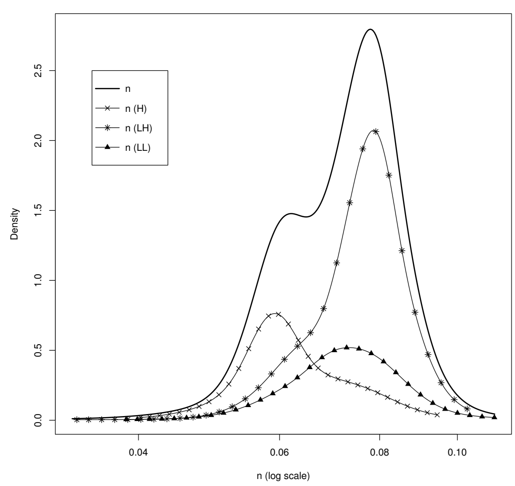


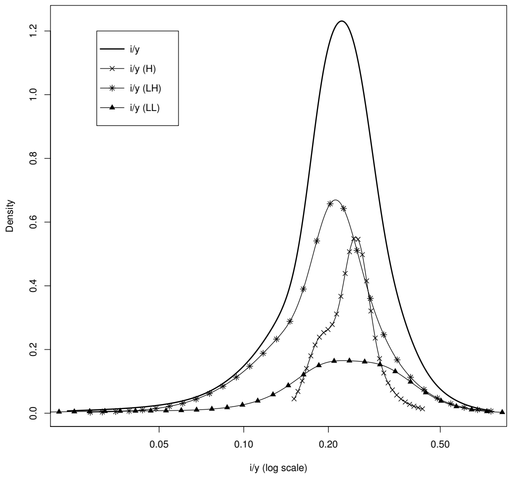
Figure .9: *Conditioned* probability of being the least false model for all possible countries' partition, given life expectancy at birth in 1960 and the % of Catholics in 1965 as partitioning variables.



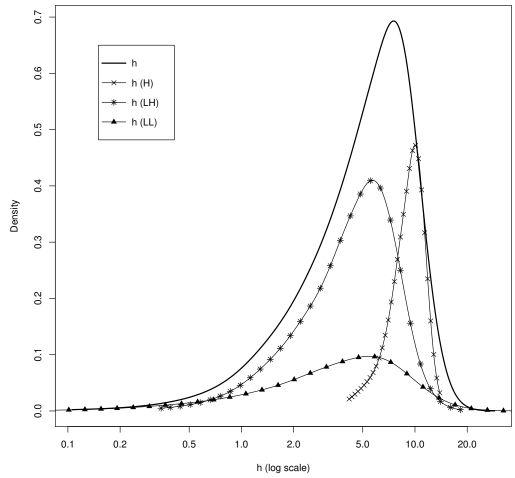
(a) Initial GDP per worker.



(b) (Augmented) Emploment growth rate.

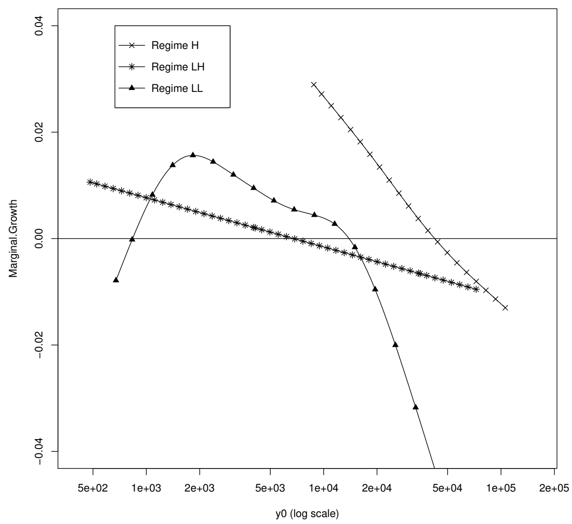


(c) Investment rate.

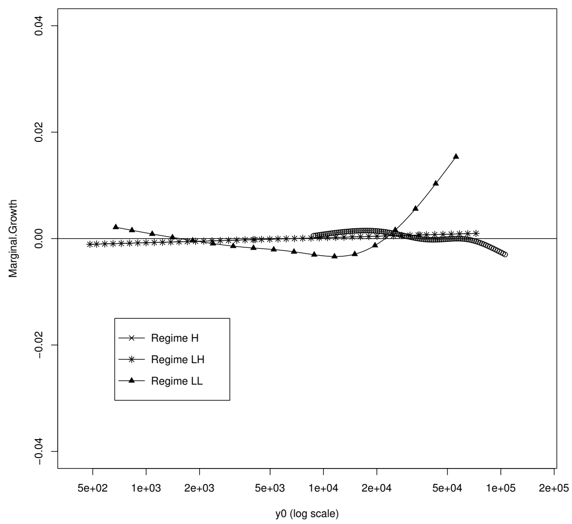


(d) Human capital.

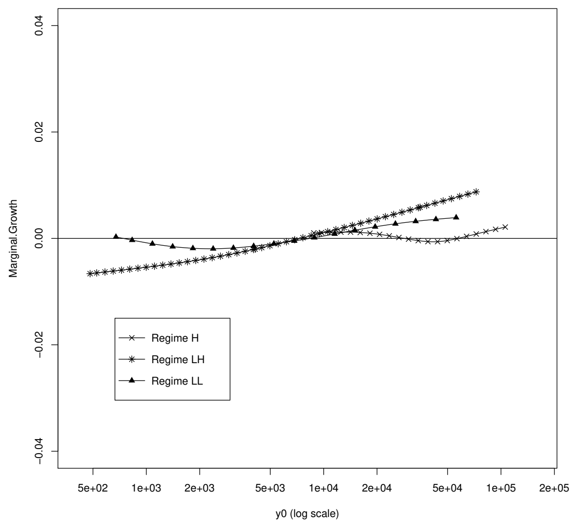
Figure .10: Distribution of proximate determinants for the whole sample and within each growth regime. Dotted vertical lines indicate the average value for the whole sample (black) and in each growth regime (blue, orange and red for Regimes H, LH, and LL respectively). Densities and averages are estimate from pooling observations of all periods for each growth regime.



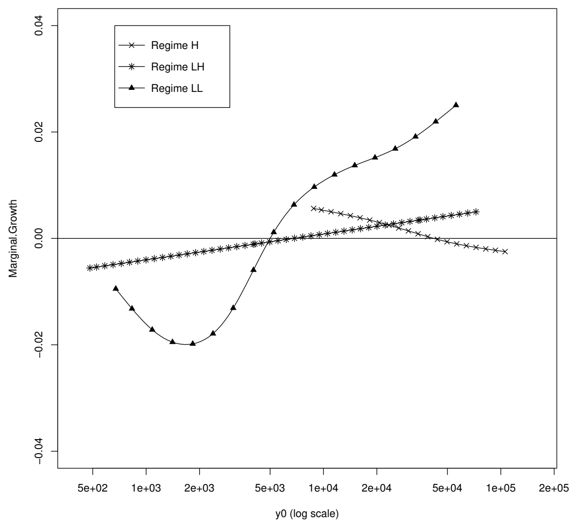
(a) Variable: log.y0.



(b) Variable: log.n.



(c) Variable: log.i/y.



(d) Variable: log.h.

Figure .11: Conditional marginal growth effect in the growth regimes.

of the dependent variable on the endogenous explanatory variable. This method provides consistent estimation of the underlying regression coefficients (see, e.g. Ng and Pinkse, 1995; Blundell and Powell, 2003). CFM is also used to perform the endogeneity test on the determinants of the growth model.

Following the CFM we use a two-stage procedure: i) first we run a semiparametric regression of each endogenous variable on the exogenous determinants and the instruments; then ii) we insert the first-stage residuals in the original semiparametric regression. To test the null hypothesis that the coefficients of the first-stage residuals are jointly equal to zero, we use a Likelihood Ratio test.

As instruments we use the initial level of each variable (i.e., in 1960) for all the sub-periods of the pooled dataset. In particular we use:

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

CFM is used in all estimations, that is for all the models estimated in the procedure to explore all potential growth regimes. Results of the first-stage regressions are reported in Table 10.

	Regime H			**Regime LH**			**Regime LL**		
Dep. Var:	**log.n**	**log.i/y**	**log.h**	**log.n**	**log.i/y**	**log.h**	**log.n**	**log.i/y**	**log.h**
	Pooled GAM	Pooled GAM	Pooled GAM	Pooled GAM	Pooled GAM	Pooled GAM	Pooled GAM	Pooled GAM	Pooled GAM
	1960-2008	**1960-2008**	**1960-2008**	**1960-2008**	**1960-2008**	**1960-2008**	**1960-2008**	**1960-2008**	**1960-2008**
Parametric	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate	Estimate
Intercept	-2.671***	-1.451***	2.019***	-2.625***	-1.692***	0.798***	-2.699***	-1.877***	0.440***
$D_{1970-1980}$	-0.061	0.0432	0.104***	0.051**	0.193***	0.350***	0.056	0.363***	0.399***
$D_{1980-1990}$	-0.101**	-0.036	0.184***	0.064***	0.068	0.663***	0.131***	0.429***	0.796***
$D_{1990-2000}$	-0.182***	-0.006	0.239***	0.067***	-0.035	0.877***	0.109***	0.416***	1.074***
$D_{2000-2008}$	-0.185**	0.067	0.278***	0.009	0.039	1.019***	0.055	0.488***	1.306***
Semi-parametric	EDF	EDF	EDF	EDF	EDF	EDF	EDF	EDF	EDF
log.y	1.000***	2.241**	1.000***	1.000*	1.000	3.192**	1.000*	1.000	4.414**
log.n.1960	2.858	3.019**	3.548***	3.725***	3.053**	1.14	2.104***	3.053**	2.7257
log.i/y.1960	7.833***	4.885***	1.000	1.000	4.020***	4.163***	2.115***	4.020***	1.000
log.h.1960	1.000	1.000	1.000***	3.930***	1.000	1.631***	3.348***	1.000	6.312***
REML score	-45.784	-40.97	-85.68	-158.53	65.823	-8.135	-37.85	30.175	24.739
Scale est.	0.011	0.014	0.006	0.011	0.014	0.006	0.014	0.076	0.050
Obs.	100	100	100	235	235	235	85	85	85
Countries.	20	20	20	47	47	47	17	17	17

Table 10: First-stage regressions of potentially endogenous determinants. Significance codes: 0.01"***" 0.05"**" 0.1"*". **EDF**: estimated degrees of freedom that 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). **REML score**: score of the restricted maximum likelihood estimation providing the fundamental information on the specification of the model. **Scale est.**: scale parameter, corresponding to the residual variance of the estimation. **Obs.**: number of observations. **Countries**: number of countries.

.7 A Test for Omitted-variable Bias in the Growth Model

We propose a test to detect the presence of possible omitted-variable bias in the growth model. In particular, from the estimation of Eq. (8) we obtain $\hat{g}_i^r$, the estimated residual growth of country i defined as $\hat{g}_i^r \equiv \log(y_{iT}/\hat{y}_{iT})$, where $\hat{y}_{iT}$ is the fitted value of the estimated growth model. Collecting the residual growth of all countries in the vector $\hat{\mathbf{g}}^r$ and the initial level of output per worker in the vector $\mathbf{y}_0$, a test of omitted-variable bias in the growth regression is expressed as follows:

$$E[\hat{\mathbf{g}}^r|\mathbf{y}_0] = E[\hat{\mathbf{g}}^r] = 0 \ , \forall \ \mathbf{y}_0. \tag{20}$$

If $\mathbf{y}_0$ is included in the set of proximate determinants, the condition in Eq. (20) ensures that no omitted-variable inconsistency related to $\mathbf{y}_0$ is present in the estimation (see Wooldridge, 2002, pp. 61-63). The violation of Eq. (20) would result in biased and inconsistent estimation of the growth model leading, in turn, to the inconsistency of the estimation of the distributional effect of proximate determinants.

The condition in Eq. (20) can be tested using a global test. In particular, consider two competing nested models for the residual growth:

$$\begin{aligned} \text{Model 1} \quad &: \quad \hat{g}_i^r = \alpha + u_i; \ \text{ and} \\ \text{Model 2} \quad &: \quad \hat{g}_i^r = \alpha + m(y_{i0}) + u_i, \end{aligned}$$

from which it is possible to formulate two alternative hypotheses H_0 and H_1:

$$\begin{aligned} H_0 \quad &: \quad E[\hat{\mathbf{g}}^r|\mathbf{y}_0] = \hat{\alpha} = \overline{\hat{\mathbf{g}}}^r = 0; \ \text{ and} \\ H_1 \quad &: \quad E[\hat{\mathbf{g}}^r|\mathbf{y}_0] = \hat{\alpha} + \hat{m}(\mathbf{y}_0). \end{aligned}$$

As in linear models we compare the residual sum of squares from two competing *nested* models by an F-test, where the *nonparametric F statistic* is given by:

$$F = \frac{(RSS_1 - RSS_2)/(df_2 - df_1)}{RSS_2/df_2}. \tag{21}$$

Although no general expression for the distribution of the nonparametric F statistic is available, Hastie and Tibshirani (1990) suggest that at least some approximate guidance can be given by referring to an F distribution with $(df_1 - df_2)$ and df_2 degrees of freedom (see also Bowman and Azzalini, 1997, pp. 153-154). When the (global) null hypothesis $E[\hat{\mathbf{g}}^r|\mathbf{y}_0] = 0$ for each $\mathbf{y}_0$ is rejected according to the F-test, a local test based on a bootstrap procedure can identify which range of values of $\mathbf{y}_0$ is responsible for the rejection.

For the estimate of the best model, the F-test cannot reject the null hypothesis of mean-independence between the residuals and GDP per worker in 1960 at the usual significance level of 5% ($F = 0.001$ with a *p-value* of 0.98). Moreover, Figure .12 reports the estimated distribution of the residual growth (i.e. $\hat{\mathbf{g}}^r$) conditional on GDP per worker in 1960 (i.e. $\mathbf{y}_0$) for the best model with growth regimes. We also report the estimated conditional mean (i.e. $E[\hat{\mathbf{g}}^r|\mathbf{y}_0]$) (thick line) with the corresponding 95% confidence bands obtained by a bootstrap procedure, and a vertical line representing the estimated unconditional mean (i.e. $E[\hat{\mathbf{g}}^r]$), which is approximately zero as expected. Figure .12 shows that for any level of GDP per worker in 1960 the conditional mean never differs from the unconditional mean at the usual level of significance of 5%. Accordingly, we conclude that the estimate of the best model with growth regimes does not suffer from omitted-variable bias, i.e. the estimated model appears correctly specified, at least conditioning on GDP per worker in 1960.

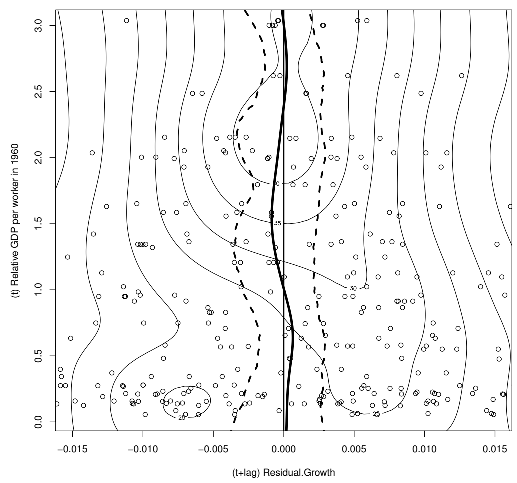


Figure .12: Conditional distribution of residual growth, the conditional mean (thick line), its confidence bands at 95% confidence level (dotted lines) and the unconditional mean (thin vertical line).

.8 The Estimation of Actual and Ergodic Distributions

In this appendix we describe the estimation of output per worker distributions (actual, counterfactual, and ergodic), and of the stochastic kernel, the operator that maps current distributions into future distributions.

Estimation of the Distribution of Output per Worker

In the estimation of the actual distribution, we use an adaptive kernel, a procedure recommended when observations vary in sparseness over the support of the distribution. 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:

$$\tilde{f}(\boldsymbol{x}) = \frac{1}{N \det(\boldsymbol{H})} \sum_{i=1}^{N} \omega_i k \left\{ \boldsymbol{H}^{-1} (\boldsymbol{x} - \boldsymbol{X}_i) \right\}, \tag{22}$$

where $k(\boldsymbol{u}) = (2\pi)^{-1} exp\left(-\frac{1}{2}\boldsymbol{u}^2\right)$ is a Gaussian kernel and the *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. $h_i = [4/(d+2)]^{1/(d+4)} \hat{\sigma_i} N^{-1/(d+4)}$, where $\hat{\sigma_i}$ 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:

$$\lambda_i = \left[\tilde{f}(\boldsymbol{X}_i)/g \right]^{-\alpha}, \tag{23}$$

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:

$$\hat{f}(\boldsymbol{x}) = \frac{1}{N \det(\boldsymbol{H})} \sum_{i=1}^{N} \lambda_i^{-d} \omega_i k \left\{ \lambda_i^{-1} \boldsymbol{H}^{-1} (\boldsymbol{x} - \boldsymbol{X}_i) \right\}. \tag{24}$$

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

Estimation of the Ergodic Distribution

The ergodic distribution is the long-run distribution of the stochastic process regulating the transitions across the state space, given in our case by relative output per worker levels. Its estimation requires the estimation of the stochastic kernel. A stochastic kernel is an operator mapping the density of a variable at time θ into its density at time $\theta + \tau$, $\tau > 0$, and indicates for each level of the variable at time θ its probability distribution at time $\theta + \tau$ over the possible values of the variable. The relation between the densities and the stochastic kernel can be summarized as: $f_{\theta+\tau}(z) = \int_0^\infty \phi_\tau(z|x) f_\theta(x)\, dx$, where z and x are values of the variable, and $\phi_\tau(z|x)$ is the stochastic kernel. To estimate the stochastic kernel $\phi_\tau(z|x) \equiv \phi_\tau(z,x)/f_\theta(x)$ we estimate the joint density of z and x, $\phi_\tau(z,x)$, and the marginal density of x, $f_\theta(x)$. In the estimation of $\phi_\tau(z,x)$ we follow Johnson (2005), and use the *adaptive kernel estimator* discussed above.

The ergodic distribution solves:

$$f_\infty(x) = \int_0^\infty g_\tau(x|z) f_\infty(z)\, dz, \tag{25}$$

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

$$\int_0^\infty f_\infty(x)\, dx = 1. \tag{26}$$

Since in our estimates GDP per worker is normalized with respect to its average, the ergodic distribution must satisfy the additional constraint:

$$\int_0^{\infty} f_{\infty}(x)\, x dx = 1. \tag{27}$$

In all computations we set $\tau = 49$. The *counterfactual stochastic kernel*, which is used to estimate the counterfactual ergodic distribution, is estimated considering as final distribution the counterfactual output per worker at T. In presence of growth regimes, an ergodic distribution is computed for each growth regime and the overall ergodic distribution is the mixture of these regime-specific ergodic distributions.

.9 Estimation of a Semiparametric Growth Model

Estimation of Eq. (6) is obtained by penalized likelihood maximization (see Wood, 2011, for details). The model is fitted by minimizing:

$$||\mathbf{y} - \mathbf{X}\beta||^2 + \sum_{k=1}^{K} \lambda_k \int_0^1 \left[\mu_k''(x)\right]^2 dx, \tag{28}$$

where $\mathbf{y}$ is the vector of response variables (g_i in our case), $\mathbf{X}$ is the matrix of determinants, 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. (6), $\mu_j(.)$, as:

$$\mu_j(X_{i,j}) = \sum_{l=1}^{q} b_l\,(X_{i,j})\,\beta_l \tag{29}$$

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

Parameters $\beta_{\mathbf{l}}$ are chosen to minimize the function in Eq. (28) 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.

Semiparametric estimation is performed following the approach proposed by Wood (2006) based on penalized regression splines. In particular, we used the *mgcv* package in R Development Core Team (2012), with the restricted maximum likelihood (REML) option (see Wood, 2011).