Catching-up, leapfrogging, and falling-back in economic growth — A nonparametric approach*

Harry Haupt[1], Joachim Schnurbus[1], and Willi Semmler[2,3]

[1]University of Passau, Germany, [2]Bielefeld University, Germany, [3]New School, New York.

Abstract. Classical growth convergence regressions fail to account for various sources of heterogeneity and nonlinearity. Recent contributions advocating nonlinear dynamic factor models remedy those problems by allowing for club-specific convergence paths. Unfortunately and similar to statistical clustering methods, those results are sensitive to choices made in the clustering mechanism. In this paper we improve existing clubbing algorithms while providing an economic rationale for time-varying heterogeneity of number, size, and composition of convergence clubs. We propose a nonparametric strategy for tackling neglected heterogeneity and nonlinearity jointly while alleviating the problem of underspecification of growth convergence regressions. This strategy, which rests on club-specific transition paths derived from a nonlinear dynamic factor model, allows estimation of convergence effects on club and even country level. The proposed approach is illustrated using a current Penn World Table data set. We find empirical evidence for leapfrogging and falling-back of countries over time. Guise and degree of nonlinearities in convergence regressions also differ substantially over time. Furthermore, some countries on club-based convergence paths exhibit, in contrast to their fellow club members, insignificant convergence effects.

Key words. Growth dynamics, club convergence, mixed kernel regression.

JEL classification. O40, O47, C14, C23, C25.

*Address correspondence to: Joachim Schnurbus. E-mail: joachim.schnurbus@uni-passau.de
Postal address: Department of Statistics, University of Passau, 94030 Passau, Germany.

1. Introduction: Convergence, heterogeneity, and nonlinearity

> The empirical study of economic growth occupies a position that is notably uneasy. Understanding the wealth of nations is one of the oldest and most important research agendas in the entire discipline. At the same time, it is also one of the areas in which genuine progress seems hardest to achieve. (see Durlauf et al., 2005).

Econometric growth convergence analysis in the sense of Barro et al. (1991) and Barro & Sala-i Martin (1992) is based on the concept of β-convergence: Poor countries grow faster than rich ones such that all countries converge to a common steady state. In its most basic form a simple linear regression model for cross-sectional data

$$g_{i,T} = \alpha - \beta \ln y_{i,0} + u_{i,T}, \quad 1 \leq i \leq n, \tag{1}$$

is considered, where $g_{i,T} = T^{-1} (\ln y_{i,T} - \ln y_{i,0})$ is the average growth rate of per capita income per period and $u_{i,T}$ is an error term. Calculated over an a priori fixed time horizon from initial period 0 to final period T, the growth rate $g_{i,T}$ is explained by the initial log per capita income $\ln y_{i,0}$. Noteworthy the cross-section regression (1) is based on panel data. However, only two time periods are used due to the fact that the experimenter a priori selects both initial and final period, $t = 0$ and $t = T$, respectively. Needless to say that results are prone to heavily depend on this selection. The parameter of paramount interest β is defined as $\beta = 1 - \exp(-\rho T)$, with convergence parameter ρ measuring the speed of convergence (see Barro & Sala-i Martin, 2004). Absolute β-convergence is assumed if $\rho > 0$ and consequently $\beta > 0$. Then countries with a smaller per capita income grow faster than rich countries such that income differences decrease.

Two strands of criticism confront convergence analysis in the spirit of the presented approaches.

> First, the concept of β-convergence could be invalid due to neglected heterogeneity. That is the relationship between $g_{i,T}$ and $\ln y_{i,0}$ in (1) may depend on factors varying in the cross-section dimension (for fixed T) (e.g., Masanjala & Papageorgiou, 2004; Alfò et al., 2008; Canarella & Pollard, 2004; Hauk Jr. & Romain, 2009; Moral-Benito, 2014; Mirestean & Tsangarides, 2015). Under the assumption of different steady states in log per capita income, Barro et al. (1991), Barro & Sala-i Martin (1992), and Mankiw et al. (1992) propose to apply the concept of conditional β-convergence, where additional covariates are included in Equation (1) to determine country-specific steady states (also compare Desdoigts, 1999; Sala-i Martin, 1996a).

Second, the lack of functional flexibility of the classical parametric formulation and estimation of (1) and hence the potential of neglected nonlinearities (e.g., Kalaitzidakis et al., 2001; Liu & Stengos, 1999; Maasoumi et al., 2007; Quah, 1993, 1997; Durlauf et al., 2008; Henderson, 2010; Cohen-Cole et al., 2012).

The second issue of potential nonlinearities in growth (and convergence) regressions has among others been addressed by applying flexible semiparametric methods (e.g., Liu & Stengos, 1999) or nonparametric methods (e.g., Fiaschi & Lavezzi, 2007; Graham & Temple, 2006; Henderson et al., 2013; Henderson & Russell, 2005; Maasoumi et al., 2007, and the literature cited therein). Applying the local linear kernel estimator with a generalized product kernel function proposed by Racine & Li (2004) and Li & Racine (2008) to the original data and models from Mankiw et al. (1992), Haupt & Petring (2011) find evidence for parametric misspecification. The essential uncertainty of what drives the wealth of nations also motivates the use of a nonparametric approach in this paper, as it also allows to address the problem that different countries (or groups of countries) may be represented by different growth models (see Cohen-Cole et al., 2012).

Addressing the first issue, Durlauf & Johnson (1995) and Canova (2004) separately estimate regression (1) for groups of countries with common convergence behavior within groups and heterogeneous behavior between groups. Identification of groups is either based on initial conditions or by Bayesian estimation of break points using the ordered cross-sectional units. Maasoumi et al. (2007) point out that heterogeneity of parameters in Equation (1) may also occur across the conditional distribution of growth rates. Motivated by a growing literature for modelling cross-section associations in panel contexts, Phillips & Sul (2003, 2007a,b, 2009), hereafter PS, argue that classical convergence analysis based on (1) is prone to deliver inconsistent results and invalid convergence tests due to potential heterogeneity in the convergence parameter β concerning time, countries, and individual technology levels. PS show that by neglecting the corresponding sources of heterogeneity, the error term in (1) includes endogenous variables and variables correlated with dependent and independent variables.[1]

[1] Kuersteiner & Prucha (2013) extend the dynamic factor framework by PS by allowing cross-sectional interactions not only in the systematic part, but also in the error component of linear panel data regression models. Omitting relevant covariates in Equation (1) for a set of countries with individual steady states results in biased estimated coefficients. Potential covariates such as investment rates into physical and human capital or infrastructure and R&D spending, might determine different levels of per capita income in the long run (e.g., Greiner & Semmler, 2005; Durlauf et al., 2005; Moral-Benito, 2014). Though research reveals many factors driving economic growth, we do not explicitly address this everlasting source of underspecification. We study club convergence regardless of what the specific factors are behind the convergence to some growth clubs.

The idea of PS is to allow transition parameter and growth rate to vary over countries and time. They propose a nonlinear dynamic factor model using panel data for $t = 0, \ldots, T$,

$$\ln(y_{i,t}) = a_{i,t} + x_{i,t}t = \left(\frac{a_{i,t} + x_{i,t}t}{\mu_t}\right)\mu_t = b_{i,t}\mu_t, \tag{2}$$

where $x_{i,t}$ is an individual technology process parameter, $b_{i,t}$ is an idiosyncratic time-varying element and μ_t is a common trend factor measuring global technological progress: $b_{i,t}$ can be interpreted as the transition path of economy i to the global growth path μ_t and is calculated as the log per capita income of country i in period t. By eliminating the global growth component, the relative transition path $h_{i,t} = \ln(y_{i,t})/n^{-1}\sum_{i=1}^{n}\ln(y_{i,t}) = b_{i,t}/n^{-1}\sum_{i=1}^{n}b_{i,t}$ measures the transition element for economy i in period t in relation to a cross-section average. In this vein global convergence is assumed to be present if

$$h_{i,t} \to 1, \quad \text{for all } i, \quad \text{as } t \to \infty, \tag{3}$$

that is, all countries have the same fraction of average global per capita income. The mean square transition differential $H_t = n^{-1}\sum_{i=1}^{n}(h_{i,t} - 1)^2$ is used to test convergence in the sense of condition (3) by means of the so called log t regression of Phillips & Sul (2007a,b, 2009),

$$\ln\left(H_0/H_t\right) - 2\ln(\ln(t)) = a + \gamma\ln(t) + u_t. \tag{4}$$

Whenever the null hypothesis $\gamma \geq 0$ of global convergence of the n countries is rejected, PS propose a clubbing algorithm based on successive log t regression based tests for detecting potential clubs and hence local convergence; cf. Subsection 2.1.[2] Note that if convergence in the sense of PS is present in the time interval from 0 to T, all countries have the steady state per capita income in T (the mean income $\overline{y}_{\bullet,T}$), independent of their initial income (in $t = 0$). This is equivalent to the linear regression $\ln(y_{i,T}) = \alpha + b\ln(y_{i,0}) + u_{i,T}$ with $b = 0$ under convergence. For a value of $b = 1 - \beta$ we obtain the classical β-convergence model (1).

Recent applications of the PS approach are Bartkowska & Riedl (2012), Panopoulou & Pantelidis (2009), while Dalgaard & Hansen (2005), De Siano & D'Uva (2006), D'Uva & De Siano (2007), and Fischer & Stirböck (2006) relate to the works of Durlauf & Johnson (1995) and Canova (2004). Alternative perspectives on growth and convergence clubs can be found in the various works of Danny Quah (compare Quah, 1996a,b,c), or in Alexiadis & Tomkins (2004), Ben-David (1998), Berthelemy & Varoudakis (1996), Chatterji (1992), Galor (1996), Howitt & Mayer-Foulkes (2005), Meliciani & Peracchi (2006), Sala-i Martin (1996b) and Startz (1998).

[2] Note that the standard error in the denominator of the t-statistic is usually based on an estimator of the long-run variance (compare Sul et al., 2005, and the literature cited therein).

We summarize the contributions of our paper by exemplarily discussing results for Germany and the United States, illustrated in Figure 1.

1. One key result is an economic explanation of why the algorithm of PS must be sensitive with respect to the choice of a variable z, used for initial sorting of countries. While PS suggest to use the final period per capita income $z_T = \ln(y_T)$, we perform a rolling calculation of the algorithm over all sorting periods by using z_s, $s = 0, \ldots, T$. The inherent sensitivity of the club structure w.r.t. the choice of s is explained in two ways:

First, in terms of a leapfrogging and falling-back interpretation, as countries can switch leading and lagging positions, within and between clubs: The trajectory of Germany shows falling-back to club 2 in 1986 and leapfrogging to club 1 in 2002, while the United States remain in the best club 1 over all sorting periods.

Second, indicated by the integers at the bottom of the plot, leapfrogging and falling-back causes a varying number of convergence clubs from period 0 to T. Driving forces discussed in the literature such as differences in technology adaption and innovation strategies over time are captured by sorting countries to clubs according to sorting variables $z_0, \ldots, z_T$, respectively.

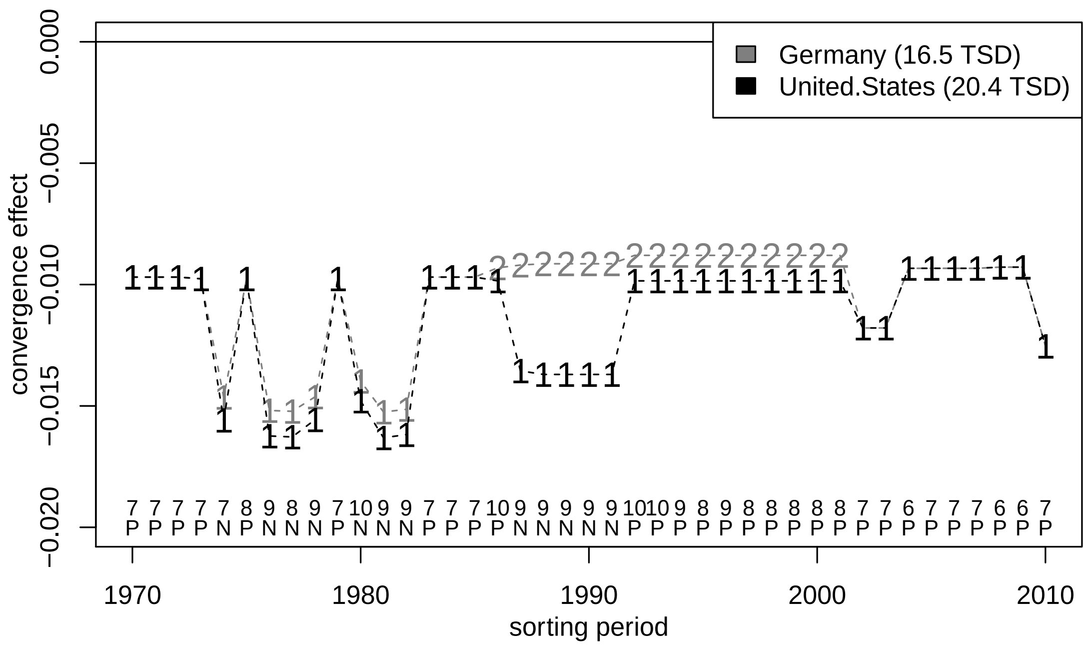

Figure 1: Estimated trajectory of country-specific convergence effects and club membership for United States and Germany. Integers at bottom correspond to total number of clubs in respective sorting period, "P" or "N" correspond to parametric or nonparametric specification, respectively. Numbers in legend correspond to initial period per capita income.

2. Furthermore we propose a framework which allows to estimate country- and club-specific convergence effects. We estimate a nonparametric version of the cross-section regression (1), where the club membership of each country is used as an additional regressor. Such a regression can be estimated for every sorting period s, $s = 0, \ldots, T$. The resulting time series of estimated patterns of convergence effects reveals considerable differences across countries and clubs, as exemplarily illustrated for the United States and Germany in Figure 1.

> First, due to different club-level convergence effects for sorting periods 1986 to 2001. That is club 1 and 2 have specific common trend factors μ_t in (2).
>
> Second, due to within-club heterogeneity in convergence effects 1974, 1976 to 1978, and 1980 to 1982. That is within club 1 the transition paths $b_{i,t}$ in (2) of Germany and the United States differ.

In detail we focus on changes of size and composition of clubs over time. Variants of a stochastic growth model establish a framework to model potential causes of catching-up, leapfrogging, falling-back, and change in club membership over time. We estimate convergence effects and to test for significance of the latter differences within a nonparametric framework, extending classical convergence regressions by allowing for club-specific or country-specific convergence effects. Outcomes of the test are indicated in Figure 1 as follows: "P" when the null of a correctly specified parametric model and hence homogeneous within-club effects cannot be rejected, "N" in the case of sufficient within-club variation and thus of country-specific convergence effects. The approach allows to tackle almost arbitrary forms of non-constant convergence effects (i.e., of nonlinear growth curves). Furthermore, for every sorting period s, the estimated bandwidths allow an overall judgement of the empirical relevance of the club structure in the sense of potential misclassification resulting from the clubbing algorithm.

The remainder of the paper is organized as follows. In Section 2, we address formation and time-dependence of the convergence club structure. Subsection 2.1 covers a review of the clubbing algorithm of PS and our extensions thereof, while Subsection 2.2 presents growth theoretical model variants to explain changing club membership and composition over time. Section 3 covers the simultaneous treatment of neglected heterogeneity and neglected nonlinearity using nonparametric growth convergence regressions. Subsection 3.1 provides technical details on mixed kernel regression, while Subsection 3.2 shows that the mixed kernel approach is able to simultaneously account for heterogeneity and nonlinearity in growth regressions. In Section 4 we apply the proposed approach to study a recent Penn World Table data release, while Section 5 concludes the paper.

2. Formation and time-dependence of convergence clubs

2.1. Convergence club identification

To identify convergence clubs, PS use a sorting (clubbing) algorithm consisting of four steps. Outlined and discussed below is an adapted version of the algorithm.[3] The main difference is that the algorithm is calculated repeatedly for all subsequent periods s, $s = 0, \dots, T$, of the panel.

<0> Set $s = 0$.

<1> Cross-section **sorting**:
Sort countries according to the Hodrick & Prescott (1997)-smoothed $z_s = \ln(y_s)$ for sorting period s.

<2> Form a **core group** of k^*, $2 \leq k^* < n$, countries:

<2.1> Find the first (i.e., highest income) two successive countries for which the log t test statistic $t_k \geq -1.65$ (i.e., the corresponding null hypothesis is not rejected). If $t_k < -1.65$ for all sequential pairs of countries, then exit the algorithm and "conclude that there are no convergence subgroups in the panel" (Phillips & Sul, 2007b).

<2.2> Start with the $k = 2$ countries identified in <2.1>, increase k (i.e., add new countries one by one) proceeding with the subsequent country from order, perform the log t test and stop increasing k if convergence hypothesis fails to hold (i.e., $t_k < -1.65$). The core group consists of the k^* countries $(k^* \in \{2, \dots, k-1\})$ that yield the highest value of the log t test statistic.

<3> Extend the core group to a **preliminary convergence club**:

<3.1> Form a complementary core group of all remaining countries (not already included in the core group or in preliminary convergence clubs).

<3.2> Add one country at a time from the complementary core group to the core group, run the log t regression, add the country to a club candidate group if the convergence test statistic is greater than a critical value $c^* = 0$. Form a preliminary convergence club of the candidate group and the core group.

<4> **Recursion** and stopping rule:
All countries that are not part of a preliminary convergence club form another group. Perform the log t test for this group. If convergence cannot be rejected, the countries of the group form the last preliminary convergence club. Otherwise, for the remaining countries start again with step <2>.

[3]The implementation in R (R Core Team, 2014) is available from the authors. It employs some minor tweaks (outlined in the code) in order to improve applicability and avoid computational issues.

<5> **Club merging** to determine final club structure:
Run log t regression for all groups of subsequent preliminary clubs. Merge those clubs fulfilling the convergence hypothesis jointly.[4]

<6> Repeat steps <1> to <5> for $s = 1, \ldots, T$.

Composing the clubs according to this algorithm does not ensure that the convergence hypothesis holds for each respective club. In step <3.2> a modified critical value of $c^* = 0$ is used, yielding an easier rejection of the convergence hypothesis (compared to the usual critical value of -1.65). However, as each country is added to the core group individually, passing an individual test (despite $c^* = 0$) does not guarantee that the countries of the preliminary convergence club pass the test with usual critical value of -1.65. Phillips & Sul (2007b) are aware of this problem and propose to increase the critical value c^* for raising the power of the corresponding test. However, such a remedy does not work automatically in general, as it requires manual tweaking of critical values. For instance when we replace the initial cross-section sorting rule by an equally plausible alternative. Thus, we augment step <3> of the algorithm in a way such that convergence is assured using a data-based criterion:

<3.3> If the convergence hypothesis holds for all countries in the preliminary convergence club, go to step <4>. If not, form a preliminary convergence club consisting of the countries of the core group and the complementary core group country, where the highest test statistic value is obtained in step <3.2>. Add the complementary core group members w.r.t. decreasing $t_{\hat{\gamma}}$ one by one to the preliminary convergence club until the corresponding $t_{\hat{\gamma}} \geq -1.65$ (i.e., convergence hypothesis is not rejected).

A crucial point is that the algorithm is sensitive w.r.t. sorting in step <1>. More precisely, if for instance the initial period income y_0 is used instead of the last period income y_T, different clubs can be obtained from using the algorithm. Differences may occur in terms of number, size, and composition of the clubs. Considering the problem of an unknown true sorting rule (compare Canova, 2004), PS usually advise to sort according to the last period income (compare Phillips & Sul, 2009).

The computational reason for such a dependence of clubbing results on the sorting time period in step <1> can be seen in the fact that the first two countries of the core group are very likely to have a large relative income in this time period, as step <2.1> starts with the respective highest income countries. From step <2.2> onward, countries are added to the two original members

[4]Note that each preliminary club is only merged once which is in accordance to the club merging proposal of Phillips & Sul (2009, compare Table II.). If the number of final clubs seems implausibly high, iteration of the last step may reduce the number of final clubs, compare Panopoulou & Pantelidis (2009).

of the core group whenever they have a common convergence behavior. Hence, all additional club members have to match the first two countries in terms of their relative transition paths (i.e., convergence behavior). In summary, sorting according to the income of different periods in step <1> is prone to create different pairs of countries in step <2.1>, and thus likely to yield different countries added in the remaining steps of the algorithm.

What basically seems to be an issue of non-robustness of the algorithm can be explained economically. In fact, it can be seen as a peculiar strength, as the algorithm is intended to identify club membership of each country based on the income order at a certain time period. Consequently, the use of different sorting periods s, $s = 0, \ldots, T$, allows to analyze potential changes in club membership over time for each country. In the following Subsection 2.2 we provide an economic rationale for such a behavior.

2.2. Clubs, leapfrogging and falling-back

Though there could be club convergence for certain time periods, countries may not stay in the same club persistently. There may be leapfrogging of countries, switching leading and lagging positions, not only within but also between clubs, leading to a change of number and composition of clubs over a longer time horizon. The first part of this Subsection briefly reviews the growth literature in terms of leapfrogging. The second part covers a growth theoretical model, while part three covers examples of club convergence and leapfrogging within this framework.

Growth literature related to leapfrogging

In the work of the founders of modern growth theory, such as in Solow (1956), or the later optimizing version of Brock & Mirman (1972), neither club convergence nor changes in club membership have been studied. Later econometric work had begun to explore different modes of convergence, as discussed in Section 1. Yet, as we argue, the complex issues of growth convergence regressions allowing for various sources of heterogeneity and nonlinearity, showing up in classical convergence tests, have not been addressed sufficiently. What forces of economic growth may help us to explain such outcomes?[5]

Earlier work on leapfrogging and switching of leading and lagging positions can be found in Brezis et al. (1993). Their model allows for switching of leading and lagging positions through new technologies and wage differentials. They argue that a leading country – that is likely to

[5]For a more detailed analysis of the forces of growth and how they may affect different levels of per capita income of countries in the long run, see Greiner & Semmler (2005).

also have higher real wages – may not stay dominant because it does not implement the new technology on time. They formulate a condition how newly emerging countries can leapfrog older incumbents: Let $A_2(0)$ denote the new technology that the new country starts to apply, whereas $A_1(K(T_2))$ denotes the incumbent country's old technology at the time period T_2 when the new technology becomes available. Then

$$\frac{A_2(0)}{A_1(K(T_2))} > \frac{w^*}{w}. \tag{5}$$

They assume that the incumbent country has, due to its technological and manufacturing leadership in the past, also increased wages but is locked into old knowledge and experiences to operate technologies. For example, they may have had technology superiority in manufacturing and agriculture, but the new competitor, with the new technology, still has relatively low wages w^* and thus cost advantages, leading to a rise of market shares in manufacturing. The model by Brezis et al. (1993) works to explain leapfrogging, if the new country does not only have lower wages but also implements the new technology faster than the incumbent. As the authors argue, this process of technological leadership through leapfrogging, relative to wage cost, may reverse at a later time period again. Hence, there might be cycles in leapfrogging. Compare also Barro & Sala-i Martin (1997) for a discussion of technological leadership and the role of intellectual property rights as well as the application in Phillips & Sul (2007a) for the relationship of speed of learning/technology implementation and growth.

Acemoglu et al. (2006), by focusing mostly on advanced countries, concentrate on innovations as the main factor of leapfrogging or falling-back. The authors stress that countries that move from purely investment-based strategies to innovation-based strategies – with innovations at the world technology frontier – may leapfrog ahead of their competitors with superior innovations. The role of relative wages, possibly determining cost advantages when a new technology is introduced, is however not considered there.

We take both previous approaches as representative for many others. More generally, factors such as investment rates into equipment, human capital, R&D spending, and infrastructure investment can account for different levels of per capita income in the long run, see Greiner & Semmler (2005). Other models view a rich resource endowment as force of economic growth and leapfrogging. A rich endowment of resources can at least lead to a temporary rise of output and capital stock, though growth may decline later due to resource scarcity.[6] So joining a club of higher rank may only be temporary, especially if these countries do not manage to

[6]See Dasgupta & Heal (1974) and Greiner & Semmler (2008). Gruene et al. (2013) show how growth due to natural resources is feasible but transient.

compete with the countries at the forefront of technological change. Furthermore, it is often stressed that a stable macroeconomic and financial development may enhance growth and lead to a steady rise of per capita income. By way of contrast, countries may fall back to a lower level club, because of macroeconomic instabilities and financial/banking crises.[7]

Growth theoretical model

It has become common to explain convergence mechanisms in the context of a stochastic optimal growth model of the Brock & Mirman (1972) type, which also has become central in Real Business Cycle (RBC) models. We start with a non-stochastic version. The framework is formulated as a discrete time maximization problem with a discounted instantaneous payoff function and dynamics given by

$$\max_{u_i \in U} \sum_{t=0}^{N-1} \beta^t g(u_t), \tag{6}$$
$$\text{s.t.} \quad x_{t+1} = A x_t^{\alpha} - u_t$$

with $g(u_t) = \ln u_t$ and constant parameters $A > 0$, $0 < \alpha < 1$ and $0 < \beta < 1$. The exact solution to this problem for $N \to \infty$ (see Santos & Vigo-Aguiar, 1995), is given by $V(x) = B + C \ln x$, with value function $V(x)$, $B = \ln((1-\alpha\beta)A) + [\alpha\beta/(1-\alpha\beta)] \ln(\alpha\beta A)/(1-\beta)$ and $C = \alpha/(1-\alpha\beta)$, where we omit the time index for the sake of simplicity. The unique optimal equilibrium for this example is given by $x^\star = (\alpha \cdot \beta \cdot A)^{1/(1-\alpha)}$.

The next issue is to allow for stochastic technology shocks and to track the respective growth paths. Moreover, there are likely to be not only stochastic shocks that move economies ahead of other economies, or back to an inferior position, but there also can be persistent leapfrogging due to new jumps in technology, as discussed in the first part of this subsection. We will explore those cases. Following again the Brock & Mirman (1972) model we can extend the above model, using a second state variable to denote the stochastic influences on the technology trend. This model variant is then given by Equation (6) with payoff function $g(u) = \ln u$, and two discrete time equations[8]

$$x_{t+1} = A \cdot x_t^{\alpha} \cdot z_t - u_t, \tag{7}$$
$$z_{t+1} = \exp\left(\rho \cdot \ln z_t + \varepsilon_{t+1}\right), \tag{8}$$

[7] Dwyer et al. (2013) empirically demonstrate that occasionally there may be large growth losses. They however also show that the size of growth losses is frequently compensated in later time periods.

[8] Using $E(\exp(a + \varepsilon_t)) = \exp(a + \sigma^2/2)$, the model used for the open loop optimization is given by $x^e_{t+1} = E\left(A \cdot (x^e_t)^{\alpha} \cdot z^e_t - u_t\right) = A \cdot (x^e_t)^{\alpha} \cdot z^e_t - u_t$ and $z^e_{t+1} = E\left(\exp\left(\rho \cdot \ln z^e_t + \varepsilon_{t+1}\right)\right) = \exp\left(\rho \cdot \ln z^e_t + \sigma^2/2\right)$.

where A, α, and ρ are real constants, and $\varepsilon_t \sim iid(0, \sigma^2)$. Following the computations in Santos & Vigo-Aguiar (1995), we get the optimally controlled dynamics

$$x_{t+1} = \alpha\beta A x_t^{\alpha} z_t, \tag{9}$$

and are in a position to derive equations for the steady state values of $E(\ln x)$ and $E(\ln z)$. Re-transformation to the original exponential variables yields the expected equilibria[9] $z^{e,*} = \exp(\sigma^2/(2 \cdot (1-\rho^2)))$ and $x^{e,*} = (\alpha \cdot \beta \cdot A)^{1/(1-\alpha)} \cdot (z^{e,*})^{1/(1-\alpha^2)}$.

Figure 2 shows six realizations of the stochastic growth path determined by Equations (8) and (9).[10] For each of the six realizations we use a different initial condition w.r.t. $x_0 \in \{0.8, 1.3, 1.6, 2, 2.5, 2.8\}$, while $z_0 = 1$ is applied. Figure 2 clearly shows β-convergence for six initial conditions. There is one growth club where growth paths converge toward the same steady state $x^{e,*} \approx 2.068729$, though intra-club leapfrogging occurs (i.e., stochastic growth paths cross). Intra-club leapfrogging yields switching of ranks within a club. The dashed lines show the paths of technology z_t which, in the non-stochastic case, was always equal to 1. One might also get the impression of σ-convergence about some eventual steady state, due to technology shocks.

— Insert Figure 2 around here —

Growth clubs and changing club membership

Next we differentiate technology levels as suggested by Brezis et al. (1993) and also Acemoglu et al. (2006), while maintaining the assumption that the relative wage cost will fulfill the Brezis et al. condition in all of the subsequent model variants. Then countries do not converge, but economic growth leads to different per capita income of countries. This is likely to be an essential reason why one observes convergence clubs. Subsequently, we will characterize the technology level or, more generally, the level of per capita income by parameter A.[11]

[9] For details see Gruene et al. (2013).

[10] As a numerical example we use the parameter values $\alpha = 0.34, \beta = 0.95, A = 5, \rho = 0.9$, and $\varepsilon_t \sim iid\ N(0, 0.015^2)$ and obtain $z^{e,*} \approx 1.000592$ and $x^{e,*} \approx 2.068729$ (indicated by a horizontal line). The variance of the technology shock is chosen roughly as proposed in the RBC literature. The numerical solution paths for a finite horizon (usually $N = 5$) model can be determined by the NMPC algorithm proposed by Gruene et al. (2013). NMPC computes the paths globally without requiring local linearization and is expected to compute solutions such that growth is steered into a neighborhood of the steady state. This is equivalent to the behavior in Figure 2, as it converges to the optimal equilibrium. Gruene et al. (2013) provide an accuracy test, computing the trajectories for paths if $N \to \infty$.

[11] As mentioned above, many factors as investment rates, human capital, R&D spending and infrastructure can account for different levels of per capita income, and thus the parameter A.

In Figure 3), the six realizations of Figure 2 are assigned to one of two growth clubs. The three countries of club 1 (black lines) have technology level $A = 6$ and thus vary around an expected equilibrium level of $x^{e,*} \approx 2.726938$, while for the countries of club 2 (grey lines) it remains that $A = 5$ and $x^{e,*} \approx 2.068729$. For both clubs we observe intra-club leapfrogging and, in earlier time periods, also a crossing of growth paths for members of different clubs because of different initial points x_0. Even though a country starts at a lower initial point, it can leapfrog the other countries and be a member of a higher growth club in the long-run. This can be motivated by the takeover criterion of Brezis et al. (1993), meaning that the technology of the competitor is eventually implemented at a higher level than the technology of the incumbent. In this case we would clearly observe convergence clubs instead of β-convergence. The eventual take-over of the leading role of higher technology may also be justified in terms of Acemoglu et al. (2006) who argue that countries may sometimes switch from an investment-based technology to an innovation-based technology.

— Insert Figure 3 around here —

For Figure 4 we increase the variance to a value of 0.045^2 keeping only one realization for each club (the remaining configuration is equivalent to Figure 3). Leapfrogging and its counterpart (falling-back) is indicated by L and F, respectively. We observe that the stochastic growth paths of both countries cross several times even though they belong to different clubs in the long-run. Though both countries exhibit a technology shock of the same variance – the country with the lower level technology, $A = 5$, can surpass the other. This temporary leapfrogging can for example result from a richer resource endowment of countries (availability of natural resources). In the long run, the technology level will remain to be decisive – about which the growth path will fluctuate. So, in the long term, without changing the technology level, the country that has moved ahead of the incumbent due to resource advantages, might fall back again in terms of per capita output.[12]

— Insert Figure 4 around here —

Note that if the countries of different growth clubs had different variances, we might observe that – though a country can have a high technology level, as the members in the upper two

[12] This is what most growth models with resources predict. See Dasgupta & Heal (1974) or Greiner & Semmler (2008). In Gruene et al. (2013) it is also numerically shown that growth is enhanced through the use of natural resources but it is likely to be transient.

graphs – yet, the higher variance of the technology shocks does not prevent it to become a member of a lower level technology club. This is likely to hold at least temporarily if a country experiences to become a laggard due to financial and banking crises, as studied in Dwyer et al. (2013). On the other hand, an incumbent, as represented by the lower level technology, can again sometimes surpass the newcomers, as Brezis et al. (1993) predicted.

The above examples are chosen for illustrative reasons. Of course, the causes for growth clubbing and membership changes of growth clubs are not exhaustively treated. Yet we could exemplify modes of convergence, changing ranks in an existing growth club due to stochastic technology shocks. There is also leapfrogging and new club formation due to persistent technology improvements, changing membership of clubs due to growth slowdowns or technology improvements, or temporary leapfrogging due to better resource endowments. So there is switching of leading and lagging positions, within and between clubs. This can depend on the technology levels, and a number of covariates we have discussed, but it can also depend on the variances of the technology shocks. Overall, the number, size and the composition of clubs may depend on the time period considered.

In real world applications, we do neither know, whether there are certain convergence clubs, nor do we know, whether a region/country truly belongs to a certain club. But we can use the algorithm of Subsection 2.1 as a tool to determine a likely club membership for each country based on the income order of all countries in a certain period. In terms of Figure 4 this means that we do neither know how the true growth paths of different countries proceed and thus do not know, when they (will) cross, nor do we know the true club membership (compare color in plots), but we can determine a likely club membership at certain points in time. Hence, we denote as leapfrogging (falling-back) when a country moves to a growth club with a lower (higher) number, as this corresponds to a higher (lower) average income (also see the discussion in Phillips & Sul, 2009, Section 5).

In terms of convergence regressions we propose to add a club covariate. Estimating the resulting regressions for each sorting periods allows to empirically assess potential leapfrogging or falling-back over the time horizon.

3. Nonlinearity and heterogeneity in growth convergence regressions

How can the previous considerations be used to extend classical convergence analysis in the sense of Mankiw et al. (1992)? More precisely, club-formation, leapfrogging and falling-back, require a regression framework allowing for time-varying within and between club heterogeneity of club- and country-specific convergence effects. In a first step, we assign each country to a club using the algorithm discussed in Subsection 2.1. The second step rests on estimation of nonparametric convergence regressions reflecting the economic rationale discussed in Subsection 2.2. The following Subsections show that the resulting modeling framework is capable of simultaneously addressing the issues of potential heterogeneities and nonlinearities raised in Section 1.

3.1. Nonparametric club-based growth regression

Starting from the concept of absolute (club-based) β-convergence we formulate a cross-section regression exploiting the panel structure of the data in the sense of PS. More precisely, for each sorting time period $s \in \{0, \ldots, T\}$ of the PS algorithm, we estimate a multiple nonparametric growth convergence regression

$$g_i = G_{(s)}(\mathbf{z}_{i,(s)}) + u_{i,(s)} \tag{10}$$

where $g_i = g_{i,T}$ is the growth rate defined in equation (1) and the covariate vector is given by $\mathbf{z}_{i,(s)} = (\ln(y_{i,0}), club_{i,(s)})$ where the categorical regressor $club_{i,(s)}$ assumes the value j if country i belongs to club j for sorting period s. In order to keep notation simple we omit the sorting index s in the following equations. The covariate vector can be extended by additional continuous or discrete (i.e., "mixed") covariates towards a conditional convergence concept, where country-specific effects are modeled via additional covariates.[13]

The mixed kernel minimization problem is a local weighted least squares problem

$$\min_{\widetilde{G}(\mathbf{z}_0), \widetilde{\beta}(\mathbf{z}_0)} \sum_{i=1}^{n} [g_i - \widetilde{G}(\mathbf{z}_0) - \widetilde{\beta}(\mathbf{z}_0)(\ln(y_{i,0}) - \ln(y_{0,0}))]^2 W(\mathbf{z}_0, \mathbf{z}_i, \boldsymbol{\lambda}), \tag{11}$$

where "local" means that the regression function is estimated at a certain position $\mathbf{z}_0$ in the covariate space[14]. The estimated local intercept $\widehat{G}(\mathbf{z}_0)$ is the fitted value. Whenever a local

[13]The usual disclaimer with respect to the curse of dimensionality applies. It lies in the eye of the beholder how the trade-off between slow convergence rate and potential omitted variable bias has to be solved. Recent work of Li and Racine suggests that the former problem often may not be crucial for empirical analysis, as frequently a low number of continuous covariates comes along with a higher number of categorical covariates, the latter not affecting the curse of dimensionality.

[14]For obtaining an estimated regression surface, the regression function has to be estimated at various positions $\mathbf{z}_0$. The usual way to conduct a mixed kernel regression is that every observed covariate configuration $\mathbf{z}_i$ is applied once for $\mathbf{z}_0$.

slope $\widehat{\beta}(\mathbf{z}_0)$ is computed, the minimization calculus corresponds to a local linear regression, if no slopes are computed a local constant regression is estimated. Note that the local slopes are only computed for continuous covariates, in the present case $\ln(y_0)$. Discrete covariates solely appear in the generalized product kernel $W(\cdot)$ that determines the weight of an observation as

$$W(\mathbf{z}_0, \mathbf{z}_i, \boldsymbol{\lambda}) = w_{\ln(y_0)}(\ln(y_{0,0}), \ln(y_{i,0}), \lambda_{\ln(y_0)}) \cdot w_{club}(club_0, club_i, \lambda_{club}).$$

The idea is that all types of covariates are smoothed with a certain weighting (kernel) function $w_k(\cdot)$, corresponding to the scale level of the covariate k. The respective degree of smoothness is determined by an individual smoothing parameter (λ_k), the bandwidth:

> We use a second order Gaussian kernel $w_{\ln(y_0)} = {\lambda_{\ln(y_0)}}^{-1}\phi\left(\ln(y_{i,0}) - \ln(y_{0,0})/\lambda_{\ln(y_0)}\right)$ for the continuous covariate $\ln(y_0)$, where ϕ is the standard normal density and $\lambda_{\ln(y_0)} \in (0, \infty)$. Due to the Gaussian distribution, large $\lambda_{\ln(y_0)}$ values yield almost equal weights for all observations, while points close to $\ln(y_{i,0})$ (i.e., $\ln(y_{i,0}) \approx \ln(y_{0,0})$) get higher weights than points near the border of the neighborhood for regular values of $\lambda_{\ln(y_0)}$.
>
> For the categorical variable $club_i$ we use a kernel function of Racine & Li (2004) $w_{club} = \lambda_{club}^{I(club_i \neq club_0)}$ with the usual indicator function $I(\cdot)$. The bandwidth λ_{club} lies in the interval $[0, 1]$. For a value of $\lambda_{club} = 0$, the kernel w_{club} is an indicator function for category $club_i$ and for $\lambda_{club} = 1$ the kernel function (i.e., the weight) is constant over all categories (i.e., clubs). As categorical covariates are not included in a local linear but a local constant fashion, constant weights across categories correspond to irrelevance of the categorical covariate.

The remaining problem consists of determining optimal values for the bandwidths contained in vector $\boldsymbol{\lambda} = (\lambda_{\ln(y_0)}, \lambda_{club})'$. We obtain $\widehat{\boldsymbol{\lambda}}$ using least-squares cross-validation, a data-driven approach where $\widehat{\boldsymbol{\lambda}}$ is the minimizer of the objective function

$$CV\left(\widetilde{\boldsymbol{\lambda}}\right) = n^{-1}\sum_{i=1}^{n}(g_{i,t} - \widehat{G}_{-i}(z_i))^2 M(z_i).$$

$\widehat{G}_{-i}(z_i)$ is the leave-one-out kernel estimator of regression function, and M is a weighting function bounded between 0 and 1, usually set to $M = 1$ (see Li & Racine, 2004). In contrast to a classical (frequency) approach, the obvious advantage of this method is potential smoothing of categorical variables. Hence, in this context it is possible to reliably estimate club and country specific convergence effects even for small clubs. With respect to the empirical explorations on transitional behavior of countries discussed in Phillips & Sul (2009, Section 5.2), which rely on a sufficiently large number of countries in each club (or group), smoothing may provide a remedy.

3.2. Simultaneous treatment of nonlinearity and heterogeneity

Mixed kernel regressions can be used to simultaneously tackle both key issues of classical growth convergence regressions raised in Section 1.

— Insert Table 1 around here —

In order to illustrate this claim Table 1 exemplarily displays four combinations of estimated bandwidths for the covariates $club$ and $\ln(y_0)$. These four cases correspond to different amounts of nonlinearity (compare Table rows) and heterogeneity (compare Table columns):

> A small[15] estimated bandwidth $\widehat{\lambda}_{\ln(y_0)}$ allows for a higher degree of curvature in the functional relationship between the continuous covariate $\ln(y_0)$ and the response variable g_i. In cases 1 and 2 of Table 1, $\widehat{\lambda}_{\ln(y_0)}$ indicates a higher degree of nonlinearity, while the relatively large bandwidth value in cases 3 and 4 corresponds to less curvature in favor of a linear relationship, respectively.
>
> Estimation of the bandwidth of the club variable λ_{club} deals with the question of uncertainty of club composition. Using the kernels proposed by Racine & Li (2004), the estimated bandwidth is bounded between 0 and 1. Cases 1 and 3 correspond to values of λ_{club} close to zero in (11), implying that for estimating club-specific effects only observations of this club are used. This occurs when the functional form is sufficiently different w.r.t. the clubs or if sufficiently different convergence behavior is present. We can interpret this in favor of club convergence as this also indicates a rather low probability of club misspecification. On the other hand, increasing values of the bandwidth indicate an increase of the misclassification probability. If λ_{club} approaches 1 as in cases 2 and 4, observations from all clubs are used to estimate club-specific effects, indicating irrelevance of the regressor $club$ and hence a very weak or even non-existent club structure (i.e., evidence for global convergence).

The present nonparametric approach allows to estimate country-specific convergence effects and to assess the uncertainty w.r.t. potential club membership misclassification via the estimated bandwidth of the regressor $club$. In practice, the estimated bandwidths for the club variable (for different sorting periods s) frequently is rather low (< 0.001), while the estimated bandwidths for the continuous covariate fluctuates between rather small values and rather large values. Hence, one has to distinguish between cases 1 and 3 of Table 1.[16]

[15] "Small" and "large" are interpreted relative to the scales of continuous covariate and response variable.

[16] Of course, the subsequent considerations remain valid for distinguishing between case 2 and case 4.

If case 1 is relevant, a nonparametric convergence regression seems preferable. If case 3 is relevant, even if the nonparametric fit can mimic the fit of a linear parametric least squares regression, the nonparametric estimator is less efficient. Hence, for case 3 our interpretations refer to the parametric case with club-specific intercept and slope. In order to distinguish between cases 1 and 3, we test the null hypothesis of a correct parametric specification as suggested by Hsiao et al. (2007). The test equation is a mixed kernel regression of the linear parametric residuals on the covariates $club$ and $\ln(y_0)$. A nonlinear fit is equivalent to a non-constant estimated partial effect w.r.t. $\ln(y_0)$ (i.e., convergence effect). In other words: If the null is not rejected, we use the parametric estimator and obtain a club-specific convergence effect that is equal for all countries of a club (of course, the convergence effect of different clubs can be equal, too). Following Durlauf & Johnson (1995), the club-based effects for each sorting time period s can be interpreted to represent averages of the underlying individual effects for each country. If we reject the null, we use the nonparametric estimator and obtain a country-specific convergence effect varying with both the respective club and initial income of the countries. We follow the suggestion of Hsiao et al. (2007) to use a wild-bootstrap, as the rate of convergence towards the asymptotic test distribution is very slow.

4. Application to Penn World Table data

In this section the proposed method is applied to a Penn World Table (PWT) data set, version 7.1, covering the 41 time periods from 1970 to 2010 for 146 countries (see Heston et al., 2012).

For the club identification algorithm for each of the 41 sorting periods we neither find global growth convergence (to one club) nor a polarization process to generate only clubs of high and low income countries. Regarding the PWT data up to 2010 we observe a convergence to at least six different club levels (in contrast to Phillips & Sul, 2009, who use PWT data, version 6.2). There seem to be a variety of forces acting to create different club levels and changing club membership. Of the 146 countries, 43 remain in the same club for all sorting periods. 42 of these countries are members of club 1, which in parts consist of high-income countries (w.r.t. initial income) like Australia and the United States and low-income countries like Botswana and China. The Democratic Republic of Congo is the only country that is a persistent member of the divergence group, as Phillips & Sul (2009) put it is "caught in a . . . pattern of transitional divergence". Some countries show a rather similar leapfrogging/falling-back behavior like Japan, France, Germany, Italy, and Switzerland, they fall-back from club 1 to club 2 in 1986 and leapfrog to club 1 in 2002.

Subsection 4.1 covers the nonparametric convergence regression results and Subsection 4.2 relates to Figure 1 and covers a detailed discussion of empirical results and interpretations exemplarily for four countries.[17]

4.1. Nonparametric growth regression results

The nonparametric regression (10) is estimated for each of the 41 sorting periods. Note that all regression variables are identical for the 41 regressions, apart from the club covariate. For every nonparametric regression we obtain an estimated bandwidth of the club-covariate < 0.001, indicating a high relevance of the club covariate and thus a low probability of club misspecification. Hence we do not have to account for cases 2 and 4 of Table 1. In the subsequent paragraphs, we detail the results of the nonparametric regression exemplarily for two sorting periods, 1973 and 1974. The latter is the first sorting period, where the test of Hsiao et al. (2007) rejects correct specification of a parametric model with club-specific intercepts and slopes. For both sorting periods, we obtain 7 categories (6 clubs and a divergence group).

Clubs based on 1974-income sorting

Using $s = 1974$ as sorting period, Figure 5 contains the estimated nonparametric regression function for each of the six clubs. Each curve is separately plotted in Figure 6 where dashed lines visualize pointwise asymptotic 95%-confidence intervals. The plots suggest a decent fit to the data points and a moderate amount of nonlinearity for clubs 1 to 4. This finding is in line with the rather small value of $\widehat{\lambda}_{\ln(y_0)} \approx 0.75$, corresponding to case 1 of Table 1. The regression curves are monotonically decreasing (apart from the leftmost region of the regression curve for club 2). This is in line with the theory on β-convergence, as poorer countries should grow faster, allowing for global (or within club) convergence. In Figure 5, with increasing club number, the fitted curves are located closer to the bottom left corner. Hence, a higher club number corresponds to countries that either have a lower initial income or a smaller growth rate. This result can be expected from the clubbing algorithm of PS, as club 1 usually consists of the countries with the highest (sorting period) income, together with the countries that have similar transition paths compared to the former. Club 2 usually consists of the highest income countries of the remaining countries (i.e., all that are not in club 1), together with countries with similar transition paths, and so forth. Each club has at least one country with a negative growth. Remarkable is the negative growth rate of all countries of club 6 (Burundi, Liberia,

[17] All computations in this paper are done using the software `R` (), version 3.1.1, and version 0.60-2 of the `np`-package of Hayfield & Racine (2008). In terms of reproducibility, a seed of 42 is used.

and Zimbabwe) despite their low initial income.

— Insert Figures 5 and 6 around here —

Due to the local linear approach of Equation (11), estimated partial effects $\widehat{\beta}(\mathbf{z}_0)$ w.r.t. the continuous covariate $\ln(y_0)$) are obtained. These estimated convergence effects are plotted in Figures 7 and 8. The asymptotic confidence bands in Figure 8 indicate that estimated partial effects are significantly different from zero for all clubs (apart from the effect for the three low initial income countries of club 2).[18] For clubs 5 and 6 we obtain a nearly constant convergence effect, which already could be expected by inspection of the corresponding approximately linear regression curves in Figure 5. Clubs 1 to 4 exhibit nonconstant and nonlinear estimated partial effects implying country-specific convergence effects. For the countries of club 1 we obtain a maximal estimated partial effect at $\ln(y_0) \approx 7.4$. This means that for countries with a log per capita income around 7.4 in 1970 (corresponding to a year 2005 worth of $\exp(7.4) \approx 1,636$ International Dollars), a marginal income increase reduces the growth rate less than for countries with very small or very large initial income. Hence, even though the 65 countries of club 1 have rather homogeneous transition paths, they behave rather heterogeneously in terms of the convergence effects. While the parametric specification allows for club-specific convergence effects due to the club-specific slopes, sorting period 1974 nicely reveals that the nonparametric specification in addition allows for data-driven estimation of country-specific convergence effects.

— Insert Figures 7 and 8 around here —

Clubs based on 1973-income sorting

For sorting according to $s = 1973$ income we obtain an estimated bandwidth $\widehat{\lambda}_{\ln(y_0)} \approx 2.3$, more than three times larger than the corresponding estimated bandwidth obtained from sorting period $s = 1974$. Given that the standard deviation of $\ln(y_0)$ is only around 1.18, this corresponds to a substantial difference in curvature. Hence, the 1973-sorting corresponds to case 3 of Table 1. The corresponding fitted regression curves are approximately linear (compare Figure 9), implying convergence effects that are constant for all club members (compare Figure 10). Equivalent results can be obtained, if a linear parametric regression is estimated,

[18] We simply used the asymptotic distribution of the nonparametric estimator (that has a lower rate of convergence than usual parametric estimators have) for determining the confidence bounds.

as long as distinct intercept and slope (w.r.t. $\ln(y_0)$) parameters are included for each club.[19] The estimated convergence effect of clubs 5 and 6 is remarkably smaller than that of the other clubs, indicating that a marginal increase in income yields a more pronounced reduction of the growth rate for these countries. A similar result is obtained for the club structure corresponding to the 1974-sorting (compare Figure 7). As the estimated bandwidth for $\ln(y_0)$ indicates, the null of a linear parametric regression cannot be rejected for the test of Hsiao et al. (2007) for the 1973-sorting, confirming the visual impression of both figures.

— Insert Figures 9 and 10 around here —

All nonparametric (local linear) estimation results emphasize the usefulness of the mixed kernel approach for the present application. The approach allows for linear or nonlinear functional relationships as well as the smoothing of discrete covariates, from weak to strong. The latter may be interpreted as indicator for the quality of the club covariate and thus allows for global or club convergence and for country-specific convergence effects, whenever $\widehat{\lambda}_{\ln(y_0)}$ is small.

4.2. Analysis for Botswana, France, Germany, and the United States

A comparison of four countries is intended to highlight the steps from the initial data to the convergence regression results and to illustrate the estimation results for countries that remain in the same club throughout (for example Botswana and the United States) and countries that fall-back or leapfrog (for example France and Germany).

— Insert Figure 11 around here —

The top panel of Figure 11 shows the per capita income time-series $y_{i,t}$ for the four countries (in International Dollars of 2005). The kink in 2009, visible in each of the four positively trending time series, indicates the financial crisis. Although the $y_{i,t}$-difference of Botswana and the United States increases over time, the difference in $\ln(y_{i,t})$ decreases.

The second panel of Figure 11 shows a scatterplot of initial log per capita income $\ln(y_{i,0})$ in 1970, against the average annual growth rate $g_{i,T}$ defined in Equation (1) and thus of covariate and response of the absolute β-convergence regression (1). We observe that France, Germany, and the United States exhibit similar growth rates, while the initial income $\ln(y_0)$ of the United

[19] Note that the slopes of clubs 1 to 4 are rather similar.

States is somewhat larger than that of France or Germany. While Botswana has by far the lowest initial income, it has a much higher average annual growth rate of almost 5 percent.

Panels 1 and 3 of Figure 11 briefly summarize the panel data information usually exploited for regression (1) and the nonlinear dynamic factor model (2). Panels 3 and 4 are intended to illustrate the main insights of the paper:

> Club formation, leapfrogging and falling-back: Integers at the bottom of panels 3 and 4 indicate the total number of clubs for the respective sorting period s, $s = 1970, \ldots, 2010$. There are periods of higher heterogeneity in convergence behavior across countries yielding a larger total number of clubs for sorting periods 1975 to 1982 and 1986 to 2001. An asterisk indicates that there is no change in club composition for the current sorting period compared to the previous one: The 7* in 1971 indicates that there are six clubs and a divergence group w.r.t. sorting period $s = 1971$ and the members in all clubs are exactly the same as for sorting according to $s = 1970$. Interestingly, France and Germany both fall-back from club 1 to club 2 in 1986 and leapfrog back to club 1 in 2002.
>
> Estimated time-varying club- and country-specific convergence effects: Depending on the results of the test of Hsiao et al. (2007), we label the use of a parametric convergence regression with club-specific intercept and club-specific convergence effect (i.e., case 3 of Table 1 is relevant) by using a "P" at the bottom of the panel. If the linear parametric null is rejected (i.e., case 1 of Table 1 is relevant), the use of a nonparametric regression allowing for country-specific convergence effects is labeled by using an "N". Use the wild bootstrap to determine the distribution of the test under the null, we employ a relatively large significance level of 0.1 to ensure that the test has enough power in detecting a misspecified parametric model. Of course, the estimated convergence effects for the "P"-periods are equal for France and Germany, due to the membership in the same club for every sorting. The country-specific convergence effects, obtained for the "N"-periods, are nearly equivalent as Germany and France have a similar initial per capita income. Asymptotic 95%-confidence bounds for the estimated effect reveal that for France and Germany, the estimated convergence effects are always significantly negative. Panel 4 shows that Botswana and the United States are permanent members of the best club 1. However, estimated convergence effects can be remarkably different over sorting periods. For Botswana, the convergence effect for these periods is smaller in absolute value than for the United States (as the corresponding fitted regression curve has a

very flat slope at the $\ln(y_0)$-position of Botswana) and is often not significantly different from 0. The convergence effects for Botswana and the United States are very different, whenever the nonparametric convergence regression is applied, while the level of the parametric convergence effect is somewhere in-between. The intra-club heterogeneity is quite high for club 1 in the "N"-periods. This is a plausible result, as the higher the intra-club heterogeneity is in terms of sufficiently different convergence speed, the higher will be the difference of country-specific and club-specific convergence effects, and thus the more likely is a rejection of the parametric specification. This seems to stand in sharp contrast with Botswana and United States being members of one club. But the clubbing algorithm is intended to combine countries that have a similar relative transition path in a club, this can even be the case even if the convergence effect is remarkably different.

5. Conclusion

Nonlinear dynamic factor panel data models offer a parsimonious framework to explain time-varying growth convergence or divergence patterns of heterogeneous cross-sections of countries or groups of countries. They improve on a substantive literature dealing either with heterogeneities or nonlinearities, both neglected in classical approaches of growth and convergence regressions.

This paper simultaneously addresses both issues by proposing a nonparametric framework building on the recent literature on (nonlinear) dynamic factor models such as Phillips & Sul (2007b, 2009) and Kuersteiner & Prucha (2013). The former deliver an economic and econometric framework to study individual growth paths relative to a useful benchmark, enabling the identification of membership in or transition between convergence clubs (or a divergence group). The latter allow cross-sectional interactions not only in the systematic but also in the error component of (linear in parameters) panel data regression models. In contrast our focus lies in the data-driven, flexible specification of the systematic component of convergence regressions: From a statistical point of view in order to avoid inconsistent marginal effects and cross-section association structures in the error components due to misspecification of the systematic part of the regression. From an economic point of view, the proposed nonparametric convergence regressions allow to estimate within club convergence effects while controlling for between club association structures. Furthermore, by allowing for leapfrogging or falling-back, the proposed modeling framework sheds light on the relationship between initial income, technological differentials or resource endowments, and the impact of this relationship on convergence, divergence, or transition (see Phillips & Sul, 2009, Section 5).

Employing an augmented version of the club identification algorithm of Phillips & Sul (2007a,b, 2009), we find that the algorithm is sensitive with respect to use of different initial sorting variables. We detail how this sensitivity can be motivated by modern growth theory. In particular, the growth theoretical model allows for an explanation of the change in number, size, and composition of clubs due to leapfrogging and falling-back of countries. Nonparametric mixed kernel growth regressions are able to simultaneously account for neglected heterogeneity and nonlinearity by smoothing discrete (e.g., club) and continuous (e.g., initial income) covariates. The approach allows for an analysis of convergence effects on both country and club level while alleviating potential misclassification in the club formation process using simultaneous smoothing over the club structure. The concept of convergence toward some benchmark (i.e., club level) acknowledges the possibility of a time-varying influence of different covariates on club formation, leapfrogging and falling-back of countries.

Our empirical exercise using Penn World Table data finds considerable evidence in favor of club-based convergence in the sense of PS, i.e., in favor of heterogeneity in convergence analysis and for neglected nonlinearity in classical growth convergence regressions. In addition, we observe changes of number, size, and composition of clubs over time. We explain the corresponding catching-up, leapfrogging, falling-back, and change in club membership over time by stochastic growth model variants detailed in Section 2.2. The club covariate and hence the mixed kernel results are different for different sorting periods, as estimated regression curves show a different amount of nonlinearity. Potential heterogeneity and nonlinearity of growth regressions can be accounted for in a data-driven fashion by the nonparametric approach without invoking potentially problematic (parametric) a priori assumptions.

The nonparametric approach can be extended in the following ways: First, the club-based nonparametric convergence regression can be adapted to the concept of conditional convergence by including additional (continuous and/or discrete) covariates, while the interpretations in the spirit of Table 1 remain valid. Second, as extension to the interpretation of estimated bandwidths, the significance of the corresponding covariates can be checked by the tests of Racine (1997) (for continuous covariates) and Racine et al. (2006) (for discrete covariates). Third, a Monte Carlo-based comparison of the prediction performance, as proposed by Haupt et al. (2010) can be conducted in order to also compare different (non- or semi-)parametric specifications.

Acknowledgements

An earlier version of the paper was presented at the 7th International Conference on Computational and Financial Econometrics in London, 2013. We thank the participants for helpful comments that led to an improvement of the paper. Special thanks go to Verena Petring for implementing a first version of the PS algorithm in `R` and further valuable research assistance and discussions. Any remaining errors are of course ours.

References

Acemoglu, D., Zilibotti, F., & Aghion, P. (2006). Distance to frontier, selection, and economic growth. *Journal of European Economic Association*, 4(1), 37–74.

Alexiadis, S. & Tomkins, J. (2004). Convergence clubs in the regions of Greece. *Applied Economics Letters*, 11, 387–391.

Alfò, M., Trovato, G., & Waldmann, R. (2008). Testing for country heterogeneity in growth models using a finite mixture approach. *Journal of Applied Econometrics*, 23, 487–514.

Barro, R. & Sala-i Martin, X. (1992). Convergence. *The Journal of Political Economy*, 100(2), 223–251.

Barro, R. & Sala-i Martin, X. (2004). *Economic Growth*. The MIT Press.

Barro, R., Sala-i Martin, X., Blanchard, O., & Hall, R. (1991). Convergence across states and regions. *Brookings Papers on Economic Activity*, 1, 107–182.

Barro, R. J. & Sala-i Martin, X. (1997). Technological diffusion, convergence, and growth. *Journal of Economic Growth*, 2, 1–27.

Bartkowska, M. & Riedl, A. (2012). Regional convergence clubs in europe: Identification and conditioning factors. *Economic Modelling*, 29(1), 22–31.

Ben-David, D. (1998). Convergence clubs and subsistance economies. *Journal of Development Economics*, 55(1), 155–171.

Berthelemy, J. & Varoudakis, A. (1996). Economic growth, convergence clubs, and the role of financial development. *Oxford Economic Papers*, 48, 300–328.

Brezis, E., Krugman, P., & Tsiddon, D. (1993). Leapfrogging in international competition: A theory of cycles in national technological leadership. *American Economic Review*, 83(5), 1211–1229.

Brock, W. & Mirman, L. (1972). Optimal economic growth and uncertainty: the discounted case. *Journal of Economic Theory*, 4, 479–513.

Canarella, G. & Pollard, S. (2004). Parameter heterogeneity in the classical growth model: A quantile regression approach. *Journal of Economic Development*, 29, 1–31.

Canova, F. (2004). Testing for convergence clubs in income per capita: A predictive density approach. *International Economic Review*, 45, 49–77.

Chatterji, M. (1992). Convergence clubs and endogenous growth. *Oxford Review of Economic Policy*, 8(4), 57–69.

Cohen-Cole, E., Durlauf, S., & Johnson, P. (2012). Nonlinearities in growth: From evidence to policy. *Journal of Macroeconomics*, 34, 42–58.

Dalgaard, C.-J. & Hansen, J. (2005). Capital utilization and the foundations of club convergence. *Economics Letters*, 87(2), 145–152.

Dasgupta, P. & Heal, G. (1974). The optimal depletion of exhaustible resources. *The Review of Economic Studies 41 (Symposium on the Economics of Exhaustible Resources)*, (pp. 3–28).

De Siano, R. & D'Uva, M. (2006). Club convergence in European regions. *Applied Economics Letters*, 13, 569–574.

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

Durlauf, S. & Johnson, P. (1995). Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics*, 10, 365–384.

Durlauf, S., Johnson, P., & Temple, J. (2005). Growth econometrics. In P. Aghion & S. Durlauf (Eds.), *Handbook of Economic Growth*, volume 1A. Amsterdam: North Holland.

Durlauf, S., Kourtellos, A., & Tan, C. (2008). Are any growth theories robust? *Economic Journal*, 118, 329–346.

D'Uva, M. & De Siano, R. (2007). Human capital and "club convergence" in Italian regions. *Economics Bulletin*, 18(1), 1–7.

Dwyer, G., Devereux, J., Baier, S., & Tamura, R. (2013). Growth, recessions and banking crises. *Journal of International Money and Finance*, 38, 18–40.

Fiaschi, D. & Lavezzi, A. (2007). Nonlinear economic growth: Some theory and cross-country evidence. *Journal of Development Economics*, 84, 271–290.

Fischer, M. & Stirböck, C. (2006). Pan-European regional income growth and club-convergence - Insights from a spatial econometric perspective. *The Annals of Regional Science*, 40, 693–721.

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

Graham, B. S. & Temple, J. R. (2006). Rich nations, poor nations: how much can multiple equilibria explain? *Journal of Economic Growth*, 11, 5–41.

Greiner, A. & Semmler, W. (2005). *The Forces of Economic Growth, A Time Series Perspective*. Princeton University Press.

Greiner, A. & Semmler, W. (2008). *The Global Environment, Natural Resources, and Economic Growth*. Oxford University Press.

Gruene, L., Semmler, W., & Stieler, M. (2013). Using nonlinear model predictive control for dynamic decision problems in economics. Working paper.

Hauk Jr., W. R. & Romain, W. (2009). A Monte Carlo study of growth regressions. *Journal of Economic Growth*, 14, 103–147.

Haupt, H. & Petring, V. (2011). Assessing parametric misspecification and heterogeneity in growth regression. *Applied Economics Letters*, 18, 389–394.

Haupt, H., Schnurbus, J., & Tschernig, R. (2010). On nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, 5, 894 – 901.

Hayfield, T. & Racine, J. (2008). Nonparametric econometrics: The `np` package. *Journal of Statistical Software*, 27, 1–32.

Henderson, D. (2010). A test for multimodality of regression derivatives with application to nonparametric growth regression. *Journal of Applied Econometrics*, 25, 458–480.

Henderson, D., Papageorgiou, C., & Parmeter, C. (2013). Who benefits from financial development? New methods, new evidence. *European Economic Review*, 63, 47–67.

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

Heston, A., Summers, R., & Aten, B. (2012). Penn World Table version 7.1 Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania. Working paper.

Hodrick, R. & Prescott, E. (1997). Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.

Howitt, P. & Mayer-Foulkes, D. (2005). R&D, implementation, and stagnation: A Schumpeterian theory of convergence clubs. *Journal of Money, Credit and Banking*, 37(1), 147–177.

Hsiao, C., Li, Q., & Racine, J. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics*, 140, 802–826.

Kalaitzidakis, P., Mamuneas, T., Savvides, A., & Stengos, T. (2001). Measures of human capital and nonlinearities in economic growth. *Journal of Economic Growth*, 6, 229–254.

Kuersteiner, G. & Prucha, I. (2013). Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics*, 174, 107–126.

Li, Q. & Racine, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14, 485–512.

Li, Q. & Racine, J. (2008). Nonparametric estimation of conditional cdf and quantile functions with mixed categorical and continuous data. *Journal of Business and Economic Statistics*, 26, 423–434.

Liu, Z. & Stengos, T. (1999). Non-linearities in cross-country growth regressions: A semi-parametric approach. *Journal of Applied Econometrics*, 14, 527–538.

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

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

Masanjala, W. & Papageorgiou, C. (2004). The solow model with CES technology: Nonlinearities and parameter heterogeneity. *Journal of Applied Econometrics*, 19, 171–201.

Meliciani, V. & Peracchi, F. (2006). Convergence in per-capita GDP across European regions: a reappraisal. *Empirical Economics*, 31, 549–568.

Mirestean, A. & Tsangarides, C. (2015). Growth determinants revisited using limited-information Bayesian model averaging. *Journal of Applied Econometrics*, to appear.

Moral-Benito, E. (2014). Growth empirics in panel data under model uncertainty and weak exogeneity. *Journal of Applied Econometrics*.

Panopoulou, E. & Pantelidis, T. (2009). Club convergence in carbon dioxide emissions. *Environmental and Resource Economics*, 44, 47–70.

Phillips, P. & Sul, D. (2003). The elusive empirical shadow of growth convergence. *Cowles Foundation Discussion Paper No. #1398*.

Phillips, P. & Sul, D. (2007a). Some empirics on economic growth under heterogeneous technology. *Journal of Macroeconomics*, 29, 455–469.

Phillips, P. & Sul, D. (2007b). Transition modeling and econometric convergence tests. *Econometrica*, 75(6), 1771–1855.

Phillips, P. & Sul, D. (2009). Economic transition and growth. *Journal of Applied Econometrics*, 24, 1153–1185.

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

Quah, D. (1996a). Empirics for economic growth and convergence. *European Economic Review*, 40, 1353–1375.

Quah, D. (1996b). Regional convergence clusters across europe. *European Economic Review*, 40, 951–958.

Quah, D. (1996c). Twin peaks: Growth and convergence in models of distribution dynamics. *The Economic Journal*, 106, 1045–1055.

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

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Racine, J. & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119, 99–130.

Racine, J. S. (1997). Consistent specification testing for nonparametric regression. *Journal of Business and Economic Statistics*, 15(3), 369 – 378.

Racine, J. S., Hart, J., & Li, Q. (2006). Testing the significance of categorical predictor variables in nonparametric regression models. *Econometric Reviews*, 25(4), 523 – 544.

Sala-i Martin, X. (1996a). The classical approach to convergence analysis. *The Economic Journal*, 106, 1019–1036.

Sala-i Martin, X. (1996b). Regional cohesion: Evidence and theories of regional growth and convergence. *European Economic Review*, 40, 1325–1352.

Santos, M. & Vigo-Aguiar, J. (1995). *Accuracy estimates for a numerical approach to stochastic growth models*. Discussion paper 107, Institute for Empirical Macroeconomics, Federal Reserve Bank of Minneapolis.

Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, 70, 65–94.

Startz, R. (1998). Growth states and shocks. *Journal of Economic Growth*, 3, 203–215.

Sul, D., Phillips, P., & Choi, C.-Y. (2005). Prewhitening bias in HAC estimation. *Oxford Bulletin of Economics and Statistics*, 67(4), 517–546.

A. Tables and Figures

Table 1: Cases w.r.t. estimated bandwidths.

	$\widehat{\lambda}_{\texttt{club}} \approx 0$	$\widehat{\lambda}_{\texttt{club}} \approx 1$	interpretation
small $\widehat{\lambda}_{\ln(y_0)}$	case 1	case 2	nonlinear regression
large $\widehat{\lambda}_{\ln(y_0)}$	case 3	case 4	linear regression
interpretation	club-based convergence	global convergence	

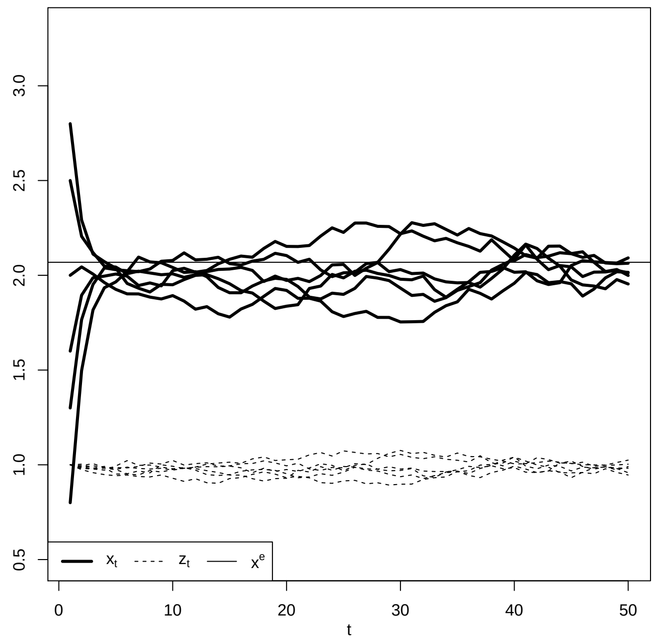

Figure 2: Stochastic growth paths. Configuration: $\alpha = 0.34, \beta = 0.95, A = 5, \sigma = 0.015, \rho = 0.9, z_0 = 1, x_0 \in \{0.8, 1.3, 1.6, 2, 2.5, 2.8\}$.

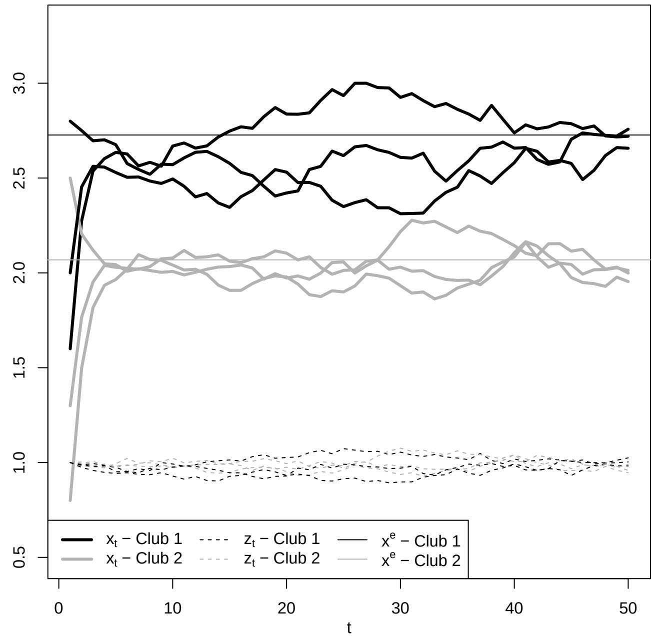

Figure 3: Club convergence behavior. Configuration: $\alpha = 0.34, \beta = 0.95, A \in \{5, 6\}, \sigma = 0.015, \rho = 0.9, z_0 = 1, x_0 \in \{0.8, 1.3, 2.5, 1.6, 2, 2.8\}$.

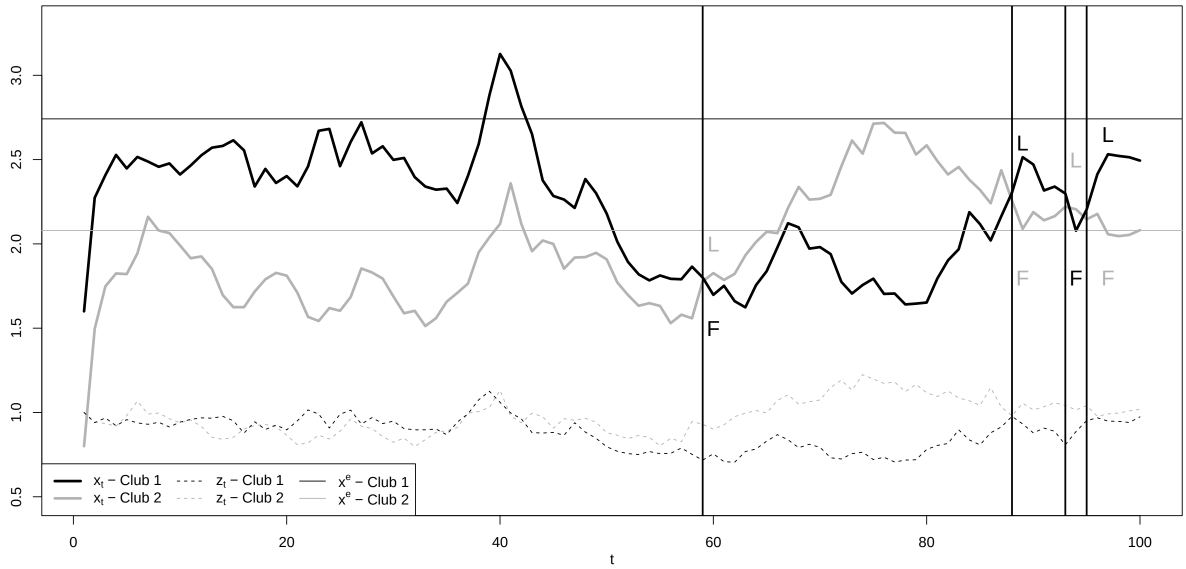

Figure 4: Leapfrogging and falling-back (switching of leading and lagging position). Configuration: $\alpha = 0.34, \beta = 0.95, A \in \{5, 6\}, \sigma = 0.045, \rho = 0.9, z_0 = 1, x_0 \in \{0.8, 1.6\}$.

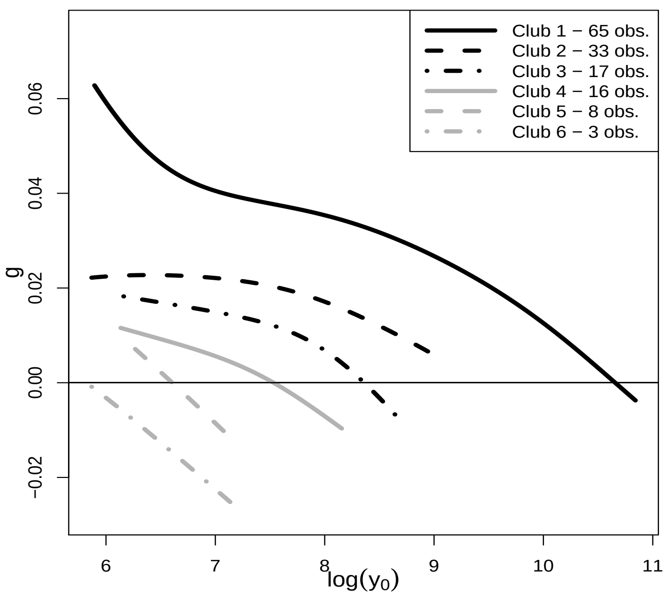

Figure 5: Nonparametrically estimated growth rates w.r.t. (log) initial income and club (sorting period: 1974).

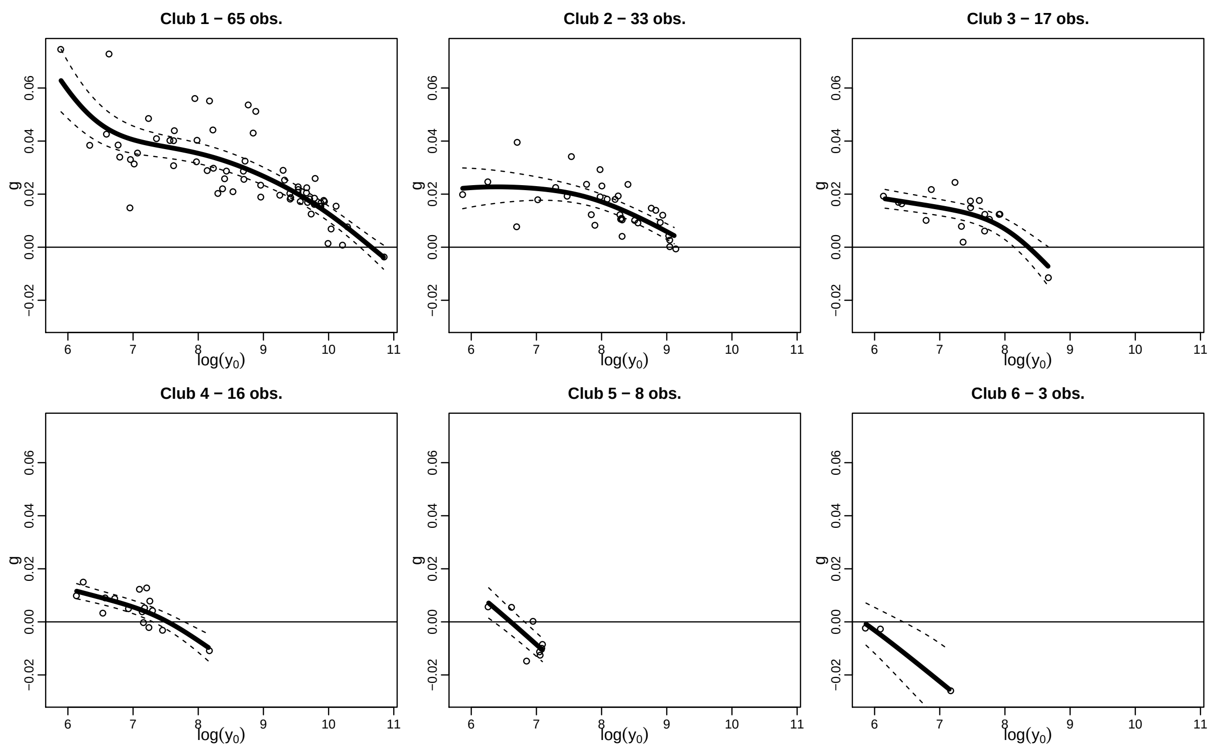

Figure 6: Nonparametrically estimated growth rates and asymptotic (pointwise) 95%-confidence intervals w.r.t. (log) initial income and club (sorting period: 1974). Observed growth rates indicated by points.

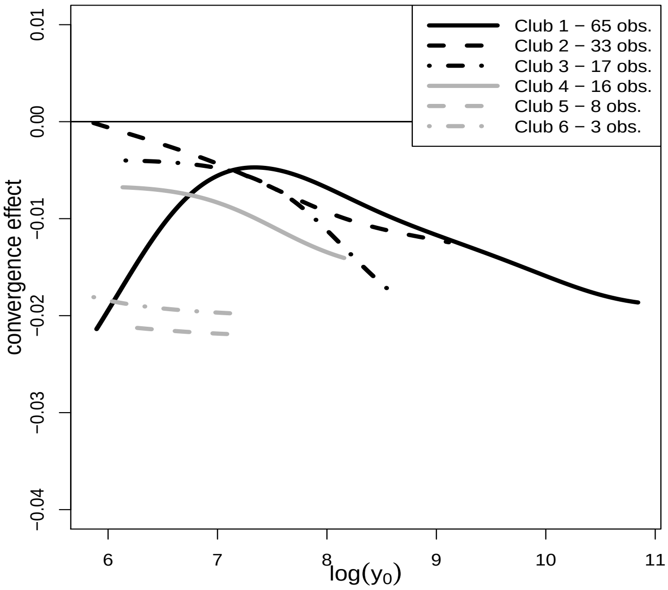

Figure 7: Nonparametrically estimated convergence effects w.r.t. (log) initial income and club (sorting period: 1974).

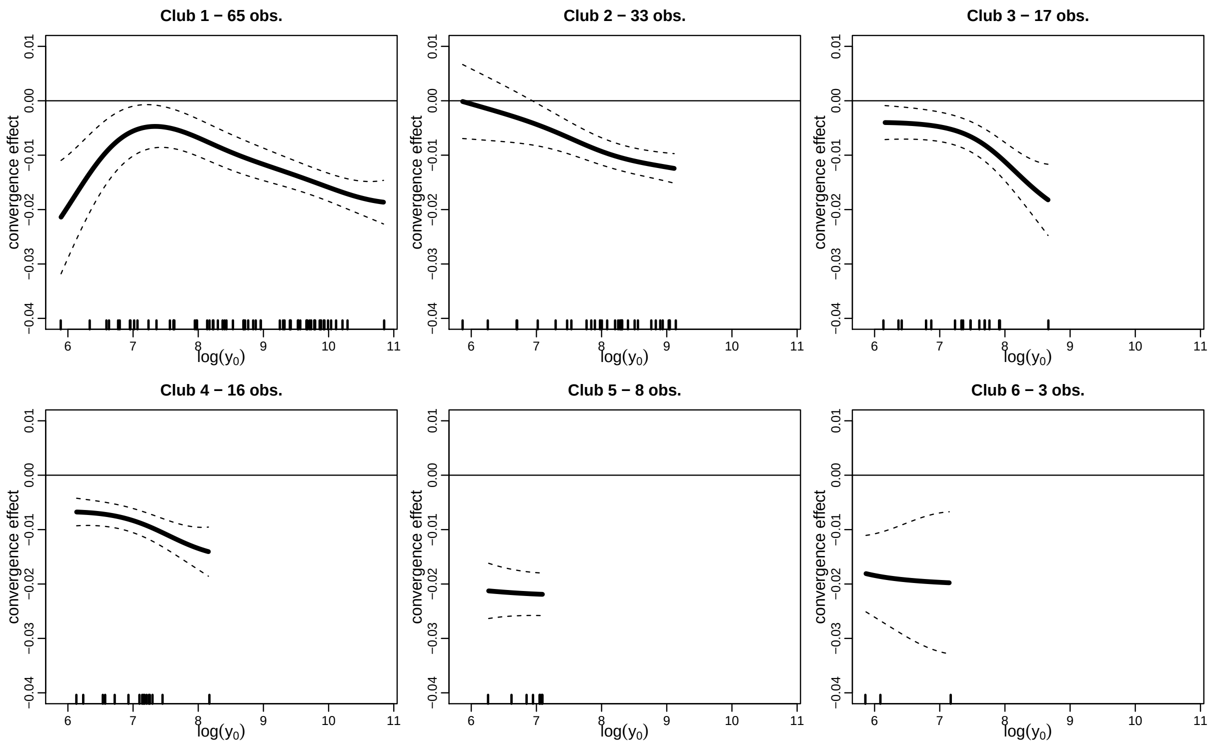

Figure 8: Nonparametrically estimated convergence effects and asymptotic (pointwise) 95%-confidence intervals w.r.t. (log) initial income and club (sorting period: 1974). Observed (log) initial income values indicated at abscissa.

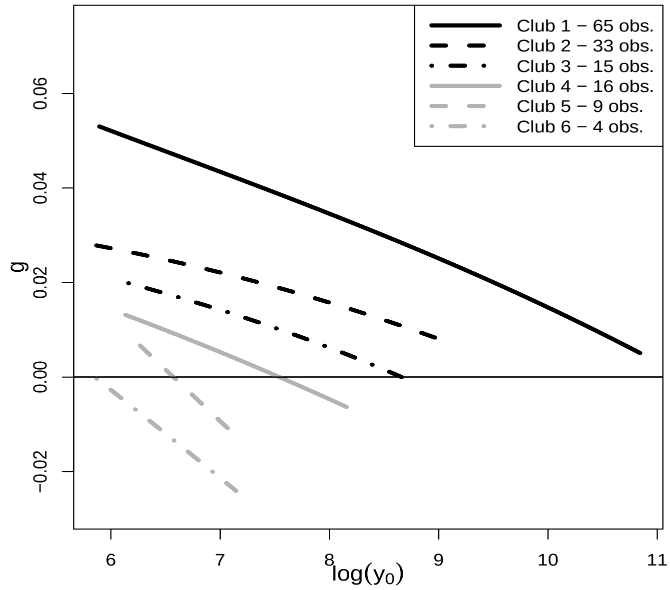

Figure 9: Nonparametrically estimated growth rates w.r.t. (log) initial income and club (sorting period: 1973).

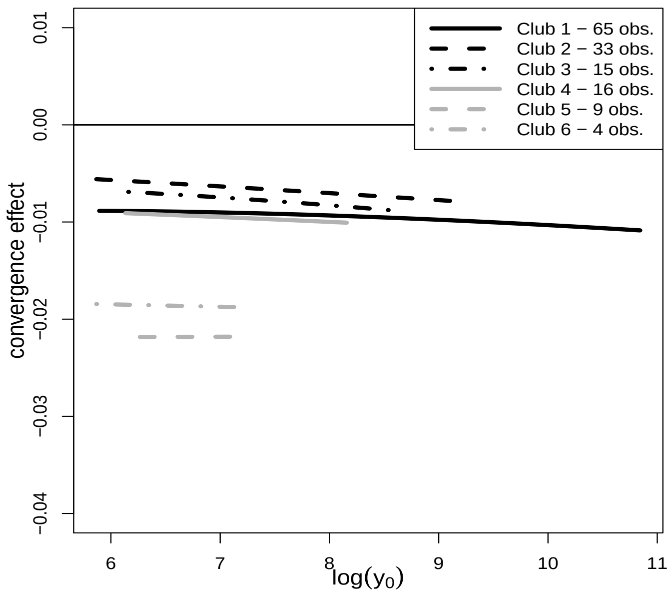

Figure 10: Nonparametrically estimated convergence effects w.r.t. (log) initial income and club (sorting period: 1973).

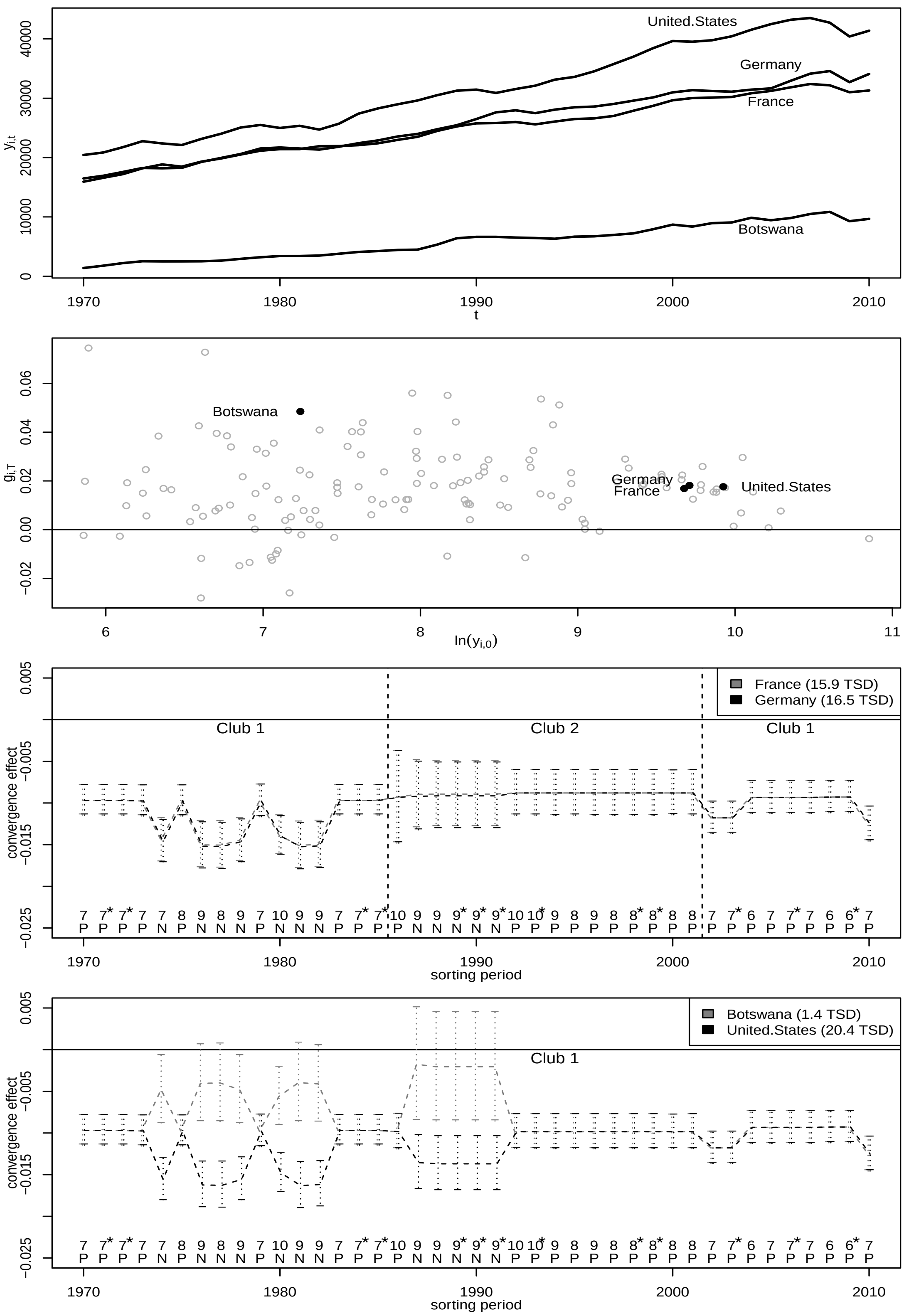

Figure 11: Top display shows time series on incomes for four countries. Second display highlights these countries positions in a plot of all income growth versus initial income data. Third (fourth) display shows convergence effects and club membership over sorting periods for France and Germany (USA and Botswana).