Econometric Analysis of Growth and Convergence

Dissertation

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften (Dr. rer. pol.) der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld

vorgelegt von

Dipl.-Vw. Verena Meier

Bielefeld, im Dezember 2011

Dekan	Prof. Dr. Herbert Dawid
Gutachter	Prof. Dr. Harry Haupt, Universität Bielefeld
Zweitutachter	apl. Prof. Dr. Peter Wolf, Universität Bielefeld
Tag der mündlichen Prüfung	25. Mai 2012

Gedruckt auf alterungsbeständigem Papier gemäß ISO 9706

Vorwort

Diese Arbeit ist während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Lehrstuhl für Ökonometrie und Statistik an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld in der Zeit von April 2008 bis Dezember 2011 entstanden. Ich möchte an dieser Stelle eine Reihe von Personen hervorheben, die auf unterschiedliche Art und Weise zum Gelingen dieser Arbeit beigetragen haben. Zu allererst möchte ich meinem Doktorvater Prof. Dr. Harry Haupt für seine hervorragende Betreuung danken. Er hat durch seine Unterstützung, sein Engagement und durch die Teilhabe an seinen Ideen wesentlich zum Erfolg dieser Arbeit beigetragen. Der Dank gebührt ebenfalls den Kollegen in V9 der Universität Bielefeld für das sehr angenehme Arbeitsklima und die hilfreichen Gespräche. Besonders hervorzuheben sind hier Joachim Schnurbus und Christian Heinze, die mir stets geduldig bei der Lösung vieler technischer Probleme behilflich waren. Herrn apl. Prof. Dr. Peter Wolf danke ich herzlich für seine Arbeit als Zweitgutachter.

Ein großer Dank geht an meinen Mann Mario und meine Eltern Ingrid und Werner für die wertvolle Unterstützung in den letzten Jahren und die Ermutigungen, an meinem Ziel festzuhalten.

Bielefeld, im Mai 2011

Verena Meier

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Motivation

Since formulating the millennium goals, reducing of global poverty and harmonization of worldwide standards of living is one of the main aims and the most complicated challenge for the United Nations. There are many questions which are relevant in this context. For transmitting the right impulses, politicians need to know how those convergence processes can be achieved and understand which determinants are crucial in this context. Thus, in addition to analyzing the theoretical context scientists are consulted dealing with the question of empirical measurability of convergence processes for developing meaningful forecasting models. Convergence processes are conditioned by the existence of economic growth in poor countries. Thus, analyzing the determinants of economic growth is a main aim in convergence analysis. Modeling and analyzing convergence processes is not only important on cross-country but also on regional level. For example, still 20 years after the reunification the convergence process between eastern and western districts in Germany dominates local and nationwide affairs illustrated in the discussion of abolishing the solidarity tax or equalization of eastern and western wage levels.

Purpose and aims

On the one hand this thesis deals with the questions: Which concepts are reliable for measuring economic growth and growth convergence, how do they work and which assumptions are made? Comparing classical and modern convergence and growth concepts on the basis of well-established cross-country datasets the questions above should be answered. Using the results and recent literature the concepts are analyzed w.r.t. their limitations and potential errors. Applying modern statistical and econometric methods the convergence concepts are combined thus the advantages of individual concepts are emphasized while the influences of potential problems are reduced or even eliminated.

On the other hand the question whether growth and convergence concepts offer similar results for different levels of aggregation is studied. Thus, in addition to cross-country data an analysis of German regional data is done.

Structure of this thesis

The thesis is organized as follows. Chapter 2 deals with the first strang of questions: which convergence concepts exist, how do they work and which restrictions exist? In detail, Section 2.1 describes economic growth models. The first one is the neoclassical growth model of Solow (1956) explaining standards of living by growth of population and saving rate and is outlined in Subsection 2.1.1. In 2.1.2 an extension of the neoclassical model proposed by Mankiw et al. (1992) is presented where human capital is considered as an additional determinant. Finally, in Subsection 2.1.3 the spatial augmented Solow model of Ertur & Koch (2007) is discussed. Section 2.2 describes classical convergence concepts. The most popular concept in this context is β-convergence, which is discussed in Subsection 2.2.1. The idea of β-convergence is that poor economies grow faster than rich ones and the concept can be put to test in a linear regression model where the growth rate of per capita income is explained by initial income (compare e.g. Barro et al., 1991). In Subsection 2.2.2

σ-convergence is outlined (see Sala-i Martin, 1996b). σ-convergence is assumed if a reduction of the standard deviation of per capita income over time is achieved. The concept of convergence in relative per capita income (e.g., Quah, 1993a,b, 1997) is presented in Subsection 2.2.3. The idea of this concept is that convergence is assumed if all economies share the same fraction of mean per capita income. Subsection 2.2.4 outlines the relationship between these three concepts. Subsection 2.2.5 describes alternative nonparametric convergence concepts of Maasoumi et al. (2007). Applying these concepts the conditional cumulated density functions (CDF) of estimated growth rates for a priori defined groups of economies are compared. Convergence is assumed if the conditional CDFs of the poorer group stochastically dominates the one of the richer group. Subsection 2.2.6 deals with the question how much of available information is used by the different convergence concepts. Although most datasets consist of yearly data, some of the presented concepts use only parts of this information, e.g. only values from start and final period or data of subperiods. This lack of used information may cause misinterpretation and misspecification. Thus, this Subsection outlines which concepts uses which degree of available information. Subsection 2.2.7 generally discusses the problem of global convergence (all countries convergence) and alternatively presents the idea of club convergence. Club convergence means that if there is no global convergence there may be at least clubs of economies with common convergence behavior. This is a very important point because empirical studies show that in many applications there is no evidence for global convergence.

Chapter 3 deals with problems and limitations of classical convergence analysis and provides methods for avoiding this problems. The first central problem discussed in Section 3.1 is the definition and selection of growth determinants. This Section summarizes different definitions for several growth determinants found in the literature and focuses on the problem, which determinants influence growth and convergence. A second problem is that many convergence concepts assume linear convergence relationships. This assumption is very restrictive and thus controversial. For example, Haupt & Petring (2011) find nonlinear convergence processes. Section 3.2 offers a review of current literature concerning the problem of nonlinearity. Subsection 3.2.1 presents a nonparametric alternative of Racine & Li (2004) allowing for nonlinearities, while Subsection 3.2.2 describes a test of Hsiao et al. (2007) which checks for parametric misspecification and may detect nonlinearities. Another widespread problem of convergence analysis is omitted heterogeneity while heterogeneity may occur in different ways. Phillips & Sul (2003, 2007a,b, 2009) find heterogeneous convergence behavior when explaining average per capita income caused by individual country-specific and time-dependent effects such as individual technology levels. Mello & Perrelli (2003) detect heterogeneous convergence behavior for different parts of cross-country income distributions. Recent literature dealing with the problem of omitted heterogeneity is summarized in Section 3.3. Allowing for heterogeneous convergence behavior over different quantiles of the income distribution quantile regression is described in Subsection 3.3.1. Subsection 3.3.2 outlines a dynamic factor model of Phillips & Sul (2003, 2007a,b, 2009) considering individual effects.

Haupt & Petring (2011) and Haupt & Meier (2011) find that nonlinearities and heterogeneity arise simultaneously in convergence analysis. The argumentation is outlined in Section 3.4. Subsection 3.4.1 describes nonparametric quantile regression as a solution for considering both problems. An alternative method in this context is the two-step procedure of Haupt & Meier (2011) outlined in Subsection 3.4.2. The method includes heterogeneity considered by convergence clubs in a flexible nonparametric convergence model. Finally, Section 3.5 deals with the problem of spatial associations. Several papers find that neighboring regions influence economic growth and convergence of an economy. Subsection 3.5.1 summarizes models capturing different kinds of spatial associations which may occur in three forms, dependence in the dependent variable, the explanatory variables or correlated other effects. Subsection 3.5.2 deals with the problem which neighbors should be considered in modeling spatial associations and which weights individual neighbors should get. Testing procedures for spatial dependence are described in Subsection 3.5.3.

In Chapter 4 the second aim of this thesis is discussed, the question whether there are differences in the analysis of different aggregation levels. The section starts with a summary of current literature. The following subsections describe the datasets on different levels of aggregation which are analyzed in the empirical part of this thesis.

Chapter 5 outlines applications of the methods described in Chapter 3 applied on the data presented in Chapter 4. The application chapter consists of three working papers, which are joint works with Harry Haupt and Joachim Schnurbus. The following descriptions are the summaries taken from the articles. Section 5.1 is a part of Haupt & Petring (2011). "A fully nonparametric analysis is applied to address the problems of nonlinearity and heterogeneity in classical growth regression models using original data from seminal contributions in this field. Nonparametric specification tests and in-sample goodness-of-fit measures, as well as cross-validation based out-of-sample measures provide considerable evidence for parametric misspecification and a superior performance of a nonparametric model, despite the small sample size. In contrast to recent contributions identifying heterogeneity as the primal source of misspecification, a formal and graphical analysis does not reveal evidence for heterogeneity in a parametric and nonparametric quantile regression framework."[1] Section 5.2 refers the application section of Haupt & Meier (2011). The methodical part of the paper is already presented in Subsections 3.3.2 and 3.4.2. "While classical growth convergence regressions fail to account for various sources of heterogeneity and nonlinearity and recent contributions are able to address either the one or the other, this paper presents a simple two-step method to address both issues. Based on a slightly augmented version of a recently proposed clubbing algorithm to identify convergence clubs, we formulate a flexible nonlinear modeling framework which allows for analyzing convergence effects on both individual and club level while alleviating potential misclassification in the club formation process using simultaneous smoothing over the club structure. The proposed method is illustrated with applications to different data."[2] The third application

[1]This abstract is cited from the summary of Haupt & Petring (2011).

[2]This abstract is cited from the summary of Haupt & Meier (2011), version: October 19, 2011.

displayed in Section 5.3 is Haupt et al. (2011). "Classical Solow-type convergence regressions have been found to suffer from at least three sources of misspecification. First, due to latent heterogeneity of convergence processes, second, due to latent spatial associations, and third, due to a too restrictive parametric functional form of the regression function. The recent literature proposes several methods to address one or two of these caveats. As all three sources of misspecification may be tightly related the present paper proposes a comprehensive modeling approach. As a first step — to allow for heterogeneities induced by non-global convergence processes — we identify convergence (and divergence) clubs from a dynamic factor model using panel data. In the second step further potential heterogeneities in the extended model are assumed to be generated by spatial associations between regions in a cross-section model. As an encompassing step we test for parametric misspecification of the extended model and check the validity of the club structure generated from panel data to capture heterogeneity of convergence processes in a cross-section model. The employed nonparametric estimation method allows to investigate potential club-specific nonlinearities. In our empirical application we study growth and convergence of the high-skilled employees using panel data for German regions. Model selection results suggest that including convergence (and divergence) club-specific effects dominate spatially augmented Solow models: The residual heterogeneity in classic models seems to be captured by the club structure identified in the first step of our analysis. If, however, the club information is neglected, spatial econometric tests suggest the existence of spatial association in the model. We check the robustness of our findings with a second application, where we analyze data from the literature used to illustrate the merits of spatial Solow models. Again the evidence is clearly in favor of our findings that spatial associations can be captured by the allowing convergence paths on the club-level."[3]

[3]This abstract is cited from the summary of Haupt et al. (2011), version: November 29, 2011.

Chapter 2

Classical Solow-type growth and convergence modeling

Analyzing economic growth and convergence is one of the main aims of economics. Hence, it is not surprising that there are many growth and convergence models whose structures differ from simple to complex. A famous and simple model is the growth model of Solow (1956). This approach explains standards of living with only two or in the extended version three variables. Since Mankiw et al. (1992), hereafter MRW, published their seminal paper, growth models also get into the focus of econometrics. The authors provide empirical evidence on the classical growth model of Solow (1956). Additionally, MRW analyze a by human capital extended version of the model and investigate the question of convergence. In this chapter classical growth and convergence concepts are presented which are of importance for the following sections in this thesis, while Section 2.1 deals with growth and Section 2.2 with convergence models.

2.1 Growth models

Solow (1956) propose one of the most popular economic growth models explaining standards of living by only two covariates. Although this simple model is about 55 years old it is still a topic in current literature. Several others provide economic extensions of the model and simultaneously its empirical validity is analyzed in econometrics (e.g. Mankiw et al., 1992; Barro, 1991; Barro & Sala-i Martin, 2004). Thus, the following section deals with the Solow model, its extensions and its empirical content. In Section 2.1.1 the Solow model and its derivation is described. Section 2.1.2 outlines the extension of the Solow model by human capital proposed by Mankiw et al. (1992). Finally, in Section 2.1.3 the spatially augmented Solow model of Ertur & Koch (2007) is presented which captures spatial dependence and spillovers.

2.1.1 Neoclassical growth model

The classical growth model of Solow (1956) can be derived by assuming a Cobb-Douglas production function

$$Y = F(K_p, L) = AK_p^{\alpha} L^{1-\alpha}, \quad 0 < \alpha < 1.$$

The total Output Y depends on the factors of production capital (the community's accumulated stock of capital) K_p and labor (the population of working age) L, where α gives the partial elasticity of production for capital and A is a constant scale factor measuring the level of technology. For assessing the growth of labor, the population growth is assumed to be exogenous at a constant rate n so that L can be expressed as

$$L(t) = L_0 e^{nt}. \tag{2.1}$$

Further, it is assumed that the total Output Y can be divided into gross investment and consumption. Gross investment is interpreted as the saving rate s_{k_p} in the sense, that this is the share of total output saved for increasing future output. The net investment results from the difference between gross investment and depreciation as

$$\dot{K}_p = s_{k_p} Y - \delta K_p, \tag{2.2}$$

where δ is the depreciation rate measuring the share of the total output which needs to be invested to hold the actual level of output. The aim of the Solow model is to explain the standard of living approximated by the per capita output

$$y = \frac{Y}{L} = \frac{A K_p^{\alpha} L^{1-\alpha}}{L^{\alpha} L^{1-\alpha}} = A \frac{K_p^{\alpha}}{L^{\alpha}} \overset{\text{def}}{=} A k_p^{\alpha}. \tag{2.3}$$

Hence, the output per capita y depends on the capital-labor ratio k_p measuring the relation between the costs for production factors capital and labor. From Equation (2.1) and (2.2) follows $\dot{K}_p / K_p = s_{k_p}(Y/K_p) - \delta$ and $\dot{L}/L = n$, respectively. Hence, the growth rate of the capital-labor ratio is given by

$$\frac{\dot{k}_p}{k_p} = s_{k_p} \frac{Y}{K_p} - \delta - n$$

and leads to

$$\dot{k}_p = s_{k_p} y - k_p(\delta + n). \tag{2.4}$$

In its steady-state the capital-labor ratio needs to be constant. Inserting (2.3) in (2.4) and simple calculus yields the steady-state value of k_p

$$k_p^* = \left(\frac{s_{k_p} A}{\delta + n} \right)^{\frac{1}{1-\alpha}}. \tag{2.5}$$

Following from equation (2.5) the equilibrium of the capital-labor ratio is positively related to the saving rate and negatively to the growth rate of the working-age population. The steady-state output per capita can be derived by substituting (2.5) into (2.3),

$$y^* = A \left(\frac{s_{k_p} A}{\delta + n} \right)^{\frac{\alpha}{1-\alpha}} = A^{\frac{1}{1-\alpha}} s_{k_p}^{\frac{\alpha}{1-\alpha}} (\delta + n)^{\frac{-\alpha}{1-\alpha}},$$

or, taking logs,

$$\log(y^*) = \frac{\log(A)}{1-\alpha} + \frac{\alpha}{1-\alpha} \log(s_{k_p}) - \frac{\alpha}{1-\alpha} \log(\delta + n). \tag{2.6}$$

Analyzing this model empirically MRW derive a corresponding regression formulation $\log(y) \simeq X'\beta$ (X is the matrix of explanatory variables) assuming

$$E(\log(y_i)|X) = \beta_1 + \beta_2 \log(s_{k_p,i}) + \beta_3 \log(\delta + n_i), \tag{2.7}$$

where $\log(y_i)$ is the observed output in economy i, $\beta_1 = \frac{\log(A)}{1-\alpha}$, $\beta_2 = \frac{\alpha}{1-\alpha}$ and $\beta_3 = -\frac{\alpha}{1-\alpha}$.

Assessing the empirical performance of the model, t-tests are used to check the correct signs of parameters for empirical data. Hence, the first check is whether β_2 is significantly positive and if β_3 is negative. Furthermore, both parameters should have roughly the same magnitude equal to $|\frac{\alpha}{1-\alpha}|$. Using several sets of cross-section data and existing empirical evidence (e.g, Jorgenson et al., 1987) MRW suggest that the capital's share of output α is roughly constant over time and economies at a level of approximately one third. Thus, β_2 and β_3 are expected to be approximately equal to one half in absolute value.

2.1.2 Extended neoclassical growth model

As an extension MRW add human capital to the classical to the Solow model. The use of human capital as a determinant of economic growth is quite common in earlier literature (e.g. Lucas, 1988). Again, a Cobb-Douglas production function is used and generalized to

$$Y = AK_p^{\alpha} K_h^{\nu} L^{1-\alpha-\nu},$$

where K_h denotes the accumulated human capital. In analogy to (2.3) the output per capita is given by

$$y = \frac{Y}{L} = \frac{AK_p^{\alpha} K_h^{\nu} L^{1-\alpha-\nu}}{L^{\alpha} L^{\nu} L^{1-\alpha-\nu}} \stackrel{\text{def}}{=} Ak_p^{\alpha} k_h^{\nu}, \tag{2.8}$$

where k_h denotes the human capital per worker and ν is the share of human capital in production.

In analogy to (2.2), the net investment in human capital results from the difference between gross investment and depreciation

$$\dot{k}_h = s_{k_h} Y - \delta K_h.$$

In the adjusted model two different saving rates, say s_{k_p} and s_{k_h}, represent gross investment for K_p and K_h. Then, in analogy to (2.4), the growth of human capital can be written as

$$\dot{k}_h = s_{k_h} y - (\delta + n) k_h. \tag{2.9}$$

The equilibrium condition in this model is that both, k_p and k_h, are constant, i.e. $\dot{k}_p = 0$ and $\dot{k}_h = 0$. The solution of this system of differential equations for k_p and k_h is given by

$$k_p^* = \left(\frac{A s_{k_p}^{1-\nu} s_{k_h}^{\nu}}{\delta + n} \right)^{\frac{1}{1-\alpha-\nu}} \tag{2.10}$$

$$k_h^* = \left(\frac{A s_{k_p}^{\alpha} s_{k_h}^{1-\alpha}}{\delta + n} \right)^{\frac{1}{1-\alpha-\nu}}. \tag{2.11}$$

Thus, the steady-state output per capita is equal to

$$y^* = A^{\frac{1}{1-\alpha-\nu}} + s_{k_p}^{\frac{\alpha}{1-\alpha-\nu}} + (\delta + n)^{\frac{\alpha+\nu}{1-\alpha-\nu}} + s_{k_h}^{\frac{\nu}{1-\alpha-\nu}}, \tag{2.12}$$

or, taking logs,

$$\begin{aligned} \log(y^*) \quad = \quad & \frac{\log(A)}{1-\alpha-\nu} + \frac{\alpha}{1-\alpha-\nu} \log(s_{k_p}) \\ & - \frac{\alpha+\nu}{1-\alpha-\nu} \log(\delta + n) + \frac{\nu}{1-\alpha-\nu} \log(s_{k_h}). \end{aligned} \tag{2.13}$$

A regression representation of model (2.13) is given by $\log(y_i) \simeq X'\beta$ assuming

$$E(\log(y_i)|X) = \beta_1 + \beta_2 \log(s_{k_p,i}) + \beta_3 \log(\delta + n_i) + \beta_4 \log(s_{k_h,i}). \tag{2.14}$$

According to (2.13), β_2 and β_4 are expected to be positive while β_3 should be negative. Obviously, the magnitude of the parameters depends on the empirical equivalent of the share of human capital ν. MRW assume $\nu = \alpha = 1/3$. Thus, β_2 and β_4 should be approximately equal to one and β_3 should be approximately equal to minus two.

2.1.3 Spatial neoclassical growth model

Again, based on the Cobb-Douglas production function (2.3) Ertur & Koch (2007) propose a spatially augmented model of Solow (1956). The authors assume that the global technology level A is not constant over time and countries and that there are interpedendencies between countries or regions. Due to knowledge spillovers these interdependencies are affected by the spatial factor. Neighboring economies influence each other more than spatially distant economies.

Ertur & Koch (2007) model the interdependencies using so-called Arrow-Romer externalities (see Arrow, 1962; Romer, 1986). The log technology level A is assumed as

$$\log(A) = log(\Omega) + \phi \log(k_p) + \gamma W \log(A). \tag{2.15}$$

A depends on a common technology level Ω with constant exogenous growth which is available for all economies. Furthermore, the technology level rises with the capital per worker (knowledge spillover), while ϕ measures the size of the effect. There is also a spatial effect: A is assumed to depend on a geometrically weighted mean of the neighbors stocks of technology. The weights are given in a nonsingular weighting matrix W and γ indicates the rate of dependence from worldwide level of technology.

Solving Equation (2.15) yields

$$\log(A) = (I - \gamma W)^{-1} \log(\Omega) + \phi (I - \gamma W)^{-1} \log(k_p). \tag{2.16}$$

For developing the steady state per capita income we build logs of (2.3)

$$\log(y) = \log(A) + \alpha \log(k_p) \tag{2.17}$$

and replace $\log(A)$ by Equation (2.16) such that

$$\log(y) = (I - \gamma W)^{-1} \log(\Omega) + \phi (I - \gamma W)^{-1} \log(k_p) + \alpha \log(k_p) \tag{2.18}$$

$$\log(y) = \log(\Omega) + (\alpha + \phi) k_p - \gamma \alpha W \log(k_p) + \gamma W \log(y). \tag{2.19}$$

Analogously to the neoclassical growth model it can be shown that the steady-state per capita income is given by

$$\begin{aligned}\log(y^*) = \frac{1}{1-\alpha-\phi}\log(\Omega) + \frac{\alpha+\phi}{1-\alpha-\phi}\log(s_{k_p}) - \frac{\alpha+\phi}{1-\alpha-\phi}\log(\delta+n) \\ -\frac{\alpha\gamma}{1-\alpha-\phi}\sum_{j\neq i} w_{i,j}\log(s_{k_p,j}) + \frac{\alpha\gamma}{1-\alpha-\phi}\sum_{j\neq i} w_{i,j}\log(\delta+n_j) + \\ \frac{\gamma(1-\alpha)}{1-\alpha-\phi}\sum_{j\neq i} w_{i,j}\log(y_j)), \quad (2.20)\end{aligned}$$

where $w_{i,j}$ is the element of the weighting matrix W giving the influence of neighbor j on economy i.

Empirically this steady-state is estimated with a regression model assuming

$$\begin{aligned}E(\log(y_i)|X) &= \beta_0 + \beta_1 \log(s_{k_p,i}) - \beta_2 \log(\delta+n_i) - \theta_1 \sum_{j\neq i} w_{i,j}\log(s_{k_p,j}) + \\ &\quad \theta_2 \sum_{j\neq i} w_{i,j}\log(\delta+n_j) + \rho\sum_{j\neq i} w_{i,j}\log(y_j)). \quad (2.21)\end{aligned}$$

The result is a spatial version of the neoclassical growth model with spatial lags of endogenous and exogenous variables. Attention should be paid to the fact that in this case X includes an endogenous variable.

2.2 Convergence concepts

Current literature focuses on the analysis of convergence models instead of growth models (e.g. Barro & Sala-i Martin, 1992; Barro et al., 1991; Sala-i Martin, 1996a,b). In this context convergence is assumed if differences in standards of living between economies become smaller or even disappear over time.

Classical literature offers different concepts of convergence. The most popular one is β-convergence meaning that poor economies growth faster than rich ones (see Sala-i Martin, 1996a,b). This concept is outlined in Subsection 2.2.1. Sala-i Martin (1996a) proposes another related concept, σ-convergence, which is described in

Subsection 2.2.2. The idea of σ-convergence is a decreasing dispersion of cross-sectional per capita income over time. Furthermore, Subsection 2.2.3 describes the concept of convergence in relative per capita income (e.g., Quah, 1993a,b, 1997) assuming an equally fraction of mean per capita income for all economies over time. Subsection 2.2.4 shows the relationship between these three concepts. The nonparametric convergence concepts proposed by Maasoumi et al. (2007) are described in Subsection 2.2.5. Applying these concepts the sample of countries is divided into different groups whose complete distribution of growth rates is estimated nonparametrically while convergence is assumed if the growth rate distribution of one group stochastically dominates the one of another group.

Subsection 2.2.6 compares the different grades of information utilization of the different concepts. Finally, 2.2.7 deals with the question, whether there is always "global convergence" or whether convergence clubs occur.

2.2.1 β-convergence

β-convergence is the most popular concept for analyzing convergence (e.g. Barro & Sala-i Martin, 1992). A main advantage of the concept is that it is founded in the theory of the Solow model and thus a theory-based interpretation of the results is possible. β-convergence means that poor economies grow faster than rich ones implying that the corresponding income gap diminishes. Therefore the linear regression model

$$\upsilon_i = \beta_0 - \beta_1 \log(y_{i,0}) + \varepsilon_i \tag{2.22}$$

is used, where $\upsilon_i = \log(y_{i,T}) - \log(y_{i,0})$ is the growth rate of per capita income in country i, T is the final period, and $\log(y_{i,0})$ is the initial income of economy i. Equation (2.22) measures so-called "absolute convergence" if β_1 is significantly greater than 0. It is called absolute convergence because only the initial incomes are used to explain growth rates. Hence, the meaning of β_1 is that a marginal reduction of initial income yields a higher growth rate. In the case of absolute β-convergence all economies converge to the same steady-state.

Analyzing the shape of β-convergence for different values of β_1 we rewrite Equation (2.22) to

$$\log(y_{i,T}) = \beta_0 + (1-\beta_1)\log(y_{i,0}) + \varepsilon_i. \tag{2.23}$$

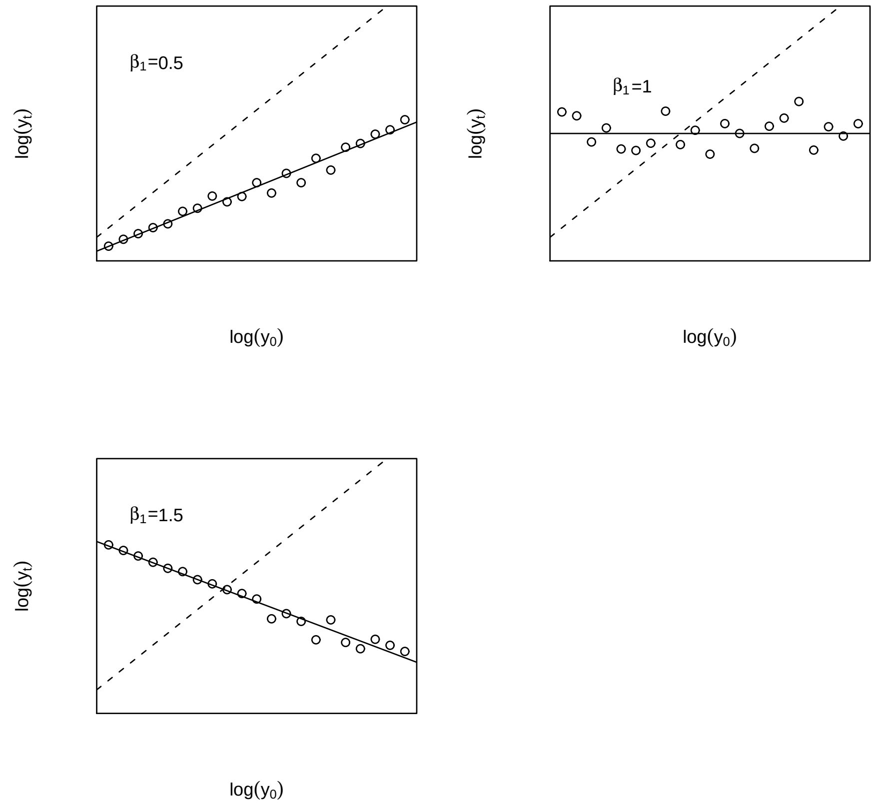


Figure 2.1: Different shapes of β-convergence depending on value of β_1. Dashed lines display the extreme case of $\beta_1 = 0$.

We differentiate between the three cases $0 < \beta_1 < 1$, $\beta_1 = 1$ and $\beta_1 > 1$. For all of them β-convergence is assumed because $\beta_1 > 0$. The first case ($\beta_1 = 0.5$) is displayed in the upper left panel in Figure 2.1. In this case the incomes of poor economies are catching up to rich ones, but in T the steady-state income has not been reached. For $\beta_1 = 1$ convergence is reached. Independently from the income in period 0, the income in period T is close to the steady-state income on cross-economy average. For $\beta_1 > 1$ the economies which are poor in 0 are rich in T

and vice versa. Sala-i Martin (1996b) rules the cases two and three out calling them “leapfrogging” and “overshooting” because poor economies systematically get ahead of rich ones which is not feasible. The author assumes $0 < \beta_1 < 1$ and excludes the other cases.

The assumption of a common steady-state implies that the economies differ only in their initial income is often implausible. The Solow model predicts different steady-states for economies depending on the values of the covariates population growth, physical capital and human capital. Thus, we can focus on another concept of convergence, namely conditional β-convergence. This concept means that poor economies return faster to their individual steady-states than rich economies. The individual steady-states depend on country specific endowments measured by different covariates, basically physical capital and human capital.

The corresponding model can be derived from the growth rates of physical (2.4) and human capital (2.9). From the production function follows that the growth rate of the output per capita y is the weighted average of the growth rates of the inputs $\dot{y}/y = \alpha(\dot{k}_p/k_p) + \nu(\dot{k}_h/k_h)$. We can rewrite (2.4) and (2.9) depending on $\log(k_p)$ and $\log(k_h)$ respectively such that

$$\dot{k}_p/k_p = s_{k_p} A e^{-(1-\alpha)\log(k_p)} e^{\nu \log(k_h)} - (\delta + n) \tag{2.24}$$

$$\dot{k}_h/k_h = s_{k_h} A e^{\alpha \log(k_p)} e^{-(1-\nu)\log(k_h)} - (\delta + n). \tag{2.25}$$

Taking a two-dimensional first-order Taylor approximation from these equations leads to

$$\begin{aligned}\dot{y}/y = {} & [\alpha s_{k_p} A(-(1-\alpha)) e^{-(1-\alpha)\log(k_p^*)} e^{\nu \log(k_h^*)} + \nu s_{k_h} A \alpha e^{\log(k_p^*)} e^{-(1-\nu)\log(k_h^*)}] \\ & \cdot [\log(k_p) - \log(k_p^*)] + [\alpha s_{k_p} A e^{-(1-\alpha)\log(k_p^*)} \nu e^{\nu \log(k_h^*)} \\ & + \nu s_{k_h} A e^{\alpha \log(k_p^*)} (-(1-\nu)) e^{-(1-\nu)\log(k_h^*)}][\log(k_h) - \log(k_h^*)]. \end{aligned} \tag{2.26}$$

By using the steady-state conditions for k_p and k_h (2.10) and (2.11) we can simplify

(2.26) to

$$\dot{y}/y = -b[\log(y) - \log(y^*)], \tag{2.27}$$

where $b = (1-\alpha-\nu)(\delta+n)$ is the convergence coefficient, which measures the speed of convergence (see Barro & Sala-i Martin, 2004).

A useful interpretation of b can be derived from the solution of the differential equation (2.27) which is given by

$$\log(y(T)) = (1-e^{-bT})\log(y^*) + e^{-bT}\log(y(0)), \tag{2.28}$$

where $y(0)$ is the initial output per capita. The half-life (the time where $\log(y(T))$ is halfway between the initial level and the steady-state) is given by the condition $e^{-bT} = 1/2$, which is equivalent to $t = \log(2)/b$. For example, if $b = 0.02$ the economy requires about 35 years to move halfway to its steady-state.

In the next step we derive the new model for analyzing convergence based on the previous approach. For that we subtract $\log(y(0))$ from both sides of (2.28) and replace $\log(y*)$ by (2.13). Thus we obtain the model

$$\begin{aligned}\log(y(T)) - \log(y(0)) = \; & - \; (1-e^{-bT})\log(y(0)) \\ & + \; (1-e^{-bT})\frac{\alpha}{b}\log(s_{k_p}) \\ & + \; (1-e^{-bT})\frac{\nu}{b}\log(s_{k_h}) \\ & - \; (1-e^{-bT})\frac{\alpha+\nu}{b}\log(\delta+n).\end{aligned} \tag{2.29}$$

Equation (2.29) can be transformed into the general regression model assuming

$$\begin{aligned}E(\upsilon_i|X) \;\; = \;\; & \beta_0 - \beta_1\log(y_{i,0}) + \beta_2\log(s_{k_p,i}) + \beta_3\log(s_{k_h,i}) \\ & + \beta_4\log(\delta+n_i),\end{aligned} \tag{2.30}$$

where

$$\beta_1 = 1 - e^{-bT}. \tag{2.31}$$

Conditional β-convergence can be assumed if b is greater than 0. For assessing b Equation (2.31) is solved for the parameter. A detailed discussion of situations where the different concepts should be used is given in Chapter 4.

The spatial augmented convergence model can be derived in the same way while the single steps are more complicated due to the large number of covariates and the problem of an endogenous explanatory variable. The details are left out here and can be found in Ertur & Koch (2007). The resulting model is

$$\begin{aligned}
\log(y(T)) - \log(y(0)) \quad = \quad & -(1-e^{-d_iT})\log(y(0)) \\
& +(1-e^{-d_iT})\frac{\alpha+\phi}{1-\alpha-\phi}\log(s_{k_p}) \\
& -(1-e^{-d_iT})\frac{\alpha+\phi}{1-\alpha-\phi}\log(\delta+n) \\
& +(1-e^{-d_iT})\frac{\gamma(1-\alpha)}{1-\alpha-\phi}\sum_{j\neq i} w_{i,j}\log(y_j(0)) \\
& -(1-e^{-d_iT})\frac{\alpha\gamma}{1-\alpha-\phi}\sum_{j\neq i} w_{i,j}\log(s_{k_p,j}) \\
& +(1-e^{-d_iT})\frac{\alpha\gamma}{1-\alpha-\phi}\sum_{j\neq i} w_{i,j}\log(\delta+n_j) \\
& +(1-e^{-d_iT})\frac{\gamma(1-\alpha)}{1-\alpha-\phi} \\
& \cdot\sum_{j\neq i}\frac{1}{(1-e^{-d_iT})}w_{i,j}[\log(y(T))-\log(y(0))],
\end{aligned}$$

where d_i is an individual convergence parameter depending on the individual level of technology.

2.2.2 σ-convergence

Another important convergence concept is σ-convergence. The idea of this concept is that convergence is assumed if the dispersion measured by the standard deviation σ of cross-economy per capita income declines over time so that

$$\sigma_{t+\tau} < \sigma_t, \tag{2.32}$$

where $t = 1, \ldots, T$. The concept of σ-convergence does not deal with the question to which steady-state the incomes convergence and thus, whether the average income rise or fall over time. For analyzing σ-convergence the only important question is whether the cross-economy variance of $\log(y_{i,t})$ decreases over time and thus if the incomes at large come closer.

The concepts of σ- and β-convergence are related (see Sala-i Martin, 1996b). Based on Equation (2.23) the error terms $u_{i,t}$ are assumed to have a zero mean and a constant variance σ_u^2 for all time periods and economies. Furthermore, the sample variance of $\log(y_{i,t})$ is given by

$$\sigma_t^2 = N^{-1} \sum_{i=1}^{N} [\log(y_{i,t}) - \overline{\log(y_{i,t})}]^2, \tag{2.33}$$

where the $\overline{\log(y_{i,t})}$ is the mean of $\log(y_{i,t})$ in t. Calculating the variance for Equation (2.23)[4] using the assumptions above yields

$$\begin{aligned} Var(\log(y_{i,t})) &= Var(\beta_0) + Var((1-\beta_1)\log(y_{i,t-1})) + Var(\varepsilon_{i,t}) \quad (2.34) \\ \sigma_t^2 &= (1-\beta_1)^2 \sigma_{t-1}^2 + \sigma_\varepsilon^2. \quad (2.35) \end{aligned}$$

Resulting from assuming that σ_u^2 is constant, σ_t^2 decreases if $\beta_1 > 0$ and thus, σ-convergence cannot occur without β-convergence. β-convergence is a necessary condition for σ-convergence. The dispersion of cross-sectional per capita income may only reduce if poor economies grow faster than rich ones. However, β-convergence is not a sufficient condition for the existence of σ-convergence. The steady-state of the linear difference equation given in Equation (2.35) is given by its inhomogeneous solution

$$\sigma_*^2 = \frac{\sigma_u^2}{1-(1-\beta_1)}. \tag{2.36}$$

Developing the first-order Taylor approximation gives the solution of the equation depending on the steady-state variance

$$\sigma_t^2 = \sigma_*^2 + (1-\beta_1)^2 [\sigma_{t-1}^2 - \sigma_*^2]. \tag{2.37}$$

Equation (2.37) suggests that β-convergence is only a necessary but not a sufficient condition for σ-convergence. If β-convergence is present, σ_t^2 can increase or decrease. The direction of change depends on whether σ_t^2 is below or above the steady-state.

[4] Here the initial period 0 is set to $t-1$.

2.2.3 Convergence in relative per capita income

Current literature deals with many more convergence concepts where only the most popular ones are presented in this thesis. In this chapter the concept of convergence in relative per capita income is described (see Quah, 1993a,b, 1997) which is used below. Quah (1993a,b, 1997) initiates the criticism on only investigating β-convergence on average. The author argues that time-average growth rates are not appropriate for deriving time dynamical implications. The problem in this context is the comparison of realizations of the same random variable at different points in time. A linear regression of those variables is clouded by Galton's Fallacy of regression towards the mean which means that economies lying above cross-economy average generally lie below the average in the second period and vice versa because the realizations above the average are only randomly higher than the average. The mean of those higher individuals will be considerably smaller in the later period, because the high values in the first period are only caused by random effects. This is the reason why a coefficient of a regression of $y_{2,i}$ on $y_{1,i}$ always tends to be negative and hence cannot imply anything useful about an assimilation of incomes over time.

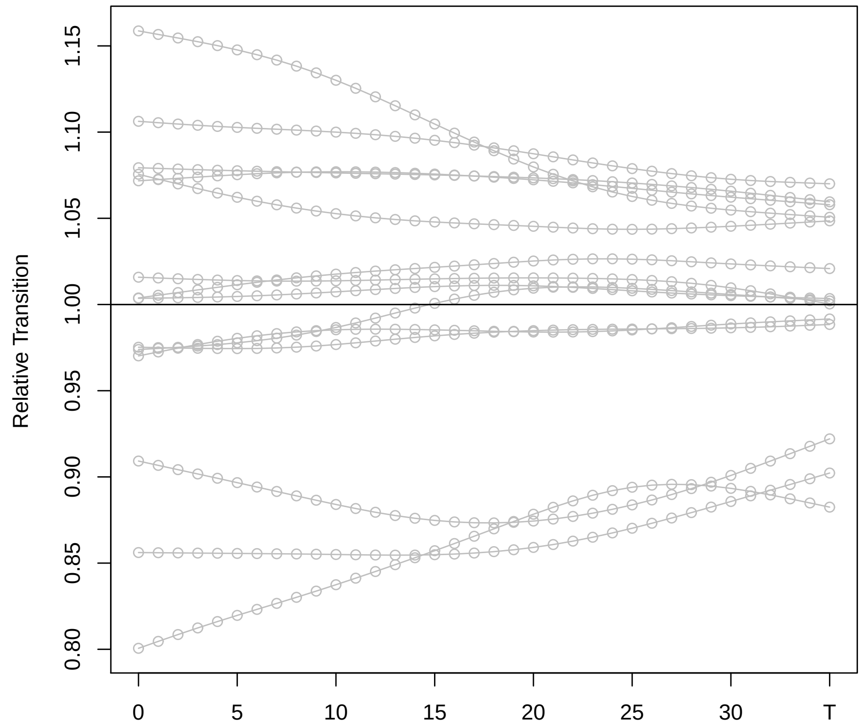
Alternatively, Quah (1993a,b, 1997) presents a new convergence concept. He normalizes the income by dividing the output per worker of every economy by a weighted cross-economy average for every year t (high weights are used for countries with large populations or simply the mean)

$$h_{i,t} = \frac{\log(y_{i,t})}{\sum_{i=1}^{N} w_i \log(y_{i,t})}, \tag{2.38}$$

where $\sum_{i=1}^{n} w_i = 1$. The interpretation of the ratio is that the output per worker of economy i is $h_{i,t}$-times as big as the weighted cross-economy average. Convergence over time is assumed if the $h_{i,t} \to 1$ for $t \to \infty$.

Figure 2.2 offers a graphical example for convergence in relative per capita income.

Figure 2.2: Example for convergence in relative per capita income



The relative transition coefficients $h_{i,t}$ for several years and economies are plotted while the coefficients from the individual economies are connected to the relative transition paths of the economies. The black line is equal to one. This group of economies converges because all transition paths are closer to one in final period T than in initial period 0.

2.2.4 The relationship between β-convergence, σ-convergence and convergence in relative per capita income

For comparing the different convergence concepts only the initial and final period are taken into account because β-convergence does not consider intermediate periods. Convergence in relative per capita income investigates the behavior of the relative transition coefficients $h_{i,t} = \log(y_{i,t})/(N^{-1}\sum_{i=1}^{N}\log(y_{i,t}))$ (Note that for simplification the unweighted average is used). Convergence is assumed if $h_{i,t} \to 1$ for all i if $t \to \infty$.

If convergence in the latter sense is fulfilled over the period from $[0,T]$ all economies have approximately the same per capita income in T, which is on the level of the mean income in T independently from their initial income in 0. Figure 2.3 shows the relationship between initial and final per capita income under the assumption of convergence in relative per capita income. Independently from the initial income, the income in the final period is approximately equal to the mean income for all economies (in this case all points lie on the regression line).

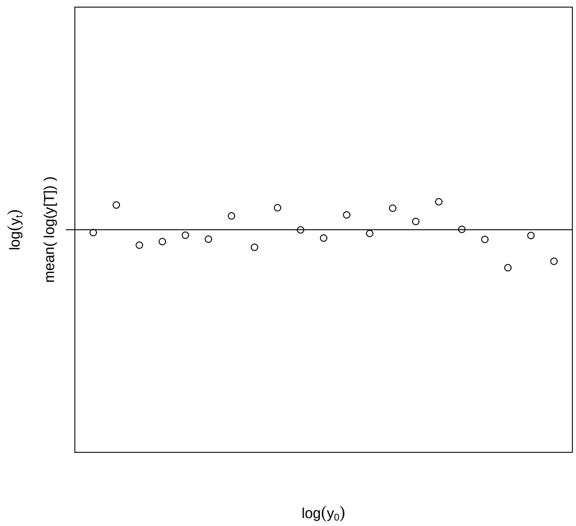


Figure 2.3: Initial and final per capita income under the assumption of convergence in relative per capita income

In empirical samples the points do not lie exactly on a line but they spread sparsely around the mean. Specifying the relationship in a linear regression model yields

$$\log(y_{i,T}) = \beta_0 + b\log(y_{i,0}) + \varepsilon_i, \tag{2.39}$$

where $b = 0$ under convergence in per capita hypothesis.

At this point the relationship between β-convergence and the convergence in relative per capita income becomes clear. Equation (2.39) can be easily transformed into the β-convergence model.

$$\log(y_{i,T}) = \beta_0 + (1 - \beta_1)\log(y_{i,0}) + \varepsilon_i, \tag{2.40}$$

Thus, convergence in the sense of Quah (1993a,b, 1997) is a special case of β-convergence where $\beta_1 = 1$. For the existence of convergence in this sense another assumption must also be fulfilled. It is not sufficient that $b = 0$ in Equation (2.39). Additionally the dispersion has to be small and unsystematic in t or at least the dispersion in T must be smaller than in 0. This is fulfilled if σ-convergence is present. Thus, β- and σ-convergence are necessary but not sufficient conditions for convergence in relative per capita income.

2.2.5 Nonparametric convergence concepts

Maasoumi et al. (2007) present two new convergence concepts using nonparametric approaches. The main idea of both concepts is to analyze convergence between a priori defined groups of economies instead of generally between poor and rich economies. On the one hand this is a very strong and restrictive assumption, because the a priori classification of the convergence direction can hide "real" directions of convergence. On the other hand the concept allows to follow specific questions about several groups. For example, German regional data are divided into east and west to check if eastern districts close the gap on the western districts.

Applying the first concept, a nonparametric regression of growth rates of per capita income on the most popular conditioning variables (population growth, human capital and investment rate) is conducted. Second, the conditional probability density function (PDF) and the conditional cumulative distribution function (CDF) of the nonparametric growth rates depending on groups and time periods are estimated.

Final stage is a check for stochastic dominance of CDFs for different groups and for different periods. The distribution of the random variable X_1 first order stochastically dominates the distribution of X_2 if the conditional cumulative distribution function of X_1 lies under the conditional CDF of X_2 at every point x ($F_1(x) \leq F_2(x)$) (see Hadar & Russell, 1969).

Two questions can be analyzed with this concept. Mainly, it can checked if per capita income from eastern and western districts converge by comparing the conditional growth rate CDFs of both groups at the same time. In this context per capita income convergence between groups is assumed if the growth rate distribution of the initially poorer group stochastically dominates the one of the other group. For the present example this means, that all parts and hence all quantiles of the distribution of east districts grow faster than districts in the west. Thus, the poorest 10% of the eastern districts grow faster than the poorest 10% of the western districts and so on. If the growth rates of the districts in the east are higher than the ones of the districts in the west at all quantiles of the growth rate distribution, the per capita income in all districts of the former German Democratic Republic (GDR) growths faster and thus the gap between income in both former parts of Germany dissolves.

Furthermore, the difference between absolute and conditional convergence can be considered. For analyzing absolute convergence, the fitted values of the nonparametric regression are used as growth rates of per capita income. Conditional convergence can be analyzed if the residuals of the nonparametric regression are used for growth rates. The residuals stand for conditional convergence because they are the growth rates after controlling for the influences of the conditioning variables. In addition to analyzing convergence between groups, it is also possible to analyze convergence within groups by regarding the conditional CDFs of one group for different periods. How the nonparametric regression works can be seen in Subsection 3.2.1.

Stochastic dominance can be analyzed graphically, though the results of a formal test are presented by Linton et al. (2005). Using an extended Kolmogorov-Smirnov test for stochastic dominance of McFadden (1989) the authors propose a consistent procedure for estimating critical values. The test idea is to build pairwise differences between the conditional CDF of X_1 and X_2 for every x and to observe the corresponding maximum. Then, differences between the conditional CDFs of X_2 and X_1 are calculated too. From both maxima the minimum is taken

$$d^* = \min_{k \neq l} \sup_{x \in X} [F_k(x) - F_l(x)].$$

The hypotheses of the test are

$$H_0 : d^* \leq 0 \text{ vs. } H_1 : d^* > 0.$$

The null hypothesis means a negative minimal difference between the CDFs which is equivalent to the fact that F_k lies under F_l and thus F_l first order dominates F_k. So, if the null is not rejected, stochastic dominance between the two distributions can be assumed. The test statistic is the empirical analogue of d^*

$$D_N = \min_{k \neq l} \sup_{x \in X} \sqrt{N}[F_{kN}(x) - F_{lN}(x)],$$

where F_{kN} denotes the number of observations of X_k, which are smaller than x, divided by N. The distribution of D_N is obtained by a subsample bootstrap. Therefore, $N-b+1$ subsamples of size b and the test statistic are computed. The null is rejected if D_N is greater than the $(1-\alpha)$-quantile of the resampled distribution.

Taking a step forward the second convergence concept of Maasoumi et al. (2007) uses entropy measures to capture the exact distances between distributions of several groups for different times t. Thus, for every period the difference between the conditional CDFs of the growth rates of eastern and western districts are calculated. Convergence is assumed if the distance became smaller or even dissolves over time. For measuring the distance the authors use an entropy which is additionally a metric proposed by Granger et al. (2004)

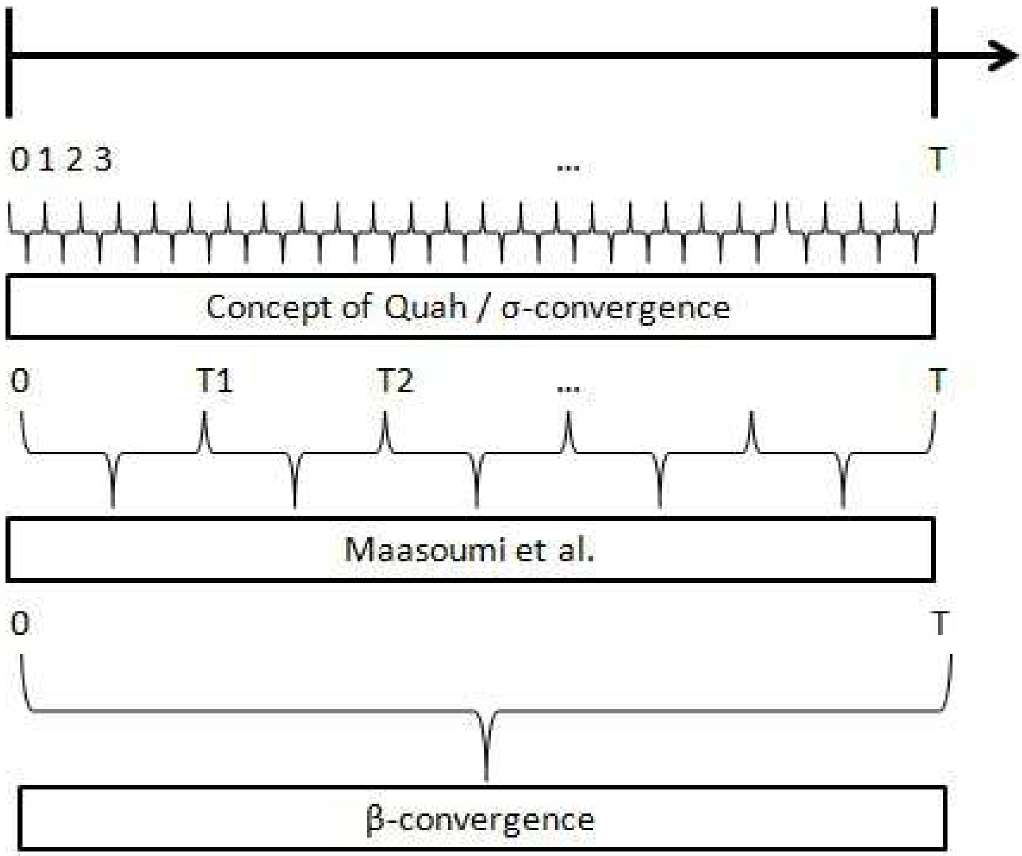
$$S_\rho = \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(f_1^{\frac{1}{2}} - f_2^{\frac{1}{2}} \right)^2 dxdy,$$

where f_1 and f_2 are the marginal densities of the growth rates for east and west. S_ρ lies between 0 and 1, while 0 means, that there is no distance between both distributions. Thus, it can be tested if the distributions are equal by testing the null hypothesis $S_\rho = 0$.

2.2.6 Utilization of available information

As described in the previous sections the different convergence concepts make differing uses of levels of available data. Illustrating this fact this Subsection contrasts the different concepts with respect to this issue. For all concepts there are available information for $t = 0, \dots, T$, but not all concepts use all information. In Figure 2.4 the considered information of the different concepts is displayed. There are

Figure 2.4: Utilization of information of different convergence concepts



three utilization-levels of available information. The concept convergence in relative per capita income and σ-convergence are arranged on the first level. Both concepts use the information for all available periods $t = 1, \dots, T$. The concepts of

Maasoumi et al. (2007) are arranged on the second level. They pool the data and divide the time horizon into several subperiods and use only information for initial and final period and the borders of the subperiods ($t \in \{0, T1, T2, ..., T\}$). Thus, these concepts ignore the information for intermediate periods. β-convergence considers only small parts of information. For estimating the concept only information of initial and final period are used $t \in \{0, T\}$. This is a clear handicap of β-convergence, because ignoring of available information may cause incorrect results and wrong conclusions. Generally, available information should be fully considered for achieving the best model performance.

2.2.7 Convergence clubs

There are many works suggesting that especially for cross-country data there is no global convergence for all countries independently from the underlying convergence concept. Thus, several authors try to identify groups of economies with common convergence behavior, the so-called convergence clubs.

The seminal paper dealing with the question whether convergence is really global is Baumol (1986). The author asks: “Does convergence [...] extend beyond the [...] countries? Or is the convergence club a very exclusive organization?” (Baumol, 1986, p. 1079). Analyzing the relationship between the Gross Domestic Product per capita in 1950 and its growth rate between 1950 and 1980, the author finds visually two convergence clubs. The first club consists of the sixteen industrialized countries in the sample and the second one contains of the centrally planned economies. As a formal validation of his assumption he finds falling Gini coefficients for the decades 1950 to 1980 inside the two clubs. A falling Gini coefficient indicates that the income distribution inside a club becomes more equal.

The work of Baumol (1986) gives rise to a circumstantial discussion about convergence clubs. The main papers in this context which are used below are summarized in this subsection.

Using panel data for 118 countries and from 1962 to 1985 Quah (1993a) finds a trend to extremes as the upper part of the income distribution converges to the richest countries, while the poor countries become poorer. The resulting bi-modality gives rise to the concept of "convergence clubs".

Durlauf & Johnson (1995) find misspecification of the linear model on the MRW data. However, using regression tree analysis the authors identify clubs for which the linear model and thus β-convergence is fulfilled on a club-level.

Seitz (1995) analyzes β-convergence in West German regions for data of the districts and district-free cities from 1980 till 1990. The author does not find "global" convergence, but he discovers convergence clubs based on the grade of urbanization. For that the districts are divided into three categories: district free cities, districts in direct neighborhood to such cities and the other districts far away from a city.

Hobijn & Franses (2000) use a consistent clustering algorithm allowing for endogenous cluster selection. For several datasets (e.g. Penn World Table) they find many clusters, but their sizes are frequently very small.

Phillips & Sul (2003, 2007a,b, 2009) develop a data-based clubbing algorithm and find convergence clubs for several sets of data.

Based on the predictive density (related with Quah, 1993a,b, 1997) Canova (2004) presents an algorithm determining the number of clubs, their location and break points between clubs. Using data on 144 European NUTS2[5] units, the authors apply the algorithm and estimate a single convergence coefficient for every club.

Juárez & Steel (2010) use an autoregressive and model-based algorithm for panel data. The advantage of the method is that it offers membership probabilities for every economy and cluster. They analyze data on 258 European NUTS2 regions from 1995 to 2004 and 738 manufacturing firms from Spain and find clusters for both data sets.

[5]NUTS (Nomenclature des unités territoriales statistiques) is a hierarchical geographical classification of European official statistics. Level 2 are units of intermediate size.

Chapter 3

Remedies for problems of classical convergence concepts

The work of MRW gives rise to an intensive and on-going discussion on the quality and validation of classical convergence models. The aim of this chapter is to offer a survey of current literature for detecting problems of classical growth and convergence modeling and potential remedies. Therefore, the problems are structured in five strands presented in the next sections. Every section starts with a survey of current literature dealing with the appropriate problem of classical modeling. The corresponding subsections describe selected economic theories and econometric methods offering improvements in details. The presented methods are the basis for the empirical analysis in Chapter 5.

Section 3.1 covers papers dealing with the question of variable definition and selection, more precisely the papers analyze which determinants really influence growth and convergence. Furthermore, in this section different definitions of the growth determinants are presented. Section 3.2 summarizes papers identifying neglected nonlinearities as a source for β-convergence being invalid for several data sets. A nonparametric alternative is discussed and a test of Hsiao et al. (2007) for parametric misspecification is described in this section. Section 3.3 deals with the problem of omitted heterogeneity and two different methods for capturing heterogeneity are presented, on the one hand quantile regression (see Koenker & Bassett, 1978) for considering location scale effects of the conditional growth distribution and on the other hand the log t regression of Phillips & Sul (2003, 2007a,b, 2009). Section 3.4 describes two methods combining the problems of neglected heterogeneity and nonlinearities. First, nonparametric quantile regression (see Li & Racine, 2008) is described which captures nonlinear location scale effects and second, a two step procedure of Haupt & Meier (2011) combining the log t regression Phillips & Sul (2003, 2007a,b, 2009) with nonparametric methods is presented. Finally, Section 3.5 summarizes papers dealing with the problem of spatial association. In this section spatial convergence models, the influence of neighboring economies, and tests for spatial dependence are presented.

3.1 Growth determinants

A first big wave of literature discusses the definition of variables in empirical analysis and the influence of potential additional covariates to augment classical growth models and conditional β-convergence.

The Solow model extended by human capital predicts a linear model where the standard of living is explained by the saving rate, the growth of working age population and human capital. All these variables can be defined in several ways and their definitions may have a noticeable influence on the estimation results.

As a measure for standards of living MRW use per capita income measured by the real GDP divided by the working-age population (15-64 years old). This is common in literature.

The first explanatory variable is the saving rate s_k. MRW proxy this variable by the GDP share of investment (including government investment) divided by 100. Taking investment rates as a measure for savings rates is not obvious at the first sight. Investment rates are the share of output which is used to replace or enlarge the stock of physical capital. Analyzing regional data several authors use other definitions of this variable. E.g. Seitz (1995) uses the absolute investment in the manufacturing industry. Kosfeld et al. (2006) use the number of newly established enterprises relative to the working population as an indicator for the investment rate.

MRW proxy the growth rate of the working-age population n by the growth rate of total population because there is no reliable data for several developing countries. The depreciation rate δ is assumed to be constant over time and countries at a value of $\delta = 0.05$. Taking population growth as a proxy is common in literature independently from the level of aggregation. The size of the depreciation rate is disputable. Especially for regional data other depreciation rates might be assumed than for cross-country data (e.g. Kosfeld et al., 2006).

The last and most complicated variable is human capital s_h. MRW use the share of working-age population, which attends the secondary school. This variable is compounded by the fraction of eligible population (12-17) visiting secondary school multiplied with the fraction of people from the working-age population in school age (15-17). Other authors (e.g. Kalaitzidakis et al., 2001) take the mean years of schooling. Analyzing German regional data these conceptions of human capital are not useful because these variables are similar for all regions in disaggregated data. Alternatively, many authors use the share of workers with academic degree (e.g. Kosfeld et al., 2006; Seitz, 1995) or the proportion of persons with “Abitur” (e.g. Herz & Röger, 1995). Funke & Niebuhr (2005) consider the density of employment in Research and Development for this covariate.

Some authors focus on the influence of human capital on output growth. For example Benhabib & Spiegel (1994) find an insignificant or even negative relationship between human capital and growth rates of per capita income by using the mean years of schooling as a proxy for human capital, while MRW find a significant positive correlation when human capital is represented by enrollment rates. Thus, the way how human capital is approximated empirically seems to play an important role.

A huge part of early convergence analysis is devoted to significance tests as there are many co-existing strands of reasoning using different but not necessarily mutually disjunct sets of explanatory variables and different data sets on different levels of aggregation (see Evans, 1998; Seitz, 1995; Herz & Röger, 1995; Islam, 1995; Lee et al., 1997 for recent contributions).

Barro (1991) is the essential reference which tries to find empirical evidence for the influence of other variables, for example geographical position, government expenditure, political stability, economic system and market distortions. Afterwards, a lot of articles introduce more and more new explanatory variables which may be related with output growth. The problem in this context seems to be that every researcher considers a certain set of variables which are only significant in the corresponding constellation. Sala-i Martin (1997) and Durlauf & Quah (1999) finally try to identify the variables which "really" influence economic growth. They do so by conducting regressions with numerous combinations of potential variables and noting which of them are significant frequently.

3.2 Nonlinearities

Current literature leaves the platform of simple replications of MRW with new data sets and variables and instead criticizes the classical convergence model for several reasons by proposing basic extensions in an economic and econometric sense. First, one main strand of this literature criticizes the lack of flexibility of MRW's (least squares) estimation of a linear regression model and starts a discussion about alternative functional forms. Second, the concept of β-convergence is criticized as it only covers one aspect - the mean effect - of initial income on the distribution of income growth rates.

Addressing the former issue "Kalaitzidakis et al. (2001) use semi-parametric estimation techniques [to the extended specification] and find a nonlinear effect of human capital measured by the mean years of schooling on economic growth. When using the enrollment rates to describe human capital, its effect is linear. Earlier results by Liu & Stengos (1999), based on ideas of Durlauf & Johnson (1995), confirm these findings, [though the authors do not back up their results with formal tests]”[6].

The problem of potential nonlinearities in growth (and convergence) regressions has also been recently addressed by applying fully nonparametric methods for regression and specification testing (see Haupt & Petring, 2011). The authors use a local linear kernel estimator with a generalized product kernel function proposed by Racine & Li (2004) and Li & Racine (2004). Using data from Penn World Tables also used by Mankiw et al. (1992) they find considerable evidence for parametric misspecification and a superior performance of a nonparametric model.

Quah (1993a,b, 1997) initiates the criticism of the concept of only investigating β-convergence on average (see Subsection 2.2.3). He explores distribution dynamics and heterogeneity by comparing the per capita income distributions over time and by estimating transition matrices. Comparing those matrices for several subperiods provides an informal basis to detect convergence or divergence in different parts of the income distribution. Using panel data for 118 countries from 1962 to 1985 he finds a trend to extremes as the upper part of the distribution converges to the most rich countries, while poor countries become poorer. The resulting bi-modality gives rise to the concept of “convergence clubs” (see Subsection 2.2.7).

Bringing both lines together Maasoumi et al. (2007) introduce a novel nonparametric concept for convergence estimation and testing (see Subsection 2.2.5). In their application to cross-country panel data over five periods they compare the distribution of growth rates for OECD and Non-OECD countries. Using stochastic dominance (SD) rankings, the idea is to assume convergence if one growth distribution dominates the other stochastically. The authors find clear evidence for both nonlinearities and convergence clubs. Furthermore, they use entropy measures to assess the numerical value of distance between the distributions for several periods. The idea of convergence in this context is that the distance between the distributions

[6] cited from Haupt & Petring (2011).

might shrink over time. In this vein Henderson (2010) applies nonparametric kernel estimation to the data of Maasoumi et al. and also finds convergence clubs by analyzing the estimated distribution of the partial effects of initial income and applying a test for multi-modality.

Funke & Niebuhr (2005) use nonparametric kernel estimates to analyze the income distribution over time in West German planning units between 1976 and 1996. They find a bimodal distribution and following this awareness they test for multiple equilibria using threshold estimation. The result are three similar groups of regions being a hint for convergence clubs. Juessen (2009) analyzes the so-called "distribution dynamics" for 271 labor market regions between 1992 and 2004 in the manner of Quah (1993a,b, 1997). Investigating the distributions of relative GDP per worker for different years with nonparametric methods shows proceeding convergence between East and West. Using a test for multimodality yields a bimodal distribution for 1992 and therefore significant differences between East and West. Analyzing the distribution of 2004 offers no longer substantial differences between both German parts.

In the following Subsections the focus is on the first issue, namely nonlinearities in growth regressions. In Subsection 3.2.1 the already mentioned concept of nonparametric kernel density estimation is described and Subsection 3.2.2 deals with testing for parametric misspecification.

3.2.1 Nonparametric kernel density estimation

A nonparametric alternative to classical linear modeling is a local linear kernel estimation with regression function

$$E[y|X] = g(X) + E[u|X],$$

where it is assumed that E[u|X]=0. In analogy to parametric modeling, the regression function estimates the conditional mean of the response variable depending on covariates. However, the specific form of the function is not restricted but it is a general function $g()$ allowing for all forms of interactions between covariates.

The idea of a local linear kernel estimation is that a model for every observation in its direct neighborhood with respect to the covariates x is estimated. The size of the neighborhood is given by the bandwidth λ. In addition, the points in the neighborhood are weighted differently, vary the kernel function $K()$.

Racine & Li (2004) consider the specific situation where both continuous and categorical data are used. They use a generalized product kernel for K. A mixed covariate vector with continuous and categorical variables is assumed while the categorical variables are divided in ordered and unordered variables. The idea of Racine & Li (2004) is that all types of variables with respect to their scale level require a specific weighting function and bandwidth.

For continuous variables a second order Gaussian kernel is used

$$l(X_i, x, \lambda) = \frac{1}{\lambda\sqrt{2\pi}} \exp\{-\frac{1}{2}\left(\frac{X_i - x}{\lambda}\right)^2\}.$$

Following from the Gaussian distribution points close to x get higher weights than boundary points. For continuous variables the bandwidth λ can get all values greater than 0. A small bandwidth close to 0 means that the neighborhood is very small. Thus, only a few points are used to estimate the local parameters and therefore the estimated parameters may vary considerably for different x. This case allows for a high degree of nonlinearity concerning the relationship between the variables. In contrast, a high bandwidth means that most points are considered for estimating the parameters which are similar for different x. The influence of the continuous variable is almost linear in this case.

For unordered categorical variables Racine & Li (2004) propose the kernel function

$$l(X_i, x, \lambda) = \begin{cases} 1, & \text{if } X_i = x \\ \lambda, & \text{otherwise} \end{cases},$$

where the bandwidth λ lies in $[0,1]$. For a bandwidth equal to zero $l(X_i, x, 0)$ is an indicator function for category x. For example, if a binary variable with categories west and east is analyzed, a the bandwidth for this variable close to 0 means that only observations lying in the west are used for this category. Vice versa for estimating the regression function for eastern Germany only observations from the east are taken into account. This implies that the influence of the variable or in other words the differences between these categories are such huge that the sample is divided into west and east both subsamples are estimated separately.

A bandwidth of 1 means that $l(X_i, x, 1)$ is a constant function. For example, for smoothing category west all observations are used, those from the west and the east. Thus, there is no difference between both categories. In this case there is no influence of the underlying covariate and it is "smoothed out".

For ordered categorical variables Racine & Li (2004) propose a kernel function

$$l(X_i, x, \lambda) = \lambda^{|X_i - x|}.$$

Again, the bandwidth λ lies in $[0,1]$. For $\lambda = 0$, $l(X_i, x, 0)$ is also an indicator function for category x and for $\lambda = 1$, the kernel function is constant. The interpretation is equal to unordered case. If the bandwidth lies between 0 and 1, observations from the same category get the weight 1. Direct neighbor categories are weighted with λ and observations with one intermediate category with λ^2. Categories close to the category of interest get higher weights than distant ones.

The product kernel K is the product of the weighting functions for P regressors

$$K = \prod_{p=1}^{P} l(X_{ip}, x_p, \lambda_p). \tag{3.1}$$

The estimation is done by local polynomial estimation. Here, the focus is on the two simplest cases of local constant and local linear estimation. In the case of local constant for every x only an intercept is estimated. In the case of local linear estimation an intercept and a slope parameter are calculated by solving the resulting minimization problem which is a weighted local least squares problem

$$\sum_{i=1}^{n} [Y_i - a(x^c) - (X_i^c - x^c)^\top b(x^c)]^2 K(\cdot),$$

while the "local" part is considered by the dependence of the parameters a and b on continuous regressors x^c. Thus, the parameter estimation on the space of the continuous regressors is weighted by all smoothed covariates.

The remaining problem is finding optimal values of the bandwidth vector

$$\lambda = (\lambda_1, ..., \lambda_p)$$

which can be obtained in two ways. One possibility is minimizing the improved Akaike Information Criterion for nonparametric methods of Hurvich et al. (1998) which is given by

$$AIC_C = \ln(\hat{\sigma}^2) + \frac{1 + tr(B)/n}{1 - \{tr(B) + 2\}/n}$$

where

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{g}(X_i)\}^2 = Y'(I-B)'(I-B)Y/n.$$

$\hat{g}(X_i)$ is the estimated nonparametric regression function and B is the $n \times n$ hat matrix including the kernel weights.

Alternatively, Li & Racine (2004) propose obtaining λ using a data-driven least squares cross-validation approach, where the objective function

$$CV(\lambda) = n^{-1}\sum_{i=1}^{n}(Y_i - \hat{g}_{-i}(X_i))^2 M(X_i)$$

is minimized while $\hat{g}_{-i}(z_i)$ is the leave-one-out kernel estimator of regression function g, and M is a weighting function bounded between 0 and 1, usually set to $M = 1$ (see Li & Racine, 2004).

The inclusion and smoothing of discrete covariates is a milestone in the area of nonparametric kernel density estimation. Basically offered for continuous variables many problems could not be answered because in empirical samples there are usually also discrete covariates. Before Racine & Li (2004) propose their new method there was only the so-called frequency approach for considering discrete data. Applying the frequency approach the sample is divided in cells each with observations offering the same combination of categories of the discrete variables. For every cell a single nonparametric kernel regression is estimated on the continuous covariates. Obviously, even a small number of discrete regressors may produce so many cells, that there are only a few or even no observations in several cells. The advantage of a generalized product kernel is that due to smoothing of continuous and discrete data always all observations are used for the estimation and so reliable information even about observations from sparsely populated cells is obtained.

Furthermore, the use of a nonparametric approach involves an important advantage in comparison to MRW's parametric regression. The nonparametric model allows all kinds of interactions between variables, both linear and higher order.

3.2.2 Testing for parametric misspecification

Thus there are arguments from an economic perspective for choosing nonparametric models, but in which cases should nonparametric models be taken into account from an econometric perspective?

When is the adaption of the more complicated nonparametric model beneficial? First hints come from the estimated bandwidths. If they propose nonlinear influences of several variables, this is a hint for a parametric misspecification. This can be tested with a formal test for parametric misspecification proposed by Hsiao et al. (2007) (hereafter Hsiao-Li-Racine-test).

The main idea of the test is that if the parametric specification is right, the conditional mean of the response variable is equal to the linear specification. This is equivalent to the null hypothesis that the conditional mean of the residuals is zero

$$H_0 : E(u_i|X_i) = 0.$$

The population test statistic for the null is given by

$$I = E[u_i E(u_i|X_i) f(X_i)] \geq 0$$

and it is zero if the null is true. Deriving the sample test statistic $E(u_i)$ is replaced by the sample mean of the residuals and the conditional mean of the residuals weighted by the density of the regressors, $E(u_i|X_i)f(X_i)$, is estimated by a leave-one-out kernel estimator. Thus, the sample test statistic is given by

$$I_n = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_i\left(\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\hat{u}_jK(\cdot)\right)$$

The distribution of I_n is obtained by resampling. Therefore, a bootstrap sample, the related residuals and the test statistic is computed a large number of times. The test decision is as follows: Reject the null if the test statistic is larger than the α-percentile of the resampled distribution.

There are a lot of set-up-parameters which can be changed, e.g. the types of kernel functions for different kinds of variables, the method for computing the bootstrap sample and bandwidth selection and so on. Haupt et al. (2010) point out that the test may be sensitive w.r.t. the test setup. Thus, different settings should be used to check potential sensitivity of a decision.

What does it mean if the parametric specification of classical β-convergence is rejected? The classical analysis is derived from economic theory and thus, the implicit linear model which is assumed to be adaptive for all convergence data. If the linearity assumption is rejected also the convergence concept is invalid. In this case other convergence concepts should be used which are adaptive for nonparametric methods.

3.3 Heterogeneity

Another main point of criticism of the classical convergence regression is that there are several forms of neglected heterogeneity causing invalid estimation results. First, heterogeneity arises from the conditional distribution of the regressand. Second, heterogeneity may occur in the cross-sectional dimension meaning countries behave heterogeneously. Third, the problem can also be caused by changes over times.

Addressing the first issue “Canarella & Pollard (2004) apply linear quantile regression [using the by human capital extended growth model (and not MRW’s original data)] and find parameter heterogeneity between lower and higher quantiles of the income distribution for all explanatory variables but homogeneity inside the lower and higher quantiles. Ram (2008), focusing on conditional β- convergence, finds heterogeneity in convergence rates (and other explanatory variables) for bottom and top quartiles, but does not back up his results with a formal test"[7]. Mello & Perrelli (2003) test for location shift in different growth models. In contrast to recent contributions identifying heterogeneity as the primal source of misspecification, the formal and graphical analysis of Haupt & Petring (2011) does not reveal evidence for heterogeneity in a parametric and nonparametric quantile regression framework.

Several works cover the latter two issues. Using panel data for 102 non-oil-producing countries Lee et al. (1997) allow for individual convergence coefficients and find considerably different coefficients. "Masanjala & Papageorgiou (2004) use nonlinearities in the production function to verify and explain potential parameter heterogeneity. The nonlinearity is introduced by using a Constant-Elasticity-of-Substitution specification instead of Cobb-Douglas as the latter is essential for the linearity of the model. Alfo et al. (2008) test for cross-country heterogeneity by using bivariate mixture models"[8]. Furthermore, Phillips & Sul (2003, 2007a,b, 2009) develop a log t regression for analyzing convergence of countries or regions under heterogeneity. For the case of divergence they also propose a clustering algorithm for searching convergence clubs instead of assuming "global" convergence.

In the following subsections two methods for capturing different kinds of heterogeneity are described. In Subsection 3.3.1 quantile regression is presented. The method is proposed by Koenker & Bassett (1978) and allows for detecting location scale effects of the conditional distribution of per capita income growth rates. Subsection 3.3.2 describes the log t regression and a clubbing algorithm of Phillips & Sul (2003, 2007a,b, 2009) considering heterogeneity by allowing for individual effects and individual technology levels.

[7] Cited from Haupt & Petring (2011).
[8] Cited from Haupt & Petring (2011).

3.3.1 Location scale effects of growth rate distribution: quantile regression

For analyzing heterogeneity over the conditional distribution of the growth rates quantile regression of Koenker & Bassett (1978) is useful. As well as OLS, with quantile regression the relationship between the growth rate of per capita income and explanatory variables can be analyzed. The difference between both methods is the aspect of the conditional distribution of the growth rates which is estimated. OLS estimates the conditional mean of the dependent variable while quantile regression estimates one conditional quantile or a set of conditional quantiles. Thus, instead of only one characteristic the whole conditional growth rate distribution can be estimated. The estimated coefficients for individual quantiles can be compared and checked for heterogeneity. A location scale effect meaning that the slope parameters of the individual conditional quantiles differ, indicates this kind heterogeneity across the growth distribution. In this case the conditional quantiles should be analyzed separately because different regression models are assumed for different quantiles. After describing quantile regression generally a procedure for detecting location scale effects is discussed at the end of this subsection.

Quantile regression uses another loss function than OLS. The idea of regression analysis is that the regression line should be estimated such that the expected "loss", the weighted difference u between the observed and estimated values ($u = y - E[y|X]$), is minimal. The loss function gives the weights for different errors. OLS is based on a symmetric and quadratic function of the form $L(u) = u^2$ proposed by Legendre and Gauss about 1800. Using the quadratic loss, the optimal predictor for y is its expectation $E(y)$. $E(y)$ also minimizes the mean squared error.

However, quantile regression is based on an asymmetric absolute loss function

$$L(u) = \begin{cases} (1-\tau)|u|, & \text{if } u < 0 \\ \tau|u|, & \text{if } u \geq 0 \end{cases} .$$

The optimal predictor in this case is the τ-th quantile of y.

In which situations does quantile regression provide additional benefit? From the discussion about loss functions it can be seen that it is the sample mean which minimizes the quadratic loss function. For computing the mean the values of the observations are used.. Thus, the instrument is highly sensitive to outliers in y. On the other hand quantile regression estimates the median or other quantiles, which are not sensitive to outliers in y because for computing quantiles the ranks of the observations are used instead their values. The main advantage of this method is that it allows estimating the whole distribution instead of a single characteristic of the dependent variable.

The minimization calculus of quantile regression is

$$\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} \tau |y_i - x_i'\beta|^+ + (1-\tau)|y_i - x_i'\beta|^-, \tag{3.2}$$

where $|y_i - x_i'\beta|^+$ is a notation for positive residuals and $|y_i - x_i'\beta|^-$ denotes the negative residuals. Because of the absolute values there is no closed-form solution, but it can be transformed into linear programming.

This yields the minimization problem

$$\min_{(\beta,u,v) \in \mathbb{R}^k \times \mathbb{R}^{2n}_+} \{\tau u + (1-\tau)v | X\beta + u - v = y\}, \tag{3.3}$$

where u is a $(n \times 1)$-vector with the positive parts of the residuals, v is a $(n \times 1)$-vector with the negative parts of the residuals and X is the $(n \times k)$ regression design matrix. Thus the sum of the absolute residuals is minimized, whereas the positive parts are weighted with τ and the negative with $(1-\tau)$ under the constraint of the validation of the regression function. The solution of the minimization problem can be found by using a simplex or interior-point algorithm (see Koenker, 2005). The optimal vector of parameters is denoted with $\hat{\beta}(\tau; y, X)$.

In addition to already described advantages, quantile regression comes with another characteristic, namely the equivariance properties of $\hat{\beta}(\tau; y, X)$. Equivariance means that shift or scale transformations on one or more variables have no fundamental influence on the interpretation of the estimates. The basic properties are:

(i) $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$

(ii) $\hat{\beta}(\tau; -ay, X) = -a\hat{\beta}(1-\tau; y, X)$

(iii) $\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma$

(iv) $\hat{\beta}(\tau; y, AX) = A^{-1}\hat{\beta}(\tau; y, X)$

The properties (i) and (ii) constitute some kind of scale equivariance. However, property (iii) indicates a shift equivariance and (iv) is called equivariance to reparameterization of design.

There is another, much more fundamental property which is elementary to understand the real possibilities of quantile regression. This is the equivariance to monotone transformations, $Q_{h(Y)}(\tau) = h(Q_Y(\tau))$, where $h(Y)$ is a nondecreasing function on $\mathbb{R}$. Hence, the quantiles of a transformed random variable are the transformed quantiles of the untransformated variable. This is a very important property which the mean does not have in general.

Under homoscedasticity conditional quantiles lie parallel to each other and thus have the same slope parameter. In this case differences in covariates shift the quantile curves but they do not change their shape or scale. Thus, there is parameter homogeneity over quantiles and analyzing the median effect is sufficient.

Testing the equality of slope parameters Koenker & Bassett (1982) propose a Wald-test. The test allows for linear restrictions of the vector of slope parameters $\beta = (\beta_{\tau_1}, ..., \beta_{\tau_m})$ which are summarized in the matrix R with rank q. The null hypothesis has the form

$$H_0 : R\beta = r,$$

where r is a vector of constants. The test statistic is given by

$$T = (R\hat{\beta} - r)'(AVar(R\hat{\beta} - r)^{-1})(R\hat{\beta} - r) \sim \chi^2_q.$$

If the null hypothesis is rejected, heterogeneity over the growth rate distribution is assumed.

3.3.2 Individual effects and technology levels

This subsection is retained from Haupt & Meier (2011).

Phillips and Sul (2003, 2007a,b, 2009), hereafter PS, argue that classical convergence analysis based on (2.23) is prone to deliver inconsistent results and invalid convergence tests due to potential heterogeneity in the convergence parameter β over time, countries, and individual technology levels. PS show that due to omitted heterogeneity the error term in (2.23) includes endogenous variables and variables which are correlated with dependent and independent variables. As a remedy PS suggest to enable a variation of the transition parameter and growth rate over districts and time[9]. They propose a nonlinear dynamic factor model

$$\log(y_{i,t}) = a_{i,t} + x_{i,t}t = \left(\frac{a_{i,t} + x_{i,t}t}{\mu_t}\right)\mu_t \stackrel{\text{def}}{=} b_{i,t}\mu_t, \tag{3.4}$$

where $x_{i,t}$ is an individual technology process parameter, $b_{i,t}$ is the idiosyncratic time-varying element and μ_t a common trend factor measuring global technological progress.

Then $b_{i,t}$ 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 district i in period t. By eliminating the global growth component, the relative transition path

$$h_{i,t} = \log(y_{i,t})/N^{-1}\sum_{i=1}^{N}\log(y_{i,t}) = b_{i,t}/N^{-1}\sum_{i=1}^{N}b_{i,t}$$

measures the transition element for economy i in period t in relation to a cross-section average. Then global convergence — all countries have the same fraction of global per capita income — is assumed to be present if

$$h_{i,t} \to 1, \quad \text{for all } i, \quad \text{as } t \to \infty. \tag{3.5}$$

[9] Note that heterogeneity of parameters in (2.22) may also occur across the conditional distribution of the growth rates g_{it}. Haupt and Petring (2011) apply quantile regression estimation and test but do not find empirical evidence in favor of such types of heterogeneity using the data from Mankiw et al. (1992). Hence this issue will not be pursued here.

The log t regression of Phillips and Sul (2007a,b, 2009)

$$\log(H_0/H_t) - 2\log(\log(t)) = a + \gamma\log(t) + u_t \tag{3.6}$$

now tests (3.5) using the mean square transition differential $H_t = \frac{1}{N}\sum_{i=1}^{N}(h_{i,t}-1)^2$. In case of global convergence $H_t \to 0$ as $t \to \infty$. The authors show that $H_t \sim A/\log(t)^2 t^{2\alpha}$ as $t \to \infty$, where $A \geq 0$ is a constant and α equals the rate of cross-section transition variation dissolving over time. Under the null hypothesis the regressor diverges to ∞ and under the alternative the regressor diverges to $-\infty$. A negative value, however, does not necessarily imply that there is divergence but that there may exist some convergence clubs instead of global convergence. Using a one-sided t-test we test the null hypothesis of $\gamma \geq 0$.

Instead of global convergence there could be some convergence clubs. To identify convergence clubs PS use a clubbing algorithm consisting of five steps

<1.> (Cross-section ordering): Order countries according to the $\log(y_{i,t})$ in final period.

<2.> (Form a core group of k^*, $2 \leq k^* < N$, countries):

- <2.1> Find the first two highest successive countries for which the log t test statistic $t_k \geq -1.65$. If the condition does not hold for any $k = 2$, drop the country with highest $\log(y_{i,t})$ and restart the procedure with the remaining countries.
- <2.2> Start with the $k = 2$ countries identified in 2.1, increase k proceeding with the subsequent country from order, run the log t regression, and calculate t_k. Stop increasing k if convergence hypothesis fails to hold (i.e. $t_k < -1.65$). Take the k^* countries with the highest test statistic from all k countries satisfying the convergence hypothesis for core group.

<3.> (Sieve the data for new club members):

- <3.1> Form a complementary core group with all remaining countries.
- <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 convergence club of the candidate group and the core group.

<4.> (Recursion and stopping rule): Form a second group from all countries which fail the sieve condition in step 3 and run log t regression. If the convergence hypothesis cannot be rejected, all remaining countries form a new convergence club. Otherwise, for the remaining countries start again with step 2 for finding a new k^*.

<5.> (Club merging): Run log t regression for all groups of subsequent clubs. Merge those clubs fulfilling the convergence hypothesis commonly.

Composing the clubs in accordance to this algorithm does not ensure that the convergence hypothesis holds for each respective club. PS (2007) are aware of this problem and propose to increase the critical value c^* for raising the power of the corresponding test. Such a remedy, however, does not work in general, for instance for the German district-level data discussed in Subsection 5.2.2 or when we replace the initial cross-section ordering rule by an (equally plausible) alternative. Thus, we may want to augment step <3.> of the algorithm in a way such that convergence is assured using a data-based criterion. A straightforward method is to search for the largest group size for whose respective members convergence holds. In a first step we leave one country out at a time and run a log t regression. We form a convergence group from the countries with highest test statistic greater than -1.65. If there is no group of countries with test statistic higher than -1.65 we leave out two countries at a time in a second step and so on. If there is more than one convergent combination of countries at one step, we choose the combination with highest test statistic. The advantage is that we get the largest group of countries satisfying convergence without the existence of path dependence. The disadvantage is that for inappropriate constellations of countries caused by high sample sizes or special sorting methods the computing time increases exorbitantly. Thus, we propose another method which we include in the clubbing algorithm as 3.3.

<3.3> If the countries from core and candidate group hold convergence hypothesis commonly, go to step 4. If not, form a convergence club with the candidate country with highest test statistic and the core group. Add one candidate country at a time to convergence club, run log t regression and add the country with highest test statistic to

the convergence club. Continue adding new countries to the convergence club until no further candidate country fulfills convergence hypothesis.

In all empirical applications discussed in Section 5.2 we find that the shape of observed points in the log t regression (3.6) to be parabolic and convex for convergence clubs. This is due to the construction of the regressor. Under the null hypothesis H_t converges to zero as $t \to \infty$ as a monotonically decreasing convex function. Calculating H_0/H_t inverts this shape into a monotonically increasing convex function. Taking the logarithm damps the curvature or even linearizes the curve. The second part of the regressor $2\log(\log(t))$ is a monotonically increasing concave curve. Subtracting this second concave part from the first convex/linear curve leads to a parabolic and convex trajectory. Thus, under the null we expect a nonlinear regression relationship. Those results suggest that the interpretation of the log t regression should be handled with care.

3.4 Interaction of nonlinearity and heterogeneity

Currently there are only two papers directly linking the problems of heterogeneity and nonlinearity in classical growth regressions.

Using parametric quantile regression Haupt & Petring (2011) analyze if there is heterogeneity over the conditional distribution of the regressand in the classical growth model applied to MRW data. In a second step they check for nonlinearities of regression quantiles using nonparametric quantile regression. In contrast to recent contributions identifying heterogeneity as the primal source of misspecification, a formal and graphical analysis does not reveal evidence for heterogeneity.

Haupt & Meier (2011) address another form of heterogeneity, namely heterogeneous behavior over time and countries. Using the algorithm of Phillips & Sul (2007a,b, 2009) they first group countries or regions in clusters with homogeneous members. In a second step they estimate a classical growth regression nonparametrically while considering the heterogeneity by capturing a variable with club information.

In the following subsections both methods, nonparametric quantile regression and the two-step procedure of Haupt & Meier (2011), are described.

3.4.1 Nonlinear location scale effects: nonparametric quantile regression

As mentioned in Subsection 3.3.1, quantile regression may be preferred to classical mean regression in several situations because the method is robust to outliers and it offers a broad overview of the whole conditional distribution of the regressand instead of a single point, the conditional mean. Thus, quantile regression allows for a detection of heterogeneity over the conditional distribution.

However, linear quantile regression is not valid in the case of a nonlinear relationship between regressand and covariates. Thus, Li & Racine (2008) propose a method for estimating conditional quantiles nonparametrically. In contrast to Koenker & Bassett (1978) the authors avoid to determine conditional quantiles by a check function given in Equation (3.2). They obtain the conditional quantile by inverting the conditional CDF of y given x at the selected portion. Thus, the conditional quantile is given by the empirical distribution function (EDF)

$$q_\tau(x) = F^{-1}(\tau|x),$$

where F is the conditional CDF of y given x which can generally be estimated by

$$\hat{F} = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq y), \tag{3.7}$$

where I is an indicator function which is equal to 1 if $Y_i \leq y$ and otherwise 0. Alternatively, the distribution function F may be estimated by a weighted version of Equation (3.7). Li & Racine (2008) propose a weighting function which is equal to their generalized product kernel. The weighted version of $\hat{F}$ is given by

$$\hat{F} = \frac{n^{-1}\sum_{i=1}^{n} I(Y_i \leq y)K(X_i, x)}{n^{-1}\sum_{i=1}^{n} K(X_i, x)},$$

where K is a generalized product kernel.

In the case of a continuous dependent variable it can also be smoothed. Therefore another estimator of the conditional CDF of y is used

$$\hat{F} = \frac{n^{-1}\sum_{i=1}^{n} G((y-Y_i)/h_0)K(X_i,x)}{n^{-1}\sum_{i=1}^{n} K(X_i,x)},$$

where G is the CDF of the underlying kernel function for continuous variables and h_0 is the bandwidth for smoothing y.

For bandwidth selection Li & Racine (2008) propose a data-driven procedure. The optimal bandwidth vector is obtained by minimizing the cross-validation objective function

$$CV = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{G}_{-i}(X_i)s(Y_i,X_i)}{n^{-1}\sum i=1^n\hat{K}_{-i}(X_i)^2} - \frac{2}{n}\sum_{i=1}^{n}\frac{\hat{g}_{-i}(Y_i,X_i)s(Y_i,X_i)}{n^{-1}\sum i=1^n\hat{K}_{-i}(X_i)},$$

where $\hat{G}_{-i}$, $\hat{g}_{-i}$ and $\hat{K}_{-i}$ are leave-one-out estimators of G, g and K and s is non-negative weighting function.

3.4.2 Nonlinear modeling with convergence clubs: a two-step procedure

This subsection is taken from Haupt & Meier (2011).

We want to apply a classical convergence analysis in the sense of Mankiw et al. (1992) while allowing for data-driven heterogeneity and nonlinearity. Thus, in a first step, we assign the regions to clubs using the algorithm discussed above. In a second step we include a categorical club variable $club_i$ in (2.22) via the j dummy variables $club_{i,j}$ which are equal to 1 if country i is in club j. The resulting baseline model allows to estimate a regression line $\delta_j + \pi_j \log(y_{i,0})$ for every club j, $1 \leq j \leq m$, i.e.

$$\upsilon_i = \sum_{j=1}^{m}\delta_j club_{i,j} + \sum_{j=1}^{m}\pi_j \log(y_{i,0}) \cdot club_{i,j} + u_{i,t}. \tag{3.8}$$

In contrast to the classical convergence model (2.22), the baseline model (3.8) allows for a considerable degree of heterogeneity. However, there are very small clubs for several applications and thus the interpretation of the parameters for those clubs should be handled with care. The main point of criticism, however, is that this model may suffer from potential misclassification of the club composition (see Subsection 3.2.2). Furthermore, the model does not allow for further nonlinearities.

In order to address the problem of potential nonlinearities we can employ a fully nonparametric alternative (see Subsection 3.2.1)

$$\upsilon_i = g(\log(y_{i,0}), club_i) + u_{i,t}. \tag{3.9}$$

This approach allows to estimate not only club-level effects — which Durlauf and Johnson (1995) interpret to represent averages of the underlying individual effects for each country — but further nonlinearities.

In contrast to model based clubbing algorithms (e.g. Juárez and Steel, 2010), the method of Phillips and Sul (2007a,b, 2009) discussed in Section 3.3.2 does not provide estimates of the misclassification probabilities for each club member. A first step towards exploring potential classification error is to check for hints on the existence of positive error probabilities by inspecting whether the "selection of core groups is robust to initial data orderings" (see Phillips and Sul, 2009, footnote 11, p. 1170). Considering the problem of an unknown true ordering rule (see Canova, 2004) we try different concepts in step <1.> and check whether considerable differences in club composition are obtained. This indicates large uncertainties which should be addressed in empirical convergence analysis.

As alternatives to the amount of final period income (final ordering), hereafter denoted as ordering rule (I), as used by Phillips and Sul (2009) we employ the following. (II) Order corresponding to the average income of all years (average ordering) for capturing potential time series volatility. Phillips and Sul (2007) propose to average over the last fraction of the sample to ensure a higher influence of recent periods. (III) Another alternative is ordering according to the difference between final period income and income in first period, capturing the income change over time (difference ordering). (IV) Finally, combining the ideas on the final period and capturing volatility, a decreasing weights ordering is employed.

We note that in all of our applications discussed below the use of different ordering rules leads to considerable differences in club sizes and composition, respectively. For evaluating the empirical performance we compare the out-of-sample performance of the convergence regression models for ordering rules (I)-(IV). As an a posteriori selection criteria for ordering rules we run a cross-validation (e.g., Haupt and Petring, 2011) and choose the model with the smallest average squared error of prediction.

While there is no obvious remedy for the misclassification problem in the parametric model (3.8), the nonparametric model (3.9) may offer one. Data-driven bandwidth selection for the club variable deals with the question of uncertainty of club composition. Using the kernels proposed by Racine and Li (2004), the optimal estimated bandwidth is bounded between 0 and 1. A bandwidth of approximately 0 means that the influence of this variable is such that for estimating the function (3.9) for a club only observations from this club are used. This occurs when the functional form is sufficiently different with respect to the different clubs or if the observations show sufficiently different convergence behavior. We can interpret this in the sense that there is a rather low probability of misspecification, thus the clubs are well chosen. With increasing values of the bandwidth the error probability for club membership rises. If the bandwidth is considerably greater than 0, observations from all clubs are used to estimated regression functions for each club and thus, there is no influence of the variable. This suggests that there is evidence in favor of an only weak or even non-existent club structure. Thus, the bandwidth of the categorical club variable serves for an a posteriori quantification of the classification (and underlying error probabilities) as a whole.

By using the nonparametric approach including the club variable we obtain individual influences of each observation while considering the uncertainty with respect to club membership, instead of a single fixed convergence regression line for each club in the parametric approach. The club-structure on the other hand has the advantage of being backed up by economic theory. Although it may produce a faulty number and/or composition of clubs, the simultaneous smoothing of the continuous and the categorical variable is capable of alleviating this problem. In summary we include heterogeneity in the sense of Phillips and Sul, reduce uncertainty of club composition, and capture potential nonlinearities, and hence are able to address the main points of criticism of convergence regressions in recent literature.

Given a set of data the initial problem a researcher faces is choosing either a parametric models such as (3.8) or a nonparametric model such as (3.9). In the context of mixed continuous and categorical covariates as in the present example this problem can be addressed by applying the test of Hsiao et al. (2007) (hereafter HLR test), which is based on the generalized product kernel estimator proposed by Racine and Li (2004) discussed in Subsection 3.2.1.

Using the same nonparametric configurations used for the nonparametric regression the HLR test checks if the parametric null model (3.8) is correctly specified. Whenever the HLR test rejects the null we apply the fully nonparametric model, enjoying the benefits discussed in the previous sections.

As the HLR test employs the bandwidths of the nonparametric regression, we are able to assess the error probabilities already after applying the test. Thus, if the test does not reject the parametric null hypothesis, we inspect the bandwidth λ_k of the cluster variable: If λ_k is close to zero, the parametric and nonparametric model work analogously and we may use the parametric model because there are no hints for club misclassification. If the bandwidth λ_k is greater than zero positive classification errors have a higher probability. In this case, however, we can still estimate a nonparametric model for the theoretical price of efficiency loss compared to the parametric model.

3.5 Spatial association

Another important point of criticism for analyzing convergence data can be seen in the assumption of spatial independence of economies.

Considering technological interdependencies between economies and knowledge spillover effects Ertur & Koch (2007) present a spatially augmented Solow model (see Subsection 2.1.3). For the Non-Oil countries of MRW the authors find a significant influence of spatial externalities for a period from 1960 to 1995. Furthermore, Moreno & Trehan (1997) test for different kinds of spillovers between neighboring economies. Using cross-country data they find spillovers in absolute and conditional β-convergence regressions. Applying three tests for spatial autocorrelation Niebuhr (2001) identifies spatial autocorrelation for 71 of the 75 West German planning units between 1976 and 1996 and proposes a spatial regression model to address this problem for absolute as well as conditional convergence. Kosfeld et al. (2006) find spatial correlation up to order three in 180 labor market regions in Germany, defined by Eckey (2001) for the period from 1992 to 2000. Using a spatial ARMA model they identify unconditional and conditional β- and σ-convergence for east German regions after 1990 but only β-convergence for west German regions. Hence, east German regions seem to catch up. Eckey et al. (2007) use geographically weighted regression to prove β-convergence for German labor productivity. For the same regions as Kosfeld et al. (2006) they find different convergence rates for several regions. In contrast to most of the other publications the authors find an emerging gap between the south and the north instead of east and west. Varying convergence coefficients leave considerable doubt on the prevalent global convergence model.

The analysis in this thesis follow the ideas of Ertur & Koch (2007). In Subsection 3.5.1 different spatial models are described. Subsection 3.5.2 discusses different kinds of weighting matrices and Subsection 3.5.3 presents testing procedures for spatial dependence.

3.5.1 Spatial patterns in convergence models

Basically, there are three forms of spatial dependence which should be taken into account (see Elhorst, 2010). First, there may exist endogenous spatial effects meaning that the dependent variable of an economy depends on the value of the dependent variable of the other economies. Second, exogenous effects of the explanatory variables of the other economies are of significance. Third, there are correlated effects of unobserved lagged covariates yielding spatial correlation.

Capturing the case of endogenous effects, the so-called spatial autoregressive process (e.g.Whittle, 1954) is used. The name follows its time series pendant where the response variable depends on its own past values. In the case of geographical data it depends on neighboring geographical units. This fact is considered in the first order spatial autoregressive process given by

$$y_i = \rho \sum_{j \neq i}^{n} w_{i,j} y_j + \beta_1 x_{i,1} + \ldots + \beta_k x_{i,k} + \varepsilon_i, 1 \leq i \leq n, \tag{3.10}$$

with $\varepsilon_i \sim N(0, \sigma^2)$. The unknown parameter ρ is assumed to measure the strength of the spatial dependence. The spatial parameter is estimated via Maximum Likelihood estimation. The term $\sum_{i=1}^{n} w_{i,j} y_j$ is denoted as spatial lag, which is a linear combination of neighboring y-values. The weights $w_{i,j}$ (can) vary for every neighboring region and can be summarized to a $n \times n$ matrix of spatial weights, W, see Subsection 3.5.2 for a thorough discussion. Thus, we can rewrite Equation (3.10) for all i in vector notation as

$$y = \rho W y + X\beta + \varepsilon, \tag{3.11}$$

with $\varepsilon \sim N(0, \sigma^2 I)$.

A second spatial model is the spatial error model which used if correlated effects in the error term occur. This model is given by a linear model with a spatial autoregressive process in the error term

$$y = X\beta + u \tag{3.12}$$

$$u = \zeta W u + \varepsilon. \tag{3.13}$$

The formulation in Equation (3.13) for the spatial error model is analogous to the spatial lag model in Equation (3.10). For notational simplicity all spatial weight matrices are denoted as W. The spatial error model is the cross-section counterpart of a moving average process in a time series context. Here, the dependent variable is not explicitly affected by spatially lagged dependent or explanatory variables, but there are other spatial effects which are correlated with the dependent and explanatory variables. These effects should be considered, otherwise OLS-estimators are biased due to omitted variable bias.

Modeling spatial error and spatial lag models as well as spatial effects of the explanatory variables of other economies, the spatial Durbin model is used (see Lesage & Pace, 2009)

$$y = \rho W y + X\beta + W X\theta + \varepsilon. \tag{3.14}$$

Spatial lag and spatial error are nested in the spatial Durbin model and obtained whenever $\theta = 0$ (spatial lag) or $\theta = -\rho\beta$ (spatial error). Thus, the estimators of the spatial Durbin model are unbiased even if the true GDP is spatial lag or spatial error (see Elhorst, 2010). The spatial Durbin model is used for estimating the spatial augmented Solow model from Subsection 2.1.3 because it includes spatial lags of dependent and explanatory variables.

3.5.2 Influence of neighboring economies

An important issue in the context of spatial modeling is the definition of the weighting matrix W as it determines direction and concrete forms of the spatial effects and all the results of the analysis depend on W.

For defining the weighting matrix W mainly neighbor or distance matrices are used (e.g Lesage & Pace, 2009; Ord, 1975). The elements $w_{i,j}$ of a neighbor matrix are defined as

$$w_{i,j} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are neighbors and } i \neq j \\ 0, & \text{otherwise} \end{cases}. \tag{3.15}$$

while two regions are called neighbors if they share a common border. This classification is not without difficulties as for example, islands need special rules. Usually the main diagonal of the weighting matrix is set to zero, because an economy is not its own neighbor per convention (see Lesage & Pace, 2009).

An alternative definition W is a distance matrix, where the weights may depend on the distance $d_{i,j}$ between economy i and j as

$$w_{i,j} = f(d_{ij}),$$

where usually $f(d_{i,j}) = 1/\sqrt{d_{i,j}}$ (see Bivand et al., 2010). Thus, economies that are far apart impact smaller than neighbors. There are several points to consider for measuring the distance between two economies. First, the reference point for measuring the distance has to be defined. This reference point could be chosen on the basis of geographical aspects (e.g. middle of the region), economical aspects (e.g. point with highest GDP), or political aspects (e.g. coordinates of the capitol). Second, the unit of distance measurement has to be chosen (e.g. geographical coordinates, kilometers). Furthermore, the fact that the economies are arranged on the earth has to be considered, thus on a curved surface and not in a plane. Ertur & Koch (2007) propose using the great-circle distance which is the shortest distance between two points on the earth surface quantified walking over the earth's surface instead of going through the earth's interior.

The difference between neighbor and distance matrices is that in neighbor matrices most entries are zero as usually a region only has few neighbors, while in distance matrices all entries are strict positive. For an easier interpretation the weighting matrix often is standardized. Therefore all entries are divided by the corresponding row sum, yielding rows that sum to one for the standardized weighting matrix. Hence, average spatial weights are obtained w.r.t. each economy, that can be interpreted straightforward. Most testing procedures for spatial dependence (discussed in the subsequent subsection) assume those row-standardized matrices.

3.5.3 Testing for spatial association

There are several tests for different kinds of spatial association. A general check for spatial association is Moran's I proposed by Moran (1950). The test does not assume a certain alternative model, it checks for general spatial correlation given a fixed spatial structure of the row-standardized weighting matrix W. Thus, Moran's I may be interpreted as the spatial pendant to the classical correlation coefficient Pearson's r and is usually calculated for OLS residuals. The test statistic

$$I = \frac{N}{\sum_i \sum_j w_{i,j}} \frac{\sum_i \sum_j w_{i,j}(\varepsilon_i - \bar{\varepsilon})(\varepsilon_j - \bar{\varepsilon})}{\sum_i (\varepsilon_i - \bar{\varepsilon})^2}$$

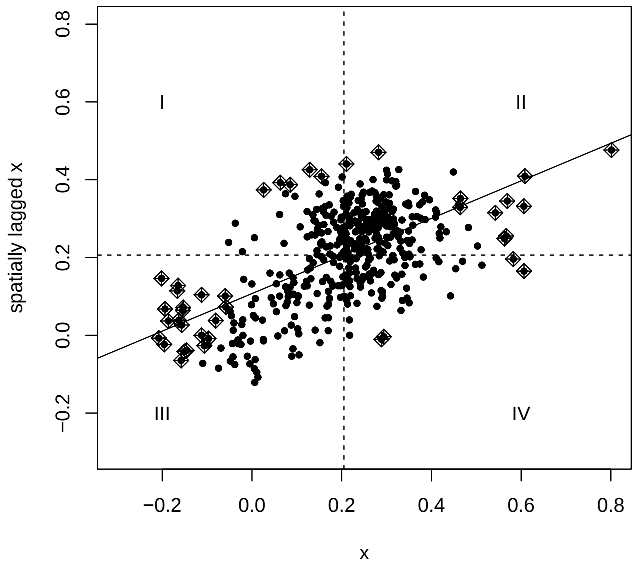
measures the correlation between the residuals and the spatially lagged residuals by dividing their covariance by the residuals variance multiplied with a variance correction factor given in the first fraction. The idea of Moran's I can be visualized by the so-called Moran scatter plot (see Anselin, 1995) which is displayed in Figure 3.1. The variable of interest, x, is plotted against its spatial lags. The dashed lines display the means of the variables such that four quadrants (I, II, III, IV) are obtained. Assuming no correlation the points should be equally dispersed over all quadrants. If there is a positive correlation there are more points in II and III than in the other quadrants meaning that an observation which is higher (lower) than the mean in x is also higher (lower) than the mean of the spatially lagged x on average. In this case a positive slope is obtained, when a regression line (the solid line) is estimated. Vice versa, a negative correlation means that there are more points in I and IV. The plot also allows for a visual outliers detection.

The remaining question is, whether the correlation is significant. Thus, the expectation of Moran's I under the null of no spatial correlation is needed which is given by $E(I) = -1/(N-1)$ (see Elhorst, 2010). The corresponding test hypotheses are

$$\mathrm{H}_0 : I = \frac{-1}{N-1} \text{ vs. } \mathrm{H}_1 : I > \frac{-1}{N-1}.$$

Transforming the test statistic to standard normal distribution the test is easily done.

Figure 3.1: A Moran Scatter plot example



There are other tests, where a specific spatial model is under the alternative. Using the Lagrange Multiplier (LM) tests proposed by Anselin (1988a,b) different spatial models can be compared in classical OLS framework. Generally, a spatial model with a spatial lag and spatial errors is assumed

$$\begin{aligned} y &= \rho W y + X\beta + u \\ u &= \zeta W u + \varepsilon, \end{aligned} \tag{3.16}$$

where it is assumed that $\varepsilon \sim N(0, \sigma^2)$. There are four models nested in this general spatial model.

i) $\rho = 0$ and $\zeta = 0$, $\quad y = X\beta + \varepsilon$ (simple linear regression model)

ii) $\rho \neq 0$ and $\zeta = 0$, $\quad y = \rho W y + X\beta + \varepsilon$ (spatial autoregressive model)

iii) $\rho = 0$ and $\zeta \neq 0$, $\quad y = X\beta + (I - \zeta W)^{-1}\varepsilon$ (spatial error model)

iv) $\rho \neq 0$ and $\zeta \neq 0$, $\quad y = \rho W y + X\beta + (I - \zeta W)^{-1}\varepsilon$ (spatial autoregressive and spatial error model)

Starting with the most restrictive model (classical OLS) in H_0, it can be tested whether the simple linear regression model (i) is preferred to one or more of the models (ii), (iii), and (iv). Therefore, model (i) and the alternative of interest (ii), (iii), or (iv) are estimated and the LM test statistic evaluated under the null is computed

$$LM = \frac{d^2}{I_0} \sim \chi^2(q),$$

where q is the number of restrictions. $d = \partial \log(L)$ is the slope of the log likelihood under the null and I_0 is the Fisher Information. For the three tests Anselin et al. (1996) develop robust alternatives where the details are skipped here. Analogously, it can be tested whether the spatial Durbin model (3.14) is preferred to the spatial lag or spatial error model. Therefore, the spatial Durbin model is estimated as alternative and compared by the LM test with the restricted models $H_0 : \theta = 0$ (spatial lag) or $H_0 : \theta = -\rho\beta$ (spatial error) (see Elhorst, 2010).

Chapter 4

Level of aggregation

Primarily, growth and convergence modeling was developed for analyzing cross-country data. The question of convergence is also interesting for lower aggregated units within a country or a group of countries, e.g. between Federal states or districts. There are a lot of publications in current literature about regional convergence analysis (e.g. Sala-i Martin, 1996b) using the same concepts for different aggregation levels. Is this reasonable or differ the results for different levels of aggregation? Barro & Sala-i Martin (2004) deal with this question.

β-convergence is divided into two approaches, absolute and conditional β-convergence (see Subsection 2.2.1). Absolute convergence means that all countries tend to a common steady state while applying conditional convergence assumes different steady states for all or even most countries. That is to say conditional convergence should be used if there are different steady states and absolute convergence is a common steady state is assumed.

What is about different aggregation levels? The basic parameters for example technology, preferences and institutions are similar for highly disaggregated regional units. Probably, there are differences but they are small. This implies that similar or even equal steady states are assumed for disaggregated units. Thus, for highly disaggregated data absolute convergence may be applied. On the contrary, different countries which are highly aggregated units (especially if they are not grouped, e.g. in OECD members) offer big differences in the factors technology, preferences and institutions. Thus, different steady states are assumed for highly aggregated units and in this case conditional β-convergence may be applied.

Within this thesis units of different levels of aggregation are regarded and empirically analyzed. In the following sections the data with different levels of aggregation used in Chapter 5 are described.

4.1 Cross-country data

On cross-country level the data sets which are analyzed in Chapter 5 are taken from Penn World Tables. From this bases MRW and PS select different groups of countries which are presented in the following subsections. The original data are used in several of our applications in Chapter 5.

4.1.1 Data of MRW

The database of MRW includes countries, which are selected using the following criteria. First, all considered variables must be available for the countries. Second, oil production may not be the dominant industry in the countries. Furthermore the authors divide the resulting countries into three overlapping groups. The first group consists of the so-called “Non-Oil countries”. The members of this group are 98 countries achieving the criteria mentioned above. The second sub-sample is called “Intermediate countries”. Included in this group are the Non-Oil countries with more than one million citizens. The name intermediate can be seen in the sense of “representative” and that is the reason why very small countries (which are not representative) are excluded. The subsample contains 75 countries. The third subsample is called “OECD-countries”. This sample includes the 22 OECD member states with more than one million citizens.

The countries are listed in Table A.1 in the appendix.

The authors analyze both absolute and conditional β-convergence. Therefore, data on standard of living, the saving rate, the growth rate of working-age population, and human capital are needed.

For per capita income the real GDP per capita is taken. Using data from the Penn World Table, Mankiw et al. (1992) take the real GDP in 1985 and divide it by the working-age population (15-64 years old) of the same year. The initial income is developed in the same manner but the data come from 1960.

The saving rate is represented by the GDP share of investment (including government investment) divided by 100. Taking investment rates as a measure for savings rates is not obvious at first sight, but the idea is comprehensible. Investment rates are the share of output which is used to replace or enlarge the stock of physical capital. This investment can be interpreted as the part of output which is not spend for the presence but which is invested or saved for future production. The data of investment rate is derived from the Real National Accounts.

The growth rate of the working-age population is assessed by the growth rate of total population. The reason for that constraint is that no reliable data is available for several developing countries. The average growth rates over the period of 1965 to 1985, divided by 100, are used[10]. The depreciation rate δ is assumed to be constant over time and countries, so that $\delta = 0.05$.

The last and most complicated variable is human capital. Mankiw et al. (1992) uses the share of working-age population, which attends the secondary school. This variable is compounded by two factors. First, there is the fraction of eligible population (12-17) visiting a secondary school, which is taken from UNESCO (1988). Second, this variable will be multiplied by the fraction of people from the working-age population in school age (15-17). The authors discuss several issues which may be critical applying such a construction of this variable.

[10] The data stem from the International Bank for Reconstruction and Development (1988).

4.1.2 Data of Phillips and Sul

For estimating their dynamic factor model considering heterogeneity in form of individual effects and technology levels Phillips & Sul (2007a,b, 2009) only consider one covariate, the per capita income over several years.

In Phillips & Sul (2009) the authors use data from Penn World Table Version 6.2. The data set includes 152 countries from 1970 to 2003. The data are summarized in Table A.2.

.

4.2 Intermediate aggregation level

The data of Japanese prefectures are used by Barro & Sala-i Martin (2004) for analyzing absolute convergence. The data set includes the income in billion yen for the 47 Japanese Prefectures from 1950 to 1990[11]. The prefectures are displayed in Table A.3 in the appendix.

4.3 Regional data

The data are taken from the regional data base of the statistics agencies of German states and the federation http://www.regionalstatistik.de. For the initial per capita income the GDP of the 439 districts and district-free cities is divided by the number of their citizens. This variable is available for the period of 1995 to 2006. For analyzing periods longer than one year, the initial income from the beginning of the period is used. For analyzing convergence, the growth of per capita income is given by the natural logarithm of the per capita income at the end of the period minus the natural logarithm of the per capita income at the beginning of the period.

The population growth is calculated by the difference between the number of cit-

[11]The data are taken from http://www.columbia.edu/∼xs23/data.htm.

izens in the final and the previous period divided by the number of citizens in the previous period. Population data can be found for 1995 until 2007. The constant depreciation rate of 5% is added to the growth rate and for longer periods the average growth rates over all years in the period are used.

It is difficult to find data for the share of investment, as there is no data available for every economic sector at regional level. Hence, the absolute investment in the manufacturing industry is used as this is common in literature, compare Seitz (1995). Here, the problem is that the share of investment is underestimated because only the investment in the manufacturing industry is considered. Although this is one of the largest sectors in most districts, it should be noted that the share of manufacturing industry on the whole economy influences the value of investment.

Another problem in this context is that the investment data contain several missing values because the statistics agencies are sworn to secrecy for some districts and times. Thus, the missing values have to be estimated. Therefore, the data are analyzed by year and federal state (Bundesland). The aggregated values for the federal states are available for every year. So, for every year and federal state, the available values are added and the difference to the aggregate is built. Next, the resulting difference is splitted on the missing districts in accordance to their proportions on investment, which is estimated by proportions from available periods. In the following the estimation of several missing observations is described.

Table A.4 lists the missing values with regard to the associated federal state and year. The last column additionally names the years in which all observations for the missing values are available. These years are used to calculate the mean proportion on investment for the missing districts. For example, in 1995 in Nordrhein-Westfalen there are no observations for the district-free cities Köln and Leverkusen. Thus, the differences between disaggregate and aggregate sums are splitted for these two cities based on shares calculated for the years, in which values of both cities are available (here 1996, 2002, 2004, 2005). Using the mean share the rest investment in 1995 is splitted on Köln and Leverkusen.

In contrast to MRW and other studies about cross-country data using shares of students visiting secondary school or years of schooling as measures for human capital recent contributions for regional data is followed. Founded in the compulsory schooling in Germany both variables can not differ significantly over the districts. Thus, human capital is measured by the number of employees liable for social insurance, who finished professional school, university of applied science, or university. Further this variable is divided by the number of all employees liable for social insurance. This is common in literature (e.g. Seitz, 1995, Niebuhr, 2001).

Chapter 5

Empirical applications

5.1 Assessing parametric misspecification in classical growth regression

This section is taken from Haupt & Petring (2011). In this section we pursue the criticisms of nonlinearity and heterogeneity, by analyzing MRW's basic growth model and using the original data set with only $n = 75$ observations. This contrasts most of the empirical contributions, who use both extended models and extended data sets or several waves of panel data. Our point is that the proposed robust and nonparametric methods work very well even in this problematic data situation with a small number of observations, where some of these observation have high leverage. Our proposal is based on the tight connection of the issues of nonlinearity and heterogeneity with the problem of potential non-robustness. The latter has been widely neglected in the growth regression literature as argued by Zaman et al. (2001). The use of full nonparametric regression is found to be the most adequate approach to MRW's classic growth model.

The remainder of the section is structured as follows. First, we briefly introduce MRW's basic growth model and carefully investigate the question whether there is evidence for parametric misspecification by applying a recently proposed nonparametric test. In addition we calculate (the distribution of) goodness-of-fit measures and perform a robustness check via repeated sample splitting for both parametric and nonparametric models of conditional mean (Section 5.2.1). Second, we investigate the heterogeneity across the growth distribution by applying both parametric and nonparametric quantile regressions (Section 5.2.2). Finally, we briefly summarize our results (Section 5.2.3).

5.1.1 Original data analysis and parametric misspecification

For the following analysis we use MRW's original data (displayed in their paper) and first consider the classical growth model (see MRW, section I.)

$$\mathtt{l85}_i = \beta_1 + \beta_2 \mathtt{ligdp}_i + \beta_3 \mathtt{lpop}_i + \varepsilon_i, \tag{5.1}$$

where `l85` is the logarithm of total output measured by the real GDP in 1985 divided by the working-age population aged 15-64 years in the same year, `ligdp` denotes the natural logarithm of the saving rate represented by the GDP share of investment divided by 100, `lpop` is population growth measured as the natural logarithm of the average growth rates over the period from 1965 until 1985 divided by 100, plus the depreciation rate assumed to be constant at 0.05, i is an index of a cross-section of countries, and ε is an error term.

Ordinary least squares regression of this model —which still is the workhorse in any growth econometrics text and works reasonably well with an R^2 of 59.9%— will serve as our parametric benchmark model. A thorough survey and discussion of the literature following MRW on growth (and convergence) regressions can be found in Durlauf & Quah (1999).

As a first check of potential misspecification we test this null hypothesis of a linear parametric model (5.1) against an alternative of parametric misspecification, using the test proposed by Hsiao et al. (2007), hereafter denoted as Hsiao-Li-Racine-test. Employing the wild bootstrap variant of the test the p-value is found to be equal to 0.09. Thus, there is some but no decisive evidence against the null (see also

Maasoumi et al., 2007 and earlier work in this vein of Durlauf & Johnson, 1995). The p-values for alternative bootstrap methods are slightly lower.

A further criterion can be found by comparing the in- and out-of-sample performance of the parametric model (5.1) and a nonparametric alternative. For the latter, the linear parametric conditional mean assumption

$$E[\texttt{l85}_i|\texttt{ligdp}_i, \texttt{lpop}_i] = \beta_1 + \beta_2 \texttt{ligdp}_i + \beta_3 \texttt{lpop}_i \tag{5.2}$$

in (5.1) is replaced by a nonparametric conditional mean assumption

$$E[\texttt{l85}_i|\texttt{ligdp}_i, \texttt{lpop}_i] = g(\texttt{ligdp}_i, \texttt{lpop}_i), \tag{5.3}$$

i.e. a general nonparametric specification of the systematic part with unknown regression function $g(.)$. The first application of a fully nonparametric approach to growth convergence regression is Maasoumi et al. (2007). Clearly, approach (5.3) allows for a very general production technology without the need to assume the validity of a local linearisation (compare e.g., Masanjala & Papageorgiou, 2004).

The (pseudo) goodness-of-fit $\tilde{R}^2$, measured by the squared correlation between actual and fitted response, shows the higher in-sample accuracy of the nonparametric model[12] compared to the parametric model with an $\tilde{R}^2$ of 0.599 for the former compared to 0.663 for the latter. This may be expected a priori due to the higher flexibility and thus potential overfitting of the nonparametric model. To avoid the latter problem we check the accuracy of both models with respect to their out-of-sample performance employing a hold-out-sample strategy.

Thus, as a cross-validation, we randomly split the sample into a 90% sub-sample for estimation, which is then used to predict the remaining 10% of the observations (see Haupt et al., 2010, 2009 for a detailed description of cross-validation with nonparametric regressions and R code). From the latter we can calculate the mean squared

[12] More specific, we estimate a local-linear model using the expected Kullback-Leibler cross-validation proposed by Hurvich et al. (1998) and a second-Order Gaussian kernel. The bandwidths for the covariates `lpop` and `ligdp` are 0.0934 (scale factor: 1.3497) and 1.1431 (scale factor: 5.2912), respectively. For all nonparametric computations in this paper we use version 0.30-1 of the `np`-package for `R` from Hayfield & Racine (2008).

error of prediction (MSEP) for both models (denoted as, say $\mathbb{M}_0$ and $\mathbb{M}_1$), and the relative MSEP as $\text{MSEP}_{\mathbb{M}_0}/\text{MSEP}_{\mathbb{M}_1}$. If this ratio is larger than one, then the nonparametric model has a superior out-of-sample performance. This calculation step is iterated 10,000 times and figure B.1 displays the empirical density of the obtained relative MSEP.

We observe that in approximately 73% of the cases the nonparametric model has a smaller MSEP compared to the parametric model. In order to test[13] for the statistical significance of this result we employ the following hypotheses:

$H_0 : E[\text{MSEP}_{\mathbb{M}_0}] - E[\text{MSEP}_{\mathbb{M}_1}] \leq 0,$
$H_1 : E[\text{MSEP}_{\mathbb{M}_0}] - E[\text{MSEP}_{\mathbb{M}_1}] > 0.$

The resulting p-values support our former result and deliver clear evidence against the null (see table A.7 below). Additionally, using the 10,000 observations of the sub-sampling distribution we calculate the median (and lower/upper quartiles) of $\tilde{R}^2$ for the nonparametric mean model with 0.663 (0.650; 0.676) and for the parametric model 0.597 (0.582; 0.612), respectively.

All in all, these results cast considerable doubt on the correct parametric specification of model (5.1). The estimated manifold of the nonparametric mean regression displayed in figure B.2, nicely reveals the different forms of local nonlinear impacts that the two covariates `lpop` and `ligdp` exert on the response variable `l85`. These findings are supported by the detailed results on cross-validations for all models presented in this and the following section (see tables 3 to 5).

5.1.2 Robustness, heterogeneity, and conditional quantiles

A RESET specification test of Ramsey (1969) based on OLS estimation of (5.1) cannot reject the null at any reasonable significance level, whereas specification tests based on subsets of the data such as Harvey & Collier (1977) and Utts (1982) clearly reject the null. The latter two, applied to the data ordered by `l85`, sug-

[13]We thank Jeff Racine for suggesting a formal test here.

gest that functional form may vary across the distribution of the response variable. Standard outlier diagnostics based on OLS residuals reveal that there are three observations (Venezuela, Tanzania, Zambia) with rather large residuals, which may cast some doubt on the validity of MRW's OLS results (see table A.5). Zaman et al. (2001) propose the use of Rousseeuw (1984) least trimmed squares (LTS) as a robust alternative. The robust estimator reported in table A.5 is calculated by applying the procedure outlined in Zaman et al. (2001). It discards only two observations (Venezuela and Zambia), and leads to a considerable increase in R^2 to 67.5%.

An alternative robust method — quantile regression (Koenker & Bassett, 1978) — allows a direct investigation of the assumption of parameter homogeneity, without the need to sacrifice certain outlier observations, which may be awkward without further subject matter knowledge or respective a priori information. For linear quantile regressions we replace assumption (5.2) by the linear parametric conditional quantile assumption

$$Q_{\vartheta}[\mathtt{l85}_i|\mathtt{ligdp}_i,\mathtt{lpop}_i] = \beta_1 + \beta_2\,\mathtt{ligdp}_i + \beta_3\,\mathtt{lpop}_i, \tag{5.4}$$

for a quantile ϑ, where $\vartheta \in (0,1)$ and $Q_{\vartheta=0.5}[.|.]$ is the conditional median. Hence, in table A.5 we compare the results of OLS estimation (MRW, Table I) of (5.1) and the robust LTS-based regression, with linear median, upper and lower quartile regressions.

At first view, there is some difference in numerical values between the slope coefficient estimates for the lower, upper, and median quartile regression, respectively. In contrast to Ram (2008), however, we wish to apply a formal test for the statistical significance of this difference. Applying the joint robust Wald test of Koenker & Bassett (1982) for the null hypothesis of identical slope parameters across quartiles gives a p-value of 0.09 and thus no clear evidence for heterogeneity using approach (5.4).

As a further generalization, we compare the results from estimation of the nonparametric conditional mean (5.3) with the results from a fully nonparametric quantile regression following as recently proposed by Li & Racine (2008). By employing the nonparametric conditional quantile assumption

$$Q_\vartheta[\mathtt{l85}_i|\mathtt{ligdp}_i, \mathtt{lpop}_i] = g(\mathtt{ligdp}_i, \mathtt{lpop}_i), \tag{5.5}$$

the latter approach combines the functional flexibility of the fully nonparametric approach with the capability to capture potential heterogeneity across the conditional growth distribution from quantile regression, and the additional benefit of robustness. Figure B.3 shows the conditional partial effects from the nonparametric quartile regression model and the nonparametric mean regression, where we observe the impact of varying one covariate while the other covariate is held constant at its mean or median value, respectively. In analogy to the linear models displayed in table A.5, we again observe differences between mean and median estimates. Although the results from quartile regression reveal some differences to the nonparametric mean approach, again there are no visible differences in (local) curvature across the conditional growth distribution.

Finally, we summarise the performance of approaches (5.2)-(5.5) with respect to goodness-of-fit and cross-validation. The cross-validation median (and interquartile ranges) of $\tilde{R}^2$ are displayed in table A.6. Table A.7 displays the p-values from pairwise t-tests on cross-validated MSEP as outlined in Section 2, where we observe that approach (5.3) dominates all other approaches with respect to both MSEP and MAEP.

Following MRW we extend our analysis to a larger dataset, including the Non-Oil countries. The different variants of the Hsiao-Li-Racine-test produces p-values smaller or equal to 0.02. The cross-validation reveals that the nonparametric model dominates the parametric model in 76% of the cases and supports the result of the Hsiao-Li-Racine-test. Using quantile regression we again find no evidence for heterogeneity, as the Wald test for the null hypothesis of identical slope parameters across quartiles has a p-value of 0.78.

5.1.3 Conclusion

The findings of our empirical analysis can be summarised in the following points. First, in accordance to some contributions in the literature (using alternative data and models), our results generate additional empirical evidence on parametric misspecification of classical growth regression models proposed by MRW. Second, in sharp contrast to recent contributions, we cannot find evidence for heterogeneity even for the extremely parsimonious model under study here. Employing the method of quantile regression we find that this holds true for both parametric and nonparametric approaches. Third, even for very small sample sizes, nonparametric approaches dominate parametric approaches with respect to in- and out-of-sample measures of fit and predictive ability, respectively. Fourth, all results also hold for an extended sample of countries.

5.2 Dealing with heterogeneity, nonlinearity and club misclassification in growth convergence: A nonparametric two-step approach

This section is retained from Haupt & Meier (2011).

In the following subsections the method proposed in Chapter 3.4.2 is illustrated with applications to three data sets based on different levels of aggregation — the countries from the Penn World Tables, the prefectures of Japan, and the districts from reunified Germany. These applications allow replication of our method and previous results in a wide sense. We use different levels of aggregation because we expect different levels of heterogeneity. Regions on district level come with similar technology and thus regions on this level converge to a similar or even the same steady state. This is the reason why the concept of absolute convergence is generally used for disaggregated data. However, different countries behave much more heterogeneously, because there are highly differing levels of technology. This is the reason why countries typically converge to different steady states. Classical convergence analysis captures this problem by extending (2.23) with additional covariates (e.g. investment rate, human capital) determining different steady states (see Sala-i Martin, 1996b). In our approach we use the concept of absolute convergence for all levels of aggregation because we allow for different steady states on club level by capturing the club variable. Independently from the aggregation level members of one club are assumed to offer homogenous convergence behavior and thus, we can assume similar steady states in a club. With respect to the aggregation level our empirical results reveal considerable differences in nonlinearity and heterogeneity, while we do not find clear evidence on the sensitivity of results with respect to the ordering rules discussed above[14].

5.2.1 Penn World Table country-level data

Using our two-step procedure we analyze convergence for Penn World Table (PWT) data of 152 countries over the years from 1970 to 2003. As global convergence is clearly rejected (p-value ≈ 0), the clubbing algorithm is applied. Table 1 displays parameter estimates and standard deviations before and after club merging for ordering rules (I)-(IV).

Final ordering (I) offers seven convergence clubs and no diverging countries[15] while one third of the countries are members of the first club. After merging six clubs remain. Using the other ordering rules we get different results. Average ordering (II) produces basically nine convergence clubs, but using club merging the number of clubs can be reduced to six clubs and the divergence group and similarly to final ordering, the first club is the biggest one and consists of 67 countries, while the other clubs are much smaller. The divergence group has six members. Differ-

[14] All computations in this paper are done using the software R, version 2.11.0, and version 0.40-4 of the np-package of Hayfield & Racine (2008). Of course, data and code are available from the authors.

[15] Using the same data, Phillips & Sul (2009) only identify five convergence clubs. For those clubs, however, we find the same parameter estimates and standard deviations.

ence ordering (III) produces only five non-mergeable convergence clubs (and one diverging country), while also 67 of the countries are members of the first convergence club. Decreasing weights ordering (IV) generates seven clubs which persist after merging. About half of the countries belong to the first club. In summary, the composition and number of convergence clubs seems to be highly sensitive with respect to the choice of the ordering rule.

The convergence behavior of the six clubs using final ordering (I) is displayed in Figure B.4, where the relative transition coefficients are plotted against time. A closer look at the respective club members listed in Table A.25) may raise some suspicion. For example Club 1 contains the USA and Botswana (e.g., Phillips & Sul, 2009). In absolute values the per capita income of the USA in 1970 is about 17429 US Dollars, compared to 1184 US Dollars in Botswana. Though in absolute values this gap rises considerably until 2003 (see Figure B.5), in relative numbers it decreases over time. While in 1970 the per capita income in Botswana is about 7% of the per capita income in the USA, in 2003 it is about 23%. Botswana also catches up in international comparison with respect to the relative transition coefficients h_{it}. In 1970 the per capita income in Botswana lies at about 80% (USA: 110%) of the cross-country average, while it rises to 91% in 2003 (USA: 105%). Thus, although the absolute incomes between these two countries differ extremely, the countries converge in the sense of Phillips and Sul as the respective h_{it} converge to 1.

In Figure B.6 the box-plots of income in final period are displayed for the six clubs found by final ordering (I). While the incomes inside the clubs are close to each other, the income distribution between the clusters is very heterogeneous. For the same clubs in Figure B.7 we display scatter plots of the log t regressions (B.8). The shape suggested by the trajectories in clubs 1 to 4 is parabolic and convex and thus may be interpreted in a way that in initial periods there are hints for divergence, while over the years we observe convergence because of a positive slope. For club 5 we detect more complex nonlinearities and convergence is assumed because γ is not significantly negative but the regressand decreases at the end of the period, indicating that there is no convergence[16]. Thus, to avoid a misinterpretation of the estimation and test results, the inspection of the log t regression scatter plots seems to be highly recommended.

The clustering algorithm may also be sensitive with respect to the respective time horizon. Thus, for the Penn World Table data we exemplarily analyze for final ordering whether number, size, and composition of clusters is constant for different time periods. We compare the results for the whole time horizon from 1970 to 2003 with consecutively shorter partial time spans, one from 1978 to 2003 and the other from 1986 to 2003. The reason why we choose both periods such that they also end in 2003 is that the income in the final period is the ordering criteria. Using the same final period enables to analyze how the length of the time horizon affects the cluster composition and the number of clusters.

Table A.9 displays numbers, sizes, and compositions of clusters for the complete time horizon 1970 to 2003 and the period from 1978-2003, respectively. The cluster sizes for the complete time horizon is given in the last column containing the row sums, the clusters of the partial period 1978 to 2003 are given in the last row containing the column sums. For the partial horizon we find an additional convergence club and a divergence group. Although the number of clusters changed, their composition is quite stable as countries belonging to the first clubs over the complete horizon predominantly also are members of the first clubs in the partial horizon (and vice versa). As can bee seen from Table A.10, the number of clusters rises to eight and one divergence group when comparing the shorter partial period from 1986 to 2003 to the complete time horizon. Again, though the number of clusters varies over time, the club composition seems to be quite stable over time. This results support the assumption that the club structure can be included as an ordered categorical variable when analyzing β-convergence.

Step one reveals hints for non-robust club sizes and club compositions with respect to ordering rules and time horzion as well as neglected nonlinearities in log t regres-

[16] Note that for other ordering rules and data sets, more clubs exhibit such a behavior.

sions. Both findings raise the question if potential misclassification of convergence clubs will affect estimation and inference within this framework. Thus, in the second step, we analyze the robustness of the club compositions resulting from the differing ordering rules (I)-(IV). For each ordering rule we estimate the parametric model (3.8) and the nonparametric model (3.9) and apply the HLR test. Finally, by running an out-of-sample cross validation analysis we select an optimal ordering rule according to lowest average squared error of prediction (ASEP).

The output for a classical β-convergence regression (2.23) is given in Table A.11. The estimated convergence coefficient is negative, but there is no statistical significance. Advancing to the baseline model (3.8) including the club variable[17] suggests strong evidence for the existence of heterogeneity. The estimated coefficients are displayed in Table A.12 and the resulting club-level regression lines can be seen in Figure B.8[18]. The convergence coefficients are significant for convergence club one to five, but not for the sixth club (which consists of only two countries). The p-values of HLR tests (see column 3 of Table A.13) are approximately equal to zero in all four cases suggesting the application of the nonparametric model (3.9). Table A.13 displays the resulting bandwidth for nonparametric models. For ordering rules (I)-(IV) the estimated bandwidth for continuous regressor is smaller or equal to its standard deviation (1.09), respectively, also indicating a nonlinear influence of the regressor $\log(y_{i,0})$.

Overall, the clubs seem to be well chosen because the bandwidth of the club variable is very small independently from the respective choice of ordering rule. An out-of-sample cross-validation, however, offers a clear ranking for ordering rules with respect to ASEP. The pairwise comparison of the models given in Table A.14 reveals that ordering rule (I) suggested by Phillips and Sul (2009) dominates all other ordering rules for the present data.

[17] We only present results for ordering rule (I) because later on we find that this rule performs best. The results for other ordering rules are similar.

[18] The estimated coefficients for quartile regression are similar to mean regression. Thus, there is no more heterogeneity over the conditional distribution of the regressand.

5.2.2 District-level data from reunified Germany

For this application data on the 439 German administrative districts are taken from the regional data base of the statistical agencies of Germany for per capita income measured as the GDP divided by the corresponding number of citizens for the years 1996 to 2005. The log t regression for German regional data suggests clear evidence against global convergence (p-value ≈ 0), but we are able to find the clubs summarized in Table A.15 for the four ordering rules and classification before merging and after merging. Again, difference ordering produces fewest number of clubs, only eight before and five after merging. Average and decreasing weights ordering reveal highest number of clubs, 24 before and ten respectively eight after merging. Notably, the first two clubs and the divergence group for almost all orderings except difference ordering are very small while for difference ordering the first club includes about 25% of data and the divergence group even one third. Figure B.9 plots the relative transition coefficients over time for the convergence clubs and the diverging group of final ordering. The plots support the convergence hypothesis for the clubs and show diverging behavior of the diverging group. The boxplots in Figure B.10 reveal the heterogeneity (homogeneity) between (within) the clubs.

Analyzing log t regression scatter-plots for regional data offers similar results to the PWT data. Figure B.11 exemplifies the results for final ordering. Most of the convergence clubs offer a parabolic and convex shape which means a nonlinear relationship but no harm for convergence interpretation. But, for the second club the regressand becomes smaller in the last period which rises doubt on the club convergence

Investigating β-convergence yields to the regression output displayed in Table A.16. The estimated coefficient is significantly negative. The estimated coefficients of the baseline model (3.8) briefed in Table A.17 offer β-convergence for all clubs, but divergence for the divergence group. The estimated regression lines for the ten convergence clubs after merging are displayed in Figure B.12. As the p-values in Table A.18) reveal, for ordering rules (I) and (II) the hypothesis of correct parametric specification of the baseline model (3.8) cannot be rejected at any reasonable significance level, while there are hints for misspecification in (III) and (IV). For assessing the quality of clubbing we estimate nonparametric models for all cases. The estimated bandwidths are displayed in Table A.18. With the exception of (II) the bandwidths for $\log(y_{i,0})$ point to nonlinear influences of the regressor. The bandwidths for the club variable are all close to zero. Thus, club compositions are well chosen for all ordering rules. The out-of-sample cross validation analysis offers a strict ranking of ordering rules (I)$\succ$(IV)$\succ$(II)$\succ$(III), where $\succ$ means that the ordering rule on the left has a lower ASEP than the rule on the right.

5.2.3 Prefecture-level data from Japan

In addition to PWT data on country-level and german regional data on district-level we analyze an in-between — data on 47 Japanese prefectures between 1956 and 1990[19]. The results on merged clubs are displayed in Table A.20. Using final ordering and difference ordering we find three convergence clubs which can be merged to two clubs. Average ordering and decreasing weights ordering propose exactly the same results. There are four convergence clubs and one divergence group consisting of three countries. After merging there are only two clubs and one divergence group. The relative transition coefficients over time for the convergence clubs after merging are exemplarily shown for final ordering in Figure B.13, where club convergence is indicated as the transition coefficients converge to one. Again, heterogeneity between clubs can be observed from Figure B.14. The scatter-plots of $\log t$ regression for final ordering displayed in Figure B.15 show quite different results than for the other examples. For all three clubs the first half of the time horizon show parabolic and convex points. In the second half, the points of clubs one and three stagnate in contrast to convergence assumption. The results of $\log t$ regressions should be handled with care. Analyzing classical β-convergence (2.23)

[19] The data of Barro and Sala-i-Martin (2004) are downloaded from http://www.columbia.edu/∼xs23/data.htm at June 15, 2011.

proposes a positive coefficient which is significant on 10%-level (see Table A.21)[20]. Thus, there are no hints for β-convergence over all prefectures. The estimated convergence coefficients for the baseline model reveal negative signs for both clubs, which are, however, not significantly different from zero (see Table A.22). The estimated regression lines are plotted in Figure B.16.

Investigating parametric misspecification the HLR test has small p-values (with maximum of around 11%) for all ordering rules. For all ordering rules the bandwidth of $club_i$ is approximately 0.01 or even smaller. Thus, the clusters seems to be well chosen for all methods. The bandwidth for $\log(y_{i,0})$ proposes a linear influence of this variable for ordering rule (I) and (III) and a nonlinear influence for (II) and (IV). The p-values and bandwidths for nonparametric regression can be found in Table A.23. The out-of-sample cross-validation offers the following sequence of ordering rules (I)$\succ$(III)$\succ$(II)=(IV).

5.2.4 Conclusion

As classical convergence regressions often fail to account for heterogeneity and nonlinearity and recent contributions are able to address either the one or the other, a simple two-step method is proposed to address both issues. Employing a slightly augmented version of the clubbing algorithm of Phillips and Sul (2007a,b, 2009) in step one, we find (i) considerable sensitivity of results on convergence club structures with respect to different initial data orderings. Further, (ii) visual inspections of log t regression scatter plots reveal that the “convergence interpretation” of the results of such a linear regression should be handled with care. As a second step we propose the use of a nonparametric test and regression which allows to analyze convergence effects on both individual and club level while alleviating potential misclassification in the club formation process using simultaneous smoothing over the club structure.

Three empirical exercises using data on different levels of aggregation, countries

[20] Again, we only present results for initial ordering.

from the Penn World Tables, Japanese prefectures, and districts from reunified Germany, respectively, illustrate the proposed two-step approach. For all applications, we find considerable evidence for club-based heterogeneity in convergence analysis by adding the clubs identified in step one as a categorical covariate. Our nonparametric estimation results suggest that the club composition is well chosen. An out-of-sample analysis reveals that initial ordering rule for starting the club identification algorithm (in step one) proposed by Phillips and Sul performs best.

5.3 Convergence of the high-skilled in German regions: Using panel and cross-section information to identify clubs, spatial patterns, and nonlinearities

This section is a joint work with Harry Haupt and Joachim Schnurbus, see Haupt et al. (2011).

5.3.1 Classical convergence regression analysis of the high-skilled employees in German regions

High-skilled employees are the basis for developing new technologies and economic growth. Lumpy provision of high-skilled labor across German regions may slow-down growth and increase already existing gaps in innovation and productivity. It is thus of obvious interest to study the spatial distribution and spatio-temporal diffusion of high-skilled labor and develop statistical methods to study existence and patterns of eventually occurring convergence and divergence processes. In our study region-specific shares of highly educated employees are used as a proxy for high-skilled labor. More precisely, the dependent variable in our model is the growth rate $\texttt{grschool}_i \stackrel{\text{def}}{=} \log(\texttt{school}_{i,2005}) - \log(\texttt{school}_{i,1996})$, where $\texttt{school}_{i,t}$ represents the share of employees liable for social security insurance in region i ($i = 1, \ldots, 439$) (as a place-of-work) and year t ($t = 1996, \ldots, 2005$), who have at least eleven years of schooling and a degree. Hereafter we will denote this share as the share of high-skilled employees.

Adapting the approach of Barro & Sala-i Martin (1992) our analysis is based on the unconditional β-convergence, where the key explanatory variable is $\texttt{school0}_i \stackrel{\text{def}}{=} \texttt{school}_{i,1996}$, the share of high-skilled employees in region i in the year 1996. Figure B.17 provides a first impression of the spatial distribution of the key variables `grschool` and $\texttt{school}_0$. Both maps reveal obvious patterns due to the former separation of Germany in Federal Republic of Germany (FRG) and German Democratic Republic (GDR), hereafter denoted as west and east. To reflect this structural information the binary variable `west` — which is equal to one for west regions and zero for east regions — is included in all subsequent analyses. Note that we primarily consider `west` as a political variable, although it is of obvious economic and as a consequence spatial — due to spill-over effects — relevance, too. Interestingly, the share of high-skilled employees in east regions in 1996 seems to be somewhat higher on average compared to the majority of west regions. In sharp contrast the growth-rate (between 1996 and 2005) is higher on average for most of the west regions compared to the east regions, where some of the latter even experienced negative growth-rates. This phenomenon, often denoted as the post-reunion brain-drain, is obviously still in progress many years after the official reunion in 1990.

Our baseline model[21] is the classical convergence regression model proposed by Mankiw et al. (1992), hereafter MRW, where in the light of our considerations above we allow for specific convergence parameters of west and east regions, re-

[21] Note that the estimation of equation (5.6) is based on cross-section data, where only information in the initial and final time period is employed. The a priori selection of $t = 0$ and $t = T$, respectively, may have a crucial impact on the outcome. We will not discuss such sources of non-robustness in this study.

spectively, that is

$$\begin{aligned} \texttt{grschool}_i &= \alpha_1 \texttt{west}_i + \alpha_2(1-\texttt{west}_i) + \beta_1 \log(\texttt{school0}_i)\texttt{west}_i \\ &+ \beta_2 \log(\texttt{school0}_i)(1-\texttt{west}_i) + \varepsilon_i, \end{aligned} \tag{5.6}$$

and β-convergence of west regions is assumed to be present if $\beta_1 < 0$ (an analogous interpretation for east regions applies to β_2). Then skill concentration differences across the regions decrease as regions with a lower concentration of high-skilled employees increase their concentration faster than regions with a higher concentration.

The OLS (ordinary least squares) estimation results for the baseline convergence regression (5.6) are displayed in Table A.29. We will not stress these preliminary results, as the baseline model obviously suffers from lack of economic content and consequently various sources of misspecification are indicated by a battery of tests. For this reason we also do not report adjusted standard errors here. Given this disclaimer, the convergence coefficient is significantly negative for both parts of Germany and the fit, measured as squared correlation of observed and fitted response values (PR^2), is moderate at about 50%. Thus, the results may be interpreted as slightly suggestive in favor of converging shares of high-skilled employees over all administrative regions.

Following the main contributions of among others Barro & Sala-i Martin (1992), Barro et al. (1991), and Mankiw et al. (1992), a plethora of works appear addressing several strands of criticism confronting the baseline Solow model (see e.g., Haupt & Petring, 2011 for a recent survey). In the following exposition we pick up three main points of criticism.

Equation (5.6) can be written compactly as $y_i = \mathbf{x}_i'\beta + \varepsilon_i$. However, let us assume that the true conditional expectation of y_i given all relevant explanatory variables is equal to $g(\mathbf{x}_i, \mathbf{z}_i)$, where g is an unknown function and $\mathbf{z}$ are unobservable explanatory variables. Then the correctly specified model is given by $y_i = g(\mathbf{x}_i, \mathbf{z}_i) + \xi_i$. When estimating the misspecified model (5.6), the error is equal to $\varepsilon_i = g(\mathbf{x}_i, \mathbf{z}_i) - \mathbf{x}_i'\beta + \xi_i$ where $\{\xi_i\}$ is an error process. The three points of criticism we will consider here reflect three potential sources of the specification error $\Delta_i = g(\mathbf{x}_i, \mathbf{z}_i) - \mathbf{x}_i'\beta$. First, neglected heterogeneity due to incorrectly assuming global convergence, while there may coexist clubs with homogeneous convergence behavior and a group of divergent regions (Section 5.3.2). Second, neglected heterogeneity induced by spatial association due to spill-over and repercussion effects between German regions (Section 5.3.3) . Third, misspecification due to neglected nonlinearities in the regression function (Section 5.3.4). Empirical evidence on all three issues is analyzed for the high-skilled employees in German regions. In order to provide a hint for the robustness of our findings we reanalyze the data of the recent exposition of Ertur & Koch (2007) in Section 5.3.5, while Section 5.3.6 concludes.

5.3.2 Heterogeneity due to convergence (and divergence) clubs

One of the main points of criticism confronting the classical convergence regression (5.6) is that there are several forms of neglected heterogeneity causing invalid estimation results (compare e.g., Masanjala & Papageorgiou, 2004; Canarella & Pollard, 2004; Ertur & Koch, 2007; Alfo et al., 2008 and
Haupt & Meier, 2011).

In a series of seminal contributions Phillips & Sul (2003, 2007a,b, 2009) (hereafter PS) build on the ideas of Durlauf & Quah (1999) and suggest that heterogeneity may occur due to individual effects and different technology levels. Considering these effects they propose a dynamic factor model based the time trajectory $\{\mathtt{school}_{i,t}\}_{t=0,\dots,T}$ of each region i. Their convergence concept — which we label as "club convergence" hereafter — is based on the idea that convergence is assumed if all regions have the (approximately) same share of high-skilled employees in the final period T. Hence club convergence is based on panel data in contrast to β-convergence, the latter only relying on cross-sections for the initial and final period 0 and T.

If there is no evidence (from a so-called log t regression test) in favor of global convergence[22], PS introduce a clustering algorithm for identifying convergence clubs empirically. The idea of convergence clubs is that there are groups of countries with common convergence (or divergence) behavior. The algorithm proposes a classification of convergence clubs, while it is not possible to analyze convergence behavior on a club-level in the sense of Mankiw et al. (1992); Barro et al. (1991); Barro & Sala-i Martin (1992). Thus we adopt the proposal of Haupt & Meier (2011) to introduce a club variable in the baseline model (5.6). Applying an augmented form (see Haupt & Meier, 2011) of the clubbing algorithm of PS to German regions yields a discrete covariate $\mathtt{club}_i$ with 11 categories, i.e. 10 convergence clubs and a divergence group.

In a first step we include the convergence clubs in the classical β-convergence analysis using m dummy variables $\mathtt{club}_{i,j}$ representing the categories of the underlying discrete variable,

$$\begin{aligned} \mathtt{grschool}_i &= \alpha_0 \mathtt{west}_i + \alpha_1 \log(\mathtt{school0}_i) \cdot \mathtt{west}_i \\ &+ \sum_{j=1}^{m} \beta_j \mathtt{club}_{i,j} + \sum_{j=1}^{m} \gamma_j \log(\mathtt{school0}_i) \cdot \mathtt{club}_{i,j} + \varepsilon_i. \end{aligned} \tag{5.7}$$

Table A.30 contains the occupation frequencies of all $\mathtt{club}$-categories for both German regions. The clubs 1 and 10 as well as the divergence group are poorly occupied each having a total of less than five observations. For clubs 8, 9, and 10, as well as for divergence group 11 there are no observations for the east regions of Germany. In Section 5.3.4 we will address potential issues of sparsely populated cells. The results for OLS estimation of Equation (5.7) are displayed in Table A.31. The estimated convergence coefficients are significantly negative for club 1 to 9. Thus, for these clubs β-convergence can be assumed. For club 10 and the divergence group 11 there is no significant convergence. Note that we only have three observations (all in west regions) in these categories. We do not find differences in the convergence behavior between west and east regions, as the coefficient of the interaction between $\mathtt{west}_{i,j}$ and $\mathtt{school0}_i$ is not significantly different from

[22] In the present case of high-skilled employees in German regions the corresponding log t regression reveals no evidence in favor of global convergence on any reasonable significance level.

0. The PR^2 is approximately 90% and also Akaike-Information-Criterion (AIC) and Schwarz-Information-Criterion (SIC) suggest a clear superiority in comparison with the baseline model (5.6). The next natural question to ask is whether the latter model is also capable of capturing potential spatial patterns in the data.

5.3.3 Spatial association

Another source of misspecification of classical convergence regressions may be neglected spatial association (e.g., Pfaffermayr, 2009; Moreno & Trehan, 1997; Südekum, 2008). We want to analyze whether our estimation results also suffer from neglected spatial association. In a regression context, spatial association can occur w.r.t. the response variable, w.r.t. the covariates, and w.r.t. the error term. We follow the approach of Ertur & Koch (2007) who propose a spatially augmented version of the classical β-convergence model. The basic idea is that interdependencies of technology and knowledge spillovers are a source of spatial association. Founded in economic theory the model includes the spatially lagged dependent variable and spatially lagged explanatory variables as

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \beta + \mathbf{W} \mathbf{X} \theta + \varepsilon, \tag{5.8}$$

with regressor matrix $\mathbf{X}$, response vector $\mathbf{y}$, error vector ε, a row-standardized $n \times n$-matrix of spatial weights $\mathbf{W}$ with corresponding parameter ρ, as well as the parameter vectors β, and θ. We switch to matrix notation, as this simplifies the subsequent derivations. Model (5.8) is denoted as spatial Durbin model and can be consistently estimated by Maximum Likelihood. The Durbin model nests the spatial lag model, which is obtained for $\theta = \mathbf{0}$ as

$$\mathbf{y} = \rho \mathbf{W} \mathbf{y} + \mathbf{X} \beta + \varepsilon, \tag{5.9}$$

and the spatial error model (for $\theta = -\rho\beta$) as

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\beta + \mathbf{u}, \\ \mathbf{u} &= \lambda \mathbf{W} \mathbf{u} + \varepsilon. \end{aligned} \tag{5.10}$$

A comprehensive summary of recent contributions concerning these models can be found in Elhorst (2010). We check for spatial association using model selection criteria and tests. For the latter we estimate the baseline regression (5.6) and the club convergence regression (5.7) without spatial effects and check the respective residuals for spatial influences using the Lagrange Multiplier tests discussed in Anselin (1988b) and Anselin et al. (1996). The tests compare the non-spatial models against a spatial lag alternative, a spatial error alternative or a combination of both, respectively. The results of the LM-tests in the upper panel of Table A.33 reveal clear evidence for neglected spatial effects in the baseline model (5.6). The test results for the club convergence regression (5.7) displayed in the lower panel suggest otherwise, however. There only seems to be some weak evidence in favor of the spatial lag alternative with a p-value of 0.07. Further we compare all models by means of the AIC and the SIC, displayed in Table A.32. Due to considerable differences in the number of parameters of the compared spatial models we follow the SIC in our argumentation because it more heavy penalized the inclusion of additional covariates.

With respect to AIC and for models based on club convergence the best model is the spatial lag model which, however, is only slightly superior to model (5.7). If the club structure is neglected there is clear evidence in favor of the spatial models, especially the spatial Durbin model. With respect to SIC model (5.7) is clearly preferred to all other models. For models without club structure the spatial error model performs best, while the model without spatial effects and the spatial lag model perform equally. Obviously there is a relation between club-based heterogeneity and spatial association. However, models including the club structure seem to be capable of capturing spatial associations.

5.3.4 Misspecification of parametric functional form

Several authors identify neglected nonlinearities as a source of invalidity of classical convergence analysis (e.g. Kalaitzidakis et al., 2001; Liu & Stengos, 1999;

Maasoumi et al., 2007; Quah, 1993a, 1997; Henderson, 2010 and Haupt & Petring, 2011). Following the proposal of Haupt & Meier (2011) we address this issue by employing a fully nonparametric approach.

The nonparametric convergence regression model

$$\texttt{grschool}_i = f\left(\log(\texttt{school0}_i), \texttt{club}_i, \texttt{west}_i\right) + \varepsilon_i, \tag{5.11}$$

allows for nonlinearities and interactions among all covariates within the regression function $f(\cdot)$. In the previous section the club membership is shown to sufficiently reflect the spatial association. Hence we include this information also as ordered discrete covariate `club` in the nonparametric regression. For the present problem we have a mix of continuous and discrete covariates. We apply the nonparametric mixed kernel regression approach of Li and Racine (compare Li & Racine, 2004, 2007; Racine & Li, 2004). Recently, Haupt & Petring (2011) found the superior performance of this approach (compared to parametric regression function specifications) in the context of growth regressions for the original data of MRW. The corresponding minimization calculus for a local linear mixed kernel regression is

$$\min_{\tilde{\alpha}(\mathbf{x}_0), \tilde{\beta}(\mathbf{x}_0)} \sum_{i=1}^{n} \Big(\texttt{grschool}_i - \tilde{\alpha}(\mathbf{x}_0) - \tilde{\beta}(\mathbf{x}_0) \cdot (\log(\texttt{school0}_i) - \log(\texttt{school0}_0)) \Big)^2 \cdot K(\mathbf{x}_0, \mathbf{x}_i, \mathbf{b}). \tag{5.12}$$

The vector $\mathbf{x}_i = (\log(\texttt{school0}_i), \texttt{club}_i, \texttt{west}_i)'$ contains the covariate values of the ith administrative region. Analogously, $\mathbf{x}_0$ refers to the covariate position $(\log(\texttt{school0}_0), \texttt{club}_0, \texttt{west}_0)$ where the regression function is estimated locally. The estimated mean regression effect at this covariate position is denoted by $\hat{\alpha}(\mathbf{x}_0)$ while the corresponding estimated first partial derivative w.r.t. $\log(\texttt{school0})$ is denoted by $\hat{\beta}(\mathbf{x}_0)$. Observations are weighted by the generalized product kernel function $K(\mathbf{x}_0, \mathbf{x}_i, \mathbf{b})$, the product of the kernels of the three covariates.

First, the continuous covariate $\log(\texttt{school0}_i)$ is weighted by a second order Gaussian kernel

$$k_{\texttt{school0}}(\texttt{school0}_0, \texttt{school0}_i, b_{\texttt{school0}}) = \frac{1}{b_{\texttt{school0}}} \phi\left(\frac{\texttt{school0}_i - \texttt{school0}_0}{b_{\texttt{school0}}}\right), \tag{5.13}$$

where $\phi(.)$ is the standard normal density and the smoothing parameter $h_k \in]0,\infty[$. The smoothing parameters are denoted bandwidths in a kernel estimation context. Small bandwidths lead to reasonable weights only for observations i where $|\texttt{school0}_i - \texttt{school0}_0|$ is small, i.e. the number of high-skilled employees ($\texttt{school0}_i$) is close to $\texttt{school0}_0$. Contrary, large bandwidths yield almost equal weights for all observations, thus indicating a linear relationship between $\log(\texttt{school0})$ and `grschool`. Second, unordered categorical covariates such as `west` are weighted by

$$k_{\texttt{west}}(\texttt{west}_0, \texttt{west}_i, b_{\texttt{west}}) = \begin{cases} 1 & \text{for} \quad \texttt{west}_i = \texttt{west}_0, \\ b_{\texttt{west}} & \text{for} \quad \texttt{west}_i \neq \texttt{west}_0, \end{cases} \tag{5.14}$$

as suggested by Li & Racine (2004). Third, the ordered variable `club` — where our reasoning in favor of a natural ordering is based on Figure B.18 — is weighted by

$$k_{\texttt{club}}(\texttt{club}_0, \texttt{club}_i, b_{\texttt{club}}) = b_{\texttt{club}}^{|\texttt{club}_i - \texttt{club}_0|}. \tag{5.15}$$

The bandwidths for both discrete kernels take values in $[0,1]$, where a value of 0 means that the regression function is separately estimated for the observations of different covariate categories, i.e. the so-called frequency approach (see Li & Racine, 2007, chapter 3). For a bandwidth of 1 we obtain equal weights for the observations of all categories of the underlying covariate, which is thus irrelevant.

The bandwidths have to be determined prior to the kernel regression estimation. In a mixed covariate context, data-driven bandwidth estimation is required. We estimate the bandwidths by least-squares cross-validation, compare Li & Racine (2007, chapter 4).

Table A.34 displays the estimated bandwidth values for the covariates. The estimated bandwidth of the continuous covariate is about half as large as the standard deviation of $\log(\texttt{school0})$ (which is 0.4816), thus the model allows for considerable nonlinearity with respect to this covariate indicating that neglected nonlinearity may indeed be a problem of the proposed approach. The estimated bandwidths of the discrete covariates are low. For `west` the total weight of about 6 ($\approx 1/0.1711$) observations of the other category corresponds to the weight of one observation of the corresponding category for an estimation. According to Haupt & Meier (2011), a bandwidth value of close to 0 for the `club` variable indicates that the convergence clubs are well chosen. Thus, as the bandwidth of 0.0027 is close to 0, there might only be small probability of club-misspecification. The PR^2 of the corresponding estimation is 0.901 and thus slightly higher than that of the OLS estimation of equation (5.7). Applying the corresponding test of Hsiao et al. (2007) for parametric misspecification as suggested by Haupt & Meier (2011) we obtain a p-value of 0.048. Hence we can reject the null hypothesis of a correct parametric specification at a 5%-level, indicating that the nonparametric approach seems preferable.

The estimated partial effects w.r.t. $\log(\texttt{school0})$ for the nonparametric mixed kernel approach are obtained as $\hat{\beta}(\mathbf{x}_0)$, compare Equation (5.12). In principle these partial effects could be evaluated for a grid covering the range of $\log(\texttt{school0})$-values for each of the 22 category combinations of the discrete covariates (or more generally for any $\mathbf{x}_0$). But, as the region-structure of Germany is rather fixed, we only focus on the evaluation of the partial effects for the given 439 observed covariate value combinations. The vertical lines indicate the estimation uncertainty and correspond to (asymptotic) confidence intervals. We also added the estimated partial effects from the OLS estimation of Equation (5.7), compare Table A.31. Here we can see a clear difference between parametric and nonparametric estimation only for the clubs 2-6, the partial effects for the other clubs (and divergence group) seem to be reasonably estimated by the parametric specification of Table A.31. For the clubs 2-6, the nonparametrically estimated partial effects are not constant.

The nonparametric mixed kernel approach of Li & Racine (2004) can also partially deal with the issue of poorly occupied category combinations. For demonstrating this we propose a new measure, the “virtual number of observations”, n_v. We have previously seen that a certain estimated bandwidth for a discrete covariate determines the extent of smoothing for this covariate (certainly, up to here this also holds true for continuous covariates), i.e. to what degree are observations of other categories used to estimate the regression function for a certain covariate category. We have seen that roughly 6 observations of East-Germany have the same weight than one observation of West-Germany for the estimation of the West-German regression function and vice versa. n_v of a certain category combination is simply the sum of all observations in all category combinations multiplied by the corresponding discrete kernel weights w.r.t. the similarity of the category combinations to that of interest. Consider e.g. the category combination $\texttt{west} = 1, \texttt{club} = 4$, according to Table A.30 this category combination is occupied by 41 observations. $n_v(\texttt{west} = 1, \texttt{club} = 4)$ is thus equal to 41 plus the observations in $\texttt{west} = 1, \texttt{club} = 3$ or 5 weighted by 0.0027 plus the observations $\texttt{west} = 1, \texttt{club} = 2$ or 6 weighted by 0.0027^2 plus ... plus the observations in all the east categories weighted by 0.1711 times the corresponding similarity to the $\texttt{club} = 4$.

Table A.35 shows that because of the very low bandwidth (that indicates a low probability of club-misspecification) we have only a rather low amount of smoothing over the categories of the discrete covariates, such that categories that were poorly occupied before also are so while estimating the nonparametric regression.

5.3.5 Re-analyzing Penn World Tables data

In a seminal paper Ertur & Koch (2007) develop the theoretical basis to allow for technological interactions between cross-sectional units (i.e. regions, countries ,...) in growth modeling. Their spatially augmented convergence model will be denoted as spatial Solow model hereafter. In order to check the robustness of our empirical findings we re-analyze data from the Penn World Tables (PWT), used by Ertur & Koch (2007) to illustrate the merits of the spatial Solow model. Ertur & Koch (2007) estimate a conditional β-convergence regression model for explaining `gy`, the average growth rate of per capita income over the period 1960 to 1995 for 91 countries. In addition to `lny60`, the log initial income in 1960, the authors employ the log saving rate `lns` and the log growth rate of working-age population `lnngd` as further explanatory variables. Based on estimation and (Moran's I) test results (see the rightmost column in Table III, p. 1051, in Ertur & Koch, 2007, and the middle column in Table A.36) the authors find clear hints for spatial structures and estimate a spatial Durbin model with lagged dependent and explanatory variables. Our results displayed in Table A.37 confirm their results to employ a spatial model with spatial lags and spatial errors. The model selection criteria displayed in the upper part of Table A.38 suggest that the spatial Durbin model (AIC) or the spatial error model (SIC) perform best, while the classical baseline model performs worst. After estimating (unconditional and conditional) spatial Solow models under a homogeneity assumption (i.e. coefficients do not vary across countries), Ertur & Koch (2007) also estimate a local spatial Durbin model to allow for country-specific parameters. They find (visual) evidence for heterogeneities in the model coefficients. In summary Ertur & Koch (2007) identify two sources of heterogeneity: First, due to spatial structures and second, due to individual effects.

Following our argumentation above we start by constructing a club variable for the PWT data. The OLS estimation (and Moran's I test) results for conditional convergence with clubs are outlined in the rightmost column of Table A.36. The LM tests displayed in Table A.37 and the model selection criteria comparisons in Table A.38 indicate a quite good performance of the baseline model including a club variable. The estimated bandwidths of a nonparametric resgression are displayed in Table A.39. Again a bandwidth close to 0 for the club variable indicates a well-chosen club classification. Applying the test of Hsiao et al. (2007) for parametric misspecification we obtain a p-value of 0.391. Hence we can not reject the null hypothesis of correct parametric specification. In analogy to the analysis of the German district data, we find the following: First, we find clear evidence for clubs being an important source of heterogeneity also in the PWT data. Second, after controlling for club effects there is only very weak empirical evidence in favor of spatial structures. Third, the club-level heterogeneity on the one hand is clearly more restrictive than the individual-level heterogeneity employed by Ertur & Koch (2007). Fourth, on the other hand the latter is based on the assumption of global convergence, whereas the former also allows for diverging countries. Fifth, nonparametric mixed kernel regression of the club-level model also allows to estimate flexible club-level effects but for PWT data the parametric specification can not be rejected.

In summary our findings are in line with Ertur & Koch (2007) that the textbook Solow model is misspecified due to neglected heterogeneity, though our modeling approaches slightly differ *for the present data*. This conclusion does not imply that the different approaches address different sources of misspecification. Of course the method proposed by us can be readily employed in the context of the spatial Solow model of Ertur & Koch (2007), whenever economic interest lies in estimation of direct, indirect, and spatial spill-over effects (e.g., Section 6 in Elhorst, 2010).

5.3.6 Conclusion

Applying classical convergence analysis of German high-skilled employees we investigate three potential sources of misspecification: Omitted heterogeneity due to convergence clubs, due to spatial associations between neighboring regions, and due to potential nonlinearities in convergenge behavior. As a first step - to allow for heterogeneities induced by non-global convergence processes - we identify convergence (and divergence) clubs from a dynamic factor model using panel data. In the second step further potential heterogeneities in the extended model are assumed to be generated by spatial associations between regions in a cross-section model. As an encompassing step we test for parametric misspecification of the extended model and check the validity of the club structure generated from panel data to capture heterogeneity of convergence processes in a cross-section model. The employed nonparametric estimation method allows to investigate potential club-specific nonlinearities.

The proposed modeling framework is applied to two data problems on different levels of spatial aggregation: Analyzing the unconditional growth convergence of high-skilled employees in German regions and analyzing a conditional growth model for countries from the Penn World Tables. Model selection results suggest that for both data examples there is no clear empirical evidence in favor of including further spatial model components. The residual heterogeneity in classical models can be captured quite good by controlling for the club structure identified in the first step of our analysis. If, however, the club information is neglected, model selection criteria and tests suggest the existence of spatial association in the model. Tests for parametric misspecification and visual inspection of estimated partial effects reveal some but not clear evidence for nonlinearities.

We stress that our findings do not suggest that there are no spatial externalities, spillovers, or repercussion effects. We just find that the convergence (and divergence) club-level parameters seem to be capable to control for these effects *for the present data sets*.

Bibliography

Alfo, M., Trovato, G., & Waldmann, R. J. (2008). Testing for country heterogeneity in growth models using a finite mixture approach. *Journal of Applied Econometrics*, 23(4), 487–514.

Anselin, L. (1988a). Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. *Geographical Analysis*, 20(1), 1–17.

Anselin, L. (1988b). *Spatial econometrics*. Studies in operational regional science ; 4. Kluwer Acad. Publ.

Anselin, L. (1995). Local indicators of spatial association. *Geographical Analysis*, 27(2), 93–115.

Anselin, L., Bera, A. K., Florax, R., & Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1), 77–104.

Arrow, K. J. (1962). The economic implications of learning by doing. *The Review of Economic Studies*, 29(3), 155–173.

Barro, R. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, 106(2), 407–443.

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

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

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

Baumol, W. (1986). Productivity growth, convergence, and welfare: What the long-run data show. *American Economic Review*, 76(5), 1072–1085.

Benhabib, J. & Spiegel, M. (1994). The role of human capital in econometric development evidence from aggregate cross-country data. *Journal of Monetary Economics*, 34, 143–173.

Bivand, R., Pebesma, E. J., & Gómez-Rubio, V. (2010). *Applied spatial data analysis with R*. New York, NY: Springer.

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*, 1, 49–77.

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

Durlauf, S. N. & Quah, D. (1999). The new empirics of economic growth. In J. B. Taylor & M. Woodford (Eds.), *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics* chapter 4, (pp. 235–308). Elsevier.

Eckey, H. (2001). Der wirtschaftliche entwicklungsstand in den regionen des vereinigten deutschlands. *Diskussion Paper No. 20/01, University of Kassel*.

Eckey, H.-F., Kosfeld, R., & Türck, M. (2007). Regional convergence in germany: a geographically weighted regression approach. *Spatial Economic Analysis*, 2(1), 45–64.

Elhorst, J. (2010). Applied spatial econometrics: Raising the bar. *Spatial Economic Analysis*, 5(1), 9–28.

Ertur, C. & Koch, W. (2007). Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics*, 22(6), 1033–1062.

Evans, P. (1998). Using panel data to evaluate growth theories. *International Economic Review*, 39(2), 295–306.

Funke, M. & Niebuhr, A. (2005). Threshold effects and regional economic growth: evidence from west germany. *Economic Modelling*, 22, 61–80.

Granger, C. W., Maasoumi, E., & Racine, J. (2004). A dependence metric for possibly nonlinear processes. *Journal of Time Series Analysis*, 25(5), 649–669.

Hadar, J. & Russell, W. R. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, 59(1), 25–34.

Harvey, A. C. & Collier, P. (1977). Testing for functional misspecification in regression analysis. *Journal of Econometrics*, 6(1), 103–119.

Haupt, H. & Meier, V. (2011). Dealing with heterogeneity, nonlinearity and club misclassification in growth convergence: A nonparametric two-step approach. Working Paper.

Haupt, H., Meier, V., & Schnurbus, J. (2011). Convergence of the high-skilled in german regions: Using panel and cross-section information to identify clubs, spatial patterns, and nonlinearities. Working Paper.

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. (2009). Statistical validation of functional form in multiple regression using r. In H. D. Vinod (Ed.), *Advances in Social Science Research Using R*, number 9 (pp. 157–168). New York: Springer.

Haupt, H., Schnurbus, J., & Tschernig, R. (2010). On nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, 25(5), 894–901.

Hayfield, T. & Racine, J. S. (2008). Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 1–32.

Henderson, D. (2010). A test for multimodality of regression derivatives with application to nonparametric growth regression. *Journal of Applied Econometrics*, 25(3), 458–480.

Herz, B. & Röger, W. (1995). Economic growth and covergence in germany. *Review of World Economics*, 131(1), 132–143.

Hobijn, B. & Franses, P. H. (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, 15, 59–81.

Hsiao, C., Li, Q., & Racine, J. S. (2007). A consistent model specification test with mixed discrete and continuous data. *Journal of Econometrics*, 140(2), 802–826.

Hurvich, C., Simonoff, J., & Tsai, C. (1998). Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society Series B*, 60, 271–293.

International Bank for Reconstruction and Development (1988). *World Tables*. John Hopkins University Press.

Islam, N. (1995). Growth empirics: a panel data approach. *Quarterly Journal of Economics*, 110(4), 1127–1170.

Jorgenson, D., Gollop, F., & Fraumeni, B. (1987). *Productivity and U.S. Economic Growth*. Harvard University Press.

Juessen, F. (2009). A distribution dynamics approach to regional gdp convergence in unified germany. *Empirical Economics*, 37, 627–652.

Juárez, M. A. & Steel, M. F. J. (2010). Model-based clustering of non-gaussian panel data based on skew-t distributions. *Journal of Business & Economic Statistics*, 28(1), 52–66.

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.

Koenker, R. (2005). *Quantile regression*. Econometric Society monographs ; 38. Cambridge Univ. Press.

Koenker, R. & Bassett, G. (1982). Tests of linear hypotheses and l_1 estimation. *Econometrica*, 50(6), 1577–1584.

Koenker, R. W. & Bassett, Gilbert, J. (1978). Regression quantiles. *Econometrica*, 46(1), 33–50.

Kosfeld, R., Eckey, H., & Dreger, C. (2006). Regional productivity and income convergence in the unified germany. *Regional Studies*, 47, 755–767.

Lee, K., Pesaran, M., & Smith, R. (1997). Growth and convergence in a multicountry stochastic solow model. *Journal of Applied Econometrics*, 12(4), 357–392.

Lesage, J. P. & Pace, R. K. (2009). *Introduction to spatial econometrics*. Statistics ; 196. CRC Press.

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, 423–34.

Li, Q. & Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Linton, O., Maasoumi, E., & Whang, Y.-J. (2005). Consistent testing for stochastic dominance: A subsampling approach. *Review of Economic Studies*, 72, 735–765.

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

Lucas, R. J. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3–42.

Maasoumi, E., Li, Q., & Racine, J. (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. H. & Papageorgiou, C. (2004). The solow model with ces technology: nonlinearities and parameter heterogeneity. *Journal of Applied Econometrics*, 19(2), 171–201.

McFadden, D. (1989). Testing for stochastic dominance. In T. Fomby & T. Seo (Eds.), *Studies in the Economics of Uncertainty*. Springer-Verlag.

Mello, M. & Perrelli, R. (2003). Growth equations: a quantile regression exploration. *The Quarterly Review of Economics and Finance*, 43(4), 643–667.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2), 17–23.

Moreno, R. & Trehan, B. (1997). Location and the growth of nations. *Journal of Economic Growth*, 2(4), 399–418.

Niebuhr, A. (2001). *Convergence and the Effects of Spatial Interaction*. Discussion Paper Series 26351, Hamburg Institute of International Economics.

Ord, K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, 70(349), pp. 120–126.

Pfaffermayr, M. (2009). Conditional β- and σ-convergence in space: A maximum likelihood approach. *Regional Science and Urban Economics*, 39(1), 63–78.

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

Phillips, P. C. B. & Sul, D. (2007a). Modeling and econometric convergence tests. *Econometrica*, 75(6), 1771 – 1855.

Phillips, P. C. B. & Sul, D. (2007b). Some empirics on economic growth under heterogeneous technology. Cowles Foundation Paper No. 1210.

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

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

Quah, D. (1993b). Galton's fallacy and tests of the convergence hypothesis. *Scandinavian Journal of Economics*, 95(4), 427–443.

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

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

Ram, R. (2008). Parametric variability in cross-country growth regressions: An application of quantile-regression methodology. *Economics Letters*, 99(2), 387–389.

Ramsey, J. (1969). Tests for specification error in classical linear least squares regression analysis. *Journal of the Royal Statistical Society Series B*, 31, 350–371.

Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5), 1002–1037.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388), 871–880.

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

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

Sala-i Martin, X. X. (1997). I just ran 2 million regressions. *American Economic Review*, 87(2), 178–183.

Südekum, J. (2008). Convergence of the skill composition across german regions. *Regional Science and Urban Economics*, 38(2), 148–159.

Seitz, H. (1995). Konvergenz: Theoretische aspekte und empirische befunde für westdeutsche regionen. *Konjunkturpolitik*, 41, 168–198.

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

UNESCO (1988). *Statistical Yearbook*. Unesco Publ. Bernan Press.

Utts, J. (1982). The rainbow test for lack of fit in regression. *Communications in Statistics - Theory and Methods*, 11, 2801–2815.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41(3/4), pp. 434–449.

Zaman, A., Rousseeuw, P. J., & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters*, 71(1), 1–8.

A Tables

Table A.1: Samples of Countries used by Mankiw et al. (1992)

Non-Oil Countries: Algeria, Angola, Argentinia, Australia, Austria, Bangladesh, Belgium, Benin, Bolivia, Botswana, Brazil, Burkina Faso, Burma, Burundi, Cameroon, Canada, Central African Republic, Chad, Chile, Colombia, Democratic Republic of Congo, Costa Rica, Denmark, Dominican Republic, Ecuador, Egypt, El Salvador, Ethiopia, Finland, France, Germany, Ghana, Greece, Guatemala, Haiti, Honduras, Hong Kong, India, Indonesia, Ireland, Israel, Italy, Ivory Coast, Jamaica, Japan, Jordan, Kenya, Liberia, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mauritius, Mexico, Morocco, Mozambique, Nepal, Netherlands, New Zealand, Nicaragua, Niger, Nigeria, Norway, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Portugal, Republic of Korea, Rwanda, South Africa, Senegal, Sierra Leone, Singapore, Somalia, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Syrian Arabian Republic, Tanzania, Thailand, Togo, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Kingdom, United States, Uruguay, Venezuela, Zaire, Zambia, Zimbabwe
Intermediate Countries: Algeria, Argentinia, Australia, Austria, Bangladesh, Belgium, Bolivia, Botswana, Brazil, Burma, Cameroon, Canada, Chile, Colombia, Costa Rica, Denmark, Dominican Republic, Ecuador, El Salvador, Ethiopia, Finland, France, Germany, Greece, Guatemala, Haiti, Honduras, Hong Kong, India, Indonesia, Ireland, Israel, Italy, Ivory Coast, Jamaica, Japan, Jordan, Kenya, Madagascar, Malawi, Malaysia, Mali, Mexico, Morocco, Netherlands, New Zealand, Nicaragua, Nigeria, Norway, Pakistan, Panama, Paraguay, Peru, Philippines, Portugal, Rebublic of Korea, South Africa, Senegal, Singapore, Spain, Sri Lanka, Sweden, Switzerland, Syrian Arabian Republic, Tanzania, Thailand, Trinidad and Tobago, Tunisia, Turkey, United Kingdom, United States, Uruguay, Venezuela, Zambia, Zimbabwe
OECD Countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, Turkey, United Kingdom, United States,

Table A.2: Penn World Table data of 152 Countries

Afghanistan, Algeria, Antigua, Argentina, Australia, Austria, Bahamas, Bahrain, Barbados, Belgium, Belize, Benin, Bermuda, Bhutan, Bolivia, Botswana, Brazil, Brunei, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Cape Verde, Central African Republic, Chad, Chile, China, Colombia, Comoros, Costa Rica, Cote d'Ivoire, Cuba, Cyprus, Democratic Republic of Korea, Democratic Republic of Congo, Denmark, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Ethiopia, Federated States of Micronesia, Fiji, Finland, France, Gabon, Germany, Ghana, Greece, Grenada, Guatemala, Guinea, Guinea Bissau, Honduras, Hong Kong, Hungary, Iceland, India, Indonesia Iran, Iraq, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Kenya, Kiribati, Kuwait, Laos, Lesotho, Liberia, Luxembourg, Macao, Madagascar Malawi, Malaysia, Maldives, Mali, Malta, Mauritania, Mauritius, Mexico, Mongolia, Morocco, Mozambique, Namibia, Nepal, Netherlands, Netherlands Antilles, New Zealand, Nicaragua, Niger, Nigeria, Norway, Oman, Pakistan, Panama, Papua New Guinea, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Republic of Congo, Republic of Korea, Romania, Rwanda, Samoa, Sao Tome and Principe, Saudi Arabia, Senegal, Sierra Leone, Singapore, Solomon Islands, Somalia, South Africa, Spain, Sri Lanka, St. Kitts and Nevis, St. Lucia, St. Vincent and the Grenadines, Sudan, Suriname, Swaziland, Sweden, Switzerland, Syria, Taiwan, Tanzania, Thailand, The Gambia, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkey, Uganda, United Arab Emirates, United Kingdom, United States, Uruguay, Vanuatu, Venezuela, Zambia, Zimbabwe

Table A.3: Intermediate aggregation level data: 47 Japanese Prefectures

Aichi, Akita, Aomori, Chiba, Ehime, Fukui, Fukuoka, Fukushima, Gifu, Gumma, Hiroshima, Hokkaido, Hyogo, Ibaraki, Ishikawa, Iwate, Kagawa, Kagoshima, Kanagawa, Kochi Kumamoto, Kyoto, Mie, Miyagi, Miyazaki, Nagano, Nagasaki, Nara, Niigata, Oita, Okayama, Okinawa, Osaka, Saga, Saitama, Shiga, Shimane, Shizuoka, Tochigi, Tokushima, Tokyo, Tottori, Toyama, Wakayama, Yamagata, Yamaguchi, Yamanashi

Table A.4: Missing observations and years used for estimating missing values.

Federal State	**year**	**Missing Observation**	**years used to estimate proportions**
Niedersachsen	1995	Wolfsburg, Gifhorn, Emden, Aurich	1998,1999,2000,2001,2002,2003,2004,2005
	1996	Wolfsburg, Gifhorn, Emden, Aurich	1998,1999,2000,2001,2002,2003,2004,2005
	1997	Wolfsburg, Gifhorn, Emden, Aurich	1998,1999,2000,2001,2002,2003,2004,2005
Nordrhein-Westfalen	1995	Köln, Leverkusen	1996,2002,2004,2005
	1997	Mühlheim, Oberhausen, Aachen, Leverkusen, Bottrop, Münster	1996,2002,2004,2005
	1998	Bonn, Leverkusen, Bottrop, Münster	1996,2002,2004,2005
	1999	Köln, Leverkusen	1996,2002,2004,2005
	2000	Köln, Leverkusen	1996,2002,2004,2005
	2001	Köln, Leverkusen	1996,2002,2004,2005
	2003	Köln, Leverkusen	1996,2002,2004,2005
Rheinland-Pfalz	2002	Neustadt a.d W., Rhein-Pfalz-Kreis	1995,1996,1997,1998,1999,2000,2001,2003,2004,2005
Bayern	2001	Ingolstadt, Freising, Neuburg, Starnberg, Regen, Rottal, Straubing, Dingolfing	1995,1996,1997,1998,1999,2000
	2002	Ingolstadt, Bad Tölz, Garmisch-Patenkirchen, Neuburg, Regen, Rottal, Straubing, Dingolfing	1995,1996,1997,1998,1999,2000
	2003	Ingolstadt, Neuburg, Regen, Rottal, Straubing, Dingolfing, Ansbach, Neustadt	1995,1996,1997,1998,1999,2000
	2004	Ingolstadt, Rosenheim, Berchtesgarden, Neuburg, Regen, Rottal, Straubing, Dingolfing, Ansbach, Neustadt	1995,1996,1997,1998,1999,2000
	2005	Ingolstadt, Rosenheim, Berchtesgarden, Neuburg, Landshut, Regen, Straubing, Dingolfing	1995,1996,1997,1998,1999,2000
Brandenburg	2004	Frankfurt, Uckermark	1995,1996,1997,1998,1999,2000,2001,2002,2003
	2005	Frankfurt, Spree-Neiße-Kreis	1995,1996,1997,1998,1999,2000,2001,2002,2003
Mecklenburg-Vorpommern	1995	Wismar, Mecklenburg	1997,1998,1999,2000,2001,2002,2003,2004,2005
	1996	Stralsund, Wismar	1997,1998,1999,2000,2001,2002,2003,2004,2005
	2003	Rügen, Uecker-Randow	1995,1996,1997,1998,1999,2000,2001,2002,2004,2005
Sachsen	1996	Plauen, Zwickauer Land	1995,1997,1998,1999,2000,2001,2002,2003,2004,2005
Thüringen	2001	Weimar, Eisenach	1995,1996,1997,1998,1999,2000
	2002	Suhl, Eisenach	1995,1996,1997,1998,1999,2000
	2003	Gera, Eisenach	1995,1996,1997,1998,1999,2000
	2004	Weimar, Eisenach	1995,1996,1997,1998,1999,2000
	2005	Kyffhäuserkreis, Eisenach	1995,1996,1997,1998,1999,2000

Table A.5: OLS (MRW, Table I), LTS based, and quartile regression coefficients.

	OLS	ROBUST	$Q_{0.25}$	$Q_{0.50}$	$Q_{0.75}$
(Intercept)	5.3459	5.6938	5.2145	5.7021	7.9470
`ligdp`	1.3176	1.4445	1.3584	1.6318	1.3543
`lpop`	-2.0172	-1.9716	-1.9795	-2.0903	-1.2347

Table A.6: Median and quartiles of pseudo $\tilde{R}^2$ based on $B = 10,000$ replications.

	Median	(lower; upper)
linear mean approach (5.2)	0.597	(0.582; 0.612)
nonparametric mean approach(5.3)	0.663	(0.650; 0.676)
linear median approach (5.4)	0.595	(0.580; 0.610)
nonparametric median approach (5.5)	0.667	(0.654; 0.681)

Table A.7: P-values for test of hypotheses $H_0 : E[g(y,\hat{y})_{row}] - E[g(y,\hat{y})_{column}] \leq 0$ vs. $H_1 : E[g(y,\hat{y})_{row}] - E[g(y,\hat{y})_{column}] > 0$ based on $B = 10,000$ replications, $g(y,\hat{y})$ =MSEP (upper display), MAEP (central display) and R^2 (lower display).

	approach (5.2)	approach (5.3)	approach (5.4)	approach (5.5)
approach (5.2)	-	0	1	0
approach (5.3)	1	-	1	0.509
approach (5.4)	0	0	-	0
approach (5.5)	1	0.491	1	-
approach (5.2)	-	0	1	0
approach (5.3)	1	-	1	1
approach (5.4)	0	0	-	0
approach (5.5)	1	0	1	-
approach (5.2)	-	1	0	1
approach (5.3)	0	-	0	1
approach (5.4)	1	1	-	1
approach (5.5)	0	0	0	-

Table A.8: Results of clubbing algorithm for PWT data. Club sizes (in brackets), estimates for γ and standard errors of the log t regression (3.6) are displayed for different ordering rules. a) of each ordering rule gives the initial classification before club merging, b) gives the final classification after merging.

(I) Final ordering				**(II) Average ordering**			
a) initial classification		**b) final classification**		**a) initial classification**		**b) final classification**	
	$\hat{\gamma}$ (SE of $\hat{\gamma}$)		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**
Club 1 [50]	0.38 (0.04)	Club 1 [50]	0.38 (0.04)	Club 1 [67]	0.09 (0.03)	Club 1 [67]	0.09 (0.03)
Club 2 [30]	0.24 (0.03)	Club 2 [30]	0.24 (0.03)	Club 2 [8]	0.36 (0.04)	Club 2 [18]	0.03 (0.03)
Club 3 [21]	0.11 (0.03)	Club 3 [21]	0.11 (0.03)	Club 3 [10]	-0.001 (0.02)		
Club 4 [24]	0.13 (0.06)	Club 4 [38]	-0.44 (0.07)	Club 4 [12]	-0.01 (0.06)	Club 3 [12]	-0.01 (0.06)
Club 5 [14]	0.19 (0.11)			Club 5 [21]	0.03 (0.05)	Club 4 [23]	0.04 (0.05)
Club 6 [11]	1.00 (0.17)	Club 5 [11]	1.00 (0.17)	Club 6 [2]	0.10 (0.31)		
Club 7 [2]	-0.47 (0.84)	Club 6 [2]	-0.47 (0.84)	Club 7 [9]	0.07 (0.05)	Club 5 [16]	0.06 (0.10)
				Club 8 [7]	0.15 (0.12)		
				Club 9 [10]	1.39 (0.15)	Club 6 [10]	1.39 (0.15)
				Group 10 [6]	-2.04^{*} (0.02)	Group 7 [6]	-2.04^{*} (0.02)
(III) Difference ordering				**(IV) Decreasing weights ordering**			
a) initial classification		**b) final classification**		**a) initial classification**		**b) final classification**	
	$\hat{\gamma}$ (SE of $\hat{\gamma}$)		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**
Club 1 [67]	-0.003 (0.007)	Club 1 [67]	-0.003 (0.007)	Club 1 [73]	0.01 (0.03)	Club 1 [73]	0.01 (0.03)
Club 2 [32]	0.71 (0.06)	Club 2 [32]	0.71 (0.06)	Club 2 [24]	0.09 (0.02)	Club 2 [24]	0.09 (0.02)
Club 3 [42]	-0.05 (0.05)	Club 3 [42]	-0.05 (0.05)	Club 3 [22]	0.05 (0.05)	Club 3 [31]	-0.05 (0.05)
Club 4 [4]	1.48 (0.09)	Club 4 [4]	1.48 (0.09)	Club 4 [9]	0.08 (0.06)		
Club 5 [6]	0.43 (0.12)	Club 5 [6]	0.43 (0.12)	Club 5 [2]	0.08 (0.19)	Club 4 [2]	0.08 (0.19)
Group 6 [1]		Group 6 [1]		Club 6 [7]	0.15 (0.11)	Club 5 [15]	-0.07 (0.12)
				Club 7 [8]	1.411 (0.18)		
				Group 8 [7]	-1.80^{*} (0.02)	Group 6 [7]	-1.80^{*} (0.02)

Table A.9: Comparison of club number, size, and composition for PWT data and final ordering (I) for different time horizons. The club structure for complete time horizon 1970 to 2003 (partial horizon from 1978 to 2003) is given in rows (columns).

	C1	C2	C3	C4	C5	C6	G7	n_c
C1	49	0	0	0	0	0	1	50
C2	13	16	1	0	0	0	0	30
C3	0	10	8	3	0	0	0	21
C4	0	0	1	23	14	0	0	38
C5	0	0	0	0	4	7	0	11
C6	0	0	0	0	0	0	2	2
n_c	62	26	10	26	18	7	3	152

Table A.10: Comparison of club number, size, and composition for PWT data and final ordering (I) for different time horizons. The club structure for complete time horizon 1970 to 2003 (partial horizon from 1986 to 2003) is given in rows (columns).

	C1	C2	C3	C4	C5	C6	C7	C8	G9	n_c
C1	43	4	2	0	0	0	0	1	0	50
C2	5	10	13	2	0	0	0	0	0	30
C3	0	0	9	9	2	1	0	0	0	21
C4	0	0	0	4	8	22	4	0	0	38
C5	0	0	0	0	0	1	10	0	0	11
C6	0	0	0	0	0	0	0	1	1	2
n_c	48	14	24	15	10	24	14	2	1	152

Table A.11: OLS estimates of classical convergence model (2.23) for PWT data.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.5997	0.4037	1.49	0.1395
$\log(y_{i,0})$	-0.0126	0.0495	-0.25	0.7995

$adj.R^2$=-0.006, AIC=-752.24, n=152

Table A.13: Estimated bandwidths for nonparametric baseline model estimation using a mixed kernel estimation for PWT data and ordering rules (I)-(IV) and p-values for Hsiao-Li-Racine tests.

	bandwidth of $\log(y_{i,0})$	bandwidth of *club*	p-value of HLR test
(I)	1.054	0.006	0.0125
(II)	0.839	0.01	≈0
(III)	0.812	≈0	≈0
(IV)	1.1205	0.028	≈0

Table A.12: OLS estimates of baseline model (3.9) for PWT data.

	Estimate	Std. Error	t value	Pr(>|t|)
Club 1	5.9291	0.3352	17.69	0.0000
Club 2	4.1480	0.5241	7.91	0.0000
Club 3	4.0241	0.6178	6.51	0.0000
Club 4	2.9553	0.5387	5.49	0.0000
Club 5	7.1889	2.3127	3.11	0.0023
Group 6	11.3003	8.4093	1.34	0.1812
Club 1:log($y_{i,0}$)	-0.5566	0.0373	-14.92	0.0000
Club 2:log($y_{i,0}$)	-0.4191	0.0620	-6.75	0.0000
Club 3:log($y_{i,0}$)	-0.4499	0.0780	-5.77	0.0000
Club 4:log($y_{i,0}$)	-0.3899	0.0750	-5.20	0.0000
Club 5:log($y_{i,0}$)	-1.0743	0.3279	-3.28	0.0013
Club 6:log($y_{i,0}$)	-1.7197	1.1323	-1.52	0.1311

$adj.R^2$=0.8486, AIC=-1014.564, N=152

Table A.14: Pairwise comparisons of cross-validation performance. Number equals share of $B = 10,000$ replications in which model in column has smaller ASEP (average squared error of prediction) than model in row for PWT data.

	(I)	(II)	(III)	(IV)
(I)	–	0.17	0.21	0.07
(II)	0.83	–	0.49	0.25
(III)	0.79	0.51	–	0.34
(IV)	0.93	0.75	0.66	–

Table A.15: Results of clubbing algorithm for German district data. Club sizes (in brackets), estimates for γ and standard errors of the log t regression (3.6) are displayed for different ordering rules. a) of each ordering rule gives the initial classification before club merging, b) gives the final classification after merging.

(I) Final ordering				**(II) Average ordering**			
a) initial classification		**b) final classification**		**a) initial classification**		**b) final classification**	
	$\hat{\gamma}$ (SE of $\hat{\gamma}$)		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**
Club 1 [3]	0.84 (0.26)	Club 1 [3]	0.84 (0.26)	Club 1 [3]	0.84 (0.26)	Club 1 [3]	0.84 (0.26)
Club 2 [5]	0.01 (0.05)	Club 2 [5]	0.01 (0.05)	Club 2 [3]	0.99 (0.30)	Club 2 [3]	0.99 (0.30)
Club 3 [4]	0.19 (0.16)	Club 3 [10]	0.03 (0.14)	Club 3 [4]	0.34 (0.16)	Club 3 [12]	-0.08 (0.12)
Club 4 [6]	0.15 (0.16)			Club 4 [8]	0.11 (0.15)		
Club 5 [30]	0.09 (0.14)	Club 4 [30]	0.09 (0.14)	Club 5 [20]	0.08 (0.14)	Club 4 [33]	0.02 (0.13)
Club 6 [24]	0.15 (0.16)	Club 5 [72]	-0.14 (0.11)	Club 6 [13]	0.13 (0.16)		
Club 7 [14]	0.12 (0.16)			Club 7 [26]	0.08 (0.15)	Club 5 [48]	0.01 (0.13)
Club 8 [14]	0.11 (0.15)			Club 8 [8]	0.05 (0.15)		
Club 9 [7]	0.03 (0.14)			Club 9 [14]	0.11 (0.15)		
Club 10 [13]	0.30 (0.15)			Club 10 [33]	0.21 (0.15)	Club 6 [86]	-0.07 (0.12)
Club 11 [16]	0.20 (0.16)	Club 6 [76]	-0.07 (0.12)	Club 11 [6]	1.42 (0.19)		
Club 12 [33]	0.11 (0.15)			Club 12 [17]	0.39 (0.17)		
Club 13 [27]	0.16 (0.16)			Club 13 [12]	0.07 (0.15)		
Club 14 [90]	0.10 (0.15)	Club 7 [90]	0.01 (0.15)	Club 14 [18]	0.11 (0.16)		
Club 15 [80]	0.15 (0.14)	Club 8 [80]	0.15 (0.14)	Club 15 [85]	-0.05 (0.13)	Club 7 [134]	-0.13 (0.11)
Club 16 [56]	0.04 (0.11)	Club 9 [56]	0.04 (0.11)	Club 16 [2]	0.66 (1.97)		
Club 17 [13]	0.09 (0.12)	Club 10 [13]	0.09 (0.12)	Club 17 [30]	0.06 (0.15)		
Group 18 [4]	-1.33* (0.03)	Group 11 [4]	-1.33* (0.03)	Club 18 [17]	0.05 (0.15)		
				Club 19 [18]	0.04 (0.15)	Club 8 [85]	-0.18 (0.11)
				Club 20 [26]	0.04 (0.14)		
				Club 21 [25]	0.03 (0.14)		
				Club 22 [16]	0.11 (0.15)		
				Club 23 [16]	0.56 (0.17)	Club 9 [16]	0.56 (0.17)
				Club 24 [8]	-0.03 (0.12)	Club 10 [8]	-0.03 (0.12)
				Group 25 [11]	-1.39* (0.02)	Group 11 [11]	-1.39* (0.02)

(III) Difference ordering				**(IV) Decreasing Weights ordering**			
a) initial classification		**b) final classification**		**a) initial classification**		**b) final classification**	
	$\hat{\gamma}$ (SE of $\hat{\gamma}$)		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**		**$\hat{\gamma}$ (SE of $\hat{\gamma}$)**
Club 1 [67]	-0.01 (0.01)	Club 1 [114]	-0.11 (0.09)	Club 1 [5]	0.31 (0.09)	Club 1 [5]	0.31 (0.09)
Club 2 [32]	0.66 (0.07)			Club 2 [8]	0.11 (0.15)	Club 2 [8]	0.11 (0.15)
Club 3 [41]	0.67 (0.15)	Club 2 [83]	-0.10 (0.07)	Club 3 [24]	0.03 (0.14)	Club 3 [32]	0.14 (0.37)
Club 4 [11]	0.36 (0.18)			Club 4 [8]	0.18 (0.16)		
Club 5 [27]	0.50 (0.09)			Club 5 [26]	0.08 (0.15)	Club 4 [48]	0.01 (0.13)
Club 6 [34]	-0.07 (0.05)	Club 3 [34]	-0.07 (0.05)	Club 6 [8]	0.05 (0.15)		
Club 7 [24]	0.35 (0.67)	Club 4 [24]	0.35 (0.67)	Club 7 [14]	0.11 (0.15)		
Club 8 [36]	-0.09 (0.12)	Club 5 [36]	-0.09 (0.12)	Club 8 [32]	0.21 (0.15)	Club 5 [94]	-0.14 (0.11)
Group 9 [148]	-1.08* (0.04)	Group 6 [148]	-1.08* (0.04)	Club 9 [14]	0.21 (0.16)		
				Club 10 [4]	1.21 (0.32)		
				Club 11 [4]	0.48 (0.12)		
				Club 12 [6]	0.34 (0.17)		
				Club 13 [12]	0.07 (0.15)		
				Club 14 [22]	-0.01 (0.14)		
				Club 15 [46]	0.38 (0.14)	Club 6 [141]	-0.17 (0.11)
				Club 16 [3]	0.52 (0.83)		
				Club 17 [81]	-0.63 (0.13)		
				Club 18 [3]	3.16 (0.96)		
				Club 19 [5]	2.52 (0.58)		
				Club 20 [3]	0.82 (0.17)		
				Club 21 [53]	-0.03 (0.11)	Club 7 [72]	-0.13 (0.11)
				Club 22 [15]	0.94 (0.25)		
				Club 23 [4]	0.42 (0.24)		
				Club 24 [22]	-0.10 (0.08)	Club 8 [22]	-0.10 (0.08)
				Group 25 [17]	-1.28* (0.03)	Group 9 [17]	-1.28* (0.03)

Table A.16: OLS estimates of classical convergence model (2.23) for German district data.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.9115	0.1337	6.82	0.0000
$\log(y_{i,0})$	-0.0734	0.0135	-5.43	0.0000

$adj.R^2$=0.0632, AIC=-790.441, $N = 439$

Table A.17: OLS estimates of baseline model (3.9) for German district data.

	Estimate	Std. Error	t value	Pr(>|t|)
Club 1	10.0508	1.3900	7.23	0.0000
Club 2	8.6208	1.2689	6.79	0.0000
Club 3	5.4437	0.9146	5.95	0.0000
Club 4	5.2388	0.3163	16.57	0.0000
Club 5	4.8481	0.2306	21.02	0.0000
Club 6	5.2649	0.3290	16.00	0.0000
Club 7	5.4732	0.2214	24.72	0.0000
Club 8	5.5146	0.2706	20.38	0.0000
Club 9	5.4455	0.3927	13.87	0.0000
Club 10	5.2947	0.9082	5.83	0.0000
Group 11	-1.5484	0.2915	-5.31	0.0000
Club 1:log($y_{i,0}$)	-0.8943	0.1284	-6.96	0.0000
Club 2:log($y_{i,0}$)	-0.7807	0.1180	-6.62	0.0000
Club 3:log($y_{i,0}$)	-0.4938	0.0862	-5.73	0.0000
Club 4:log($y_{i,0}$)	-0.4851	0.0303	-16.00	0.0000
Club 5:log($y_{i,0}$)	-0.4589	0.0228	-20.11	0.0000
Club 6:log($y_{i,0}$)	-0.5093	0.0330	-15.41	0.0000
Club 7:log($y_{i,0}$)	-0.5397	0.0226	-23.88	0.0000
Club 8:log($y_{i,0}$)	-0.5519	0.0280	-19.74	0.0000
Club 9:log($y_{i,0}$)	-0.5546	0.0412	-13.46	0.0000
Club 10:log($y_{i,0}$)	-0.5499	0.0960	-5.73	0.0000
Group 11:log($y_{i,0}$)	0.1780	0.0293	6.07	0.0000

$adj.R^2$=0.852, AIC=-1567.249, $N = 439$

Table A.18: Estimated bandwidths for nonparametric baseline model estimation using a mixed kernel estimation for German district data and ordering rules (I)-(IV) and p-values for Hsiao-Li-Racine tests.

	bandwidth of $\log(y_{i,0})$	bandwidth of *club*	p-value of HLR test
(I)	0.133	0.0002	0.61
(II)	16092882	0.007	0.99
(III)	0.121	0.0013	0.01
(IV)	0.176	0.0053	≈ 0

Table A.19: Pairwise comparisons of cross-validation performance. Number equals share of $B = 10,000$ replications in which model in column has smaller ASEP (average squared error of prediction) than model in row for regional data.

	(I)	(II)	(III)	(IV)
(I)	–	0.24	0.18	0.25
(II)	0.76	–	0.16	0.56
(III)	0.82	0.84	–	0.89
(IV)	0.75	0.44	0.11	–

Table A.20: Results of clubbing algorithm for Japanese prefecture-level data. Club sizes (in brackets), estimates for γ and standard errors of the log t regression (3.6) are displayed for different ordering rules. a) of each ordering rule gives the initial classification before club merging, b) gives the final classification after merging.

(I) Final ordering				**(II) Average ordering**			
a) initial classification		**b) final classification**		**a) initial classification**		**b) final classification**	
	$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**		$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**		$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**		$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**
Club 1 [28]	0.09 (0.01)	Club 1 [28]	0.09 (0.01)	Club 1 [17]	0.12 (0.01)	Club 1 [35]	0.01 (0.01)
Club 2 [17]	0.09 (0.01)	Club 2 [19]	0.00 (0.01)	Club 2 [10]	0.10 (0.01)		
Club 3 [2]	0.10 (0.02)			Club 3 [8]	0.05 (0.01)		
				Club 4 [9]	0.18 (0.00)	Club 2 [9]	0.18 (0.00)
				Group 5 [3]	-0.47 (0.01)	Group 3 [3]	-0.47 (0.01)

(III) Difference ordering				**(IV) Decreasing Weights ordering**			
a) initial classification		**b) final classification**		**a) initial classification**		**b) final classification**	
	$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**		$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**		$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**		$\hat{\gamma}$ **(SE of** $\hat{\gamma}$**)**
Club 1 [12]	0.23 (0.01)	Club 1 [29]	0.08 (0.01)	Club 1 [17]	0.12 (0.01)	Club 1 [35]	0.01 (0.01)
Club 2 [17]	0.22 (0.01)			Club 2 [10]	0.10 (0.01)		
Club 3 [18]	0.02 (0.01)	Club 2 [18]	0.02 (0.01)	Club 3 [8]	0.05 (0.01)		
				Club 4 [9]	0.18 (0.00)	Club 2 [9]	0.18 (0.00)
				Group 5 [3]	-0.47 (0.01)	Group 3 [3]	-0.47 (0.01)

Table A.21: OLS estimates of classical convergence model (2.23) for Japanese prefecture-level data.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.1863	0.2885	11.05	0.0000
$\log(y_{i,0})$	0.1037	0.0597	1.74	0.0891

$adj.R^2$=0.042, AIC=21.17, $N = 47$

Table A.22: OLS estimates of baseline model (3.9) for Japanese prefecture-level data.

	Estimate	Std. Error	t value	Pr(>|t|)
Club 1	3.9644	0.3179	12.47	0.0000
Club 2	3.6674	0.8260	4.44	0.0001
Club 1:log($y_{i,0}$)	-0.0273	0.0619	-0.44	0.6615
Club 2:log($y_{i,0}$)	-0.0453	0.1898	-0.24	0.8124

$adj.R^2$=0.2615, AIC=7.5555, $N = 47$

Table A.23: Estimated bandwidths for nonparametric baseline model estimation using a mixed kernel estimation for Japanese prefecture-level data and ordering rules (I)-(IV) and p-values for Hsiao-Li-Racine tests.

	bandwidth of $\log(y_{i,0})$	bandwidth of *club*	p-value of HLR test
(I)	2721352	0.01	0.04
(II)/(IV)[23]	2.366	0.01	0.11
(III)	1776859	0.003	0.05

Table A.24: Pairwise comparisons of cross-validation performance. Number equals share of $B = 10,000$ replications in which model in column has smaller ASEP (average squared error of prediction) than model in row for Japanese prefecture-level data.

	(I)	(II)/(IV)	(III)
(I)	–	0.43	0.48
(II)/(IV)	0.57	–	0.57
(III)	0.52	0.43	–

Table A.25: Resulting clubs for PWT country level data

Club 1: Antigua, Australia, Austria, Belgium, Bermuda, Botswana, Brunei, Canada, Cape Verde, Chile, China, Cyprus, Denmark, Dominica, Equatorial Guinea, Finland, France, Germany, Hong Kong, Iceland, Ireland, Israel, Italy, Japan, Republic of Korea, Kuwait, Luxembourg, Macao, Malaysia, Maldives, Malta, Mauritius, Netherlands, New Zealand, Norway, Oman, Portugal, Puerto Rico, Qatar, Singapore, Spain, St. Kitts and Nevis, St. Vincent and the Grenadines, Sweden, Switzerland, Taiwan, Thailand United Arab Emirates, United Kingdom, United States	
Club 2: Argentina, Bahamas, Bahrain, Barbados, Belize, Brazil, Colombia, Costa Rica, Dominican Republic, Egypt, Gabon, Greece, Grenada, Hungary, India, Indonesia, Mexico, Netherlands Antilles, Panama, Poland, Saudi Arabia, South Africa, Sri Lanka, St. Lucia, Swaziland, Tonga, Trinidad and Tobago, Tunisia, Turkey, Uruguay	
Club 3: Algeria, Bhutan, Cuba, Ecuador, El Salvador, Fiji, Guatemala, Iran, Jamaica, Lesotho, Federated States of Micronesia, Morocco, Namibia, Pakistan, Papua New Guinea, Paraguay, Peru, Philippines, Romania, Suriname, Venezuela	
Club 4: Benin, Bolivia, Burkina Faso, Cameroon, Cote d'Ivoire, Ethiopia, Ghana, Guinea, Honduras, Jordan, Democratic Republic of Korea, Laos, Mali, Mauritania, Mozambique, Nepal, Nicaragua, Samoa, Solomon Islands, Syria, Tanzania, Uganda, Vanuatu, Zimbabwe, Cambodia, Chad, Comoros, Republic of Congo, The Gambia, Iraq, Kenya, Kiribati, Malawi, Mongolia, Nigeria, Sao Tome and Principe, Senegal, Sudan	
Club 5: Afghanistan, Burundi, Central African Republic, Guinea Bissau, Madagascar, Niger, Rwanda, Sierra Leone, Somalia, Togo, Zambia	
Club 6: Democratic Republic of Congo, Liberia	

Table A.26: Resulting clubs for German district level data

Club 1: Wolfsburg(DFC)[24], Frankfurt am Main(DFC), Schweinfurt(DFC)
Club 2: Düsseldorf(DFC), Ludwigshafen am Rhein(DFC),Stuttgart(DFC), Ingolstadt(DFC), Regensburg(DFC)
Club 3: Mannheim(DFC), München(DFC), Erlangen(DFC), Aschaffenburg(DFC), Darmstadt(DFC), Koblenz(DFC), Ulm(DFC),Passau(DFC), Dingolfing-Landau, Bamberg(DFC)
Club 4: Hamburg(DFC), Bremen(DFC), Köln(DFC), Leverkusen(DFC), Münster(DFC), Offenbach am Main(DFC), Wiesbaden(DFC), Kassel(DFC), Mainz(DFC), Heilbronn(DFC), Baden.Baden(DFC), Karlsruhe(DFC), Heidelberg(DFC), Altötting, Freising, Landshut(DFC), Straubing(DFC), Amberg(DFC), Weiden in der Oberpfalz(DFC), Bayreuth(DFC), Coburg(DFC), Ansbach(DFC), Fürth(DFC), Nürnberg(DFC), Würzburg(DFC), Augsburg(DFC), Kempten im Allgäu(DFC), Memmingen(DFC), Teltow-Fläming, Merseburg-Querfurt
Club 5: Braunschweig(DFC), Salzgitter(DFC), Emden(DFC), Oldenburg(DFC), Osnabrück(DFC), Essen(DFC), Krefeld(DFC), Rhein-Kreis Neuss, Bonn(DFC), Hochtaunuskreis, Main-Taunus-Kreis, Trier(DFC), Main-Taunus-Kreis, Kaiserslautern(DFC), Landau in der Pfalz(DFC), Freiburg im Breisgau(DFC), Biberach, Rosenheim(DFC), Hof(DFC), Schwabach(DFC), Donau-Ries, Wismar(DFC), Dresden(DFC), Jena(DFC), Sömmerda, Kiel(DFC), Vechta, Aachen(DFC), Dortmund(DFC), Offenbach, Böblingen, Rastatt, Bodenseekreis, Ravensburg, Günzburg, Saarbrücken(DFC), Saarpfalz-Kreis, Zwickau(DFC), Eisenach(DFC), Flensburg(DFC), Bremerhaven(DFC), Fulda, Speyer(DFC), Heilbronn, Pforzheim(DFC), Ortenaukreis, Rottweil, Tuttlingen, Pfaffenhofen an der Ilm, Starnberg, Weilheim-Schongau, Cottbus(DFC), Schwerin(DFC), Region Hannover, Wesermarsch, Hersfeld-Rotenburg, Pirmasens(DFC), Hohenlohekreis, Ostalbkreis, Frankfurt Oder(DFC), Wilhelmshaven(DFC), Bielefeld(DFC), Olpe, Zweibrücken, Ludwigsburg, Schwäbisch Hall, Reutlingen, Mühldorf am Inn, Lichtenfels, Main-Spessart, Neu-Ulm, Potsdam(DFC), Ohrekreis

[24] DFC=District-free city

Club 6: Duisburg(DFC), Gütersloh, Bochum(DFC), Lahn-Dill-Kreis, Waldeck-Frankenberg, Esslingen, Bad-Tölz-Wolfratshausen, Rosenheim, Deggendorf, Cham, Neumarkt in der Oberpfalz, Kaufbeuren(DFC), Augsburg, Neubrandenburg(DFC), Dessau(DFC), Erfurt(DFC), Lübeck(DFC), Neumünster(DFC), Pinneberg, Stormarn, Osterode am Harz, Stade, Emsland, Mülheim an der Ruhr(DFC), Remscheid.(DFC), Wuppertal(DFC), Herford, Hagen(DFC), Siegen-Wittgenstein, Bergstraße, Main-Kinzig-Kreis, Gießen, Marburg-Biedenkopf, Main-Tauber-Kreis, Karlsruhe, Schwarzwald-Baar-Kreis, Konstanz, Dachau, Neuburg-Schrobenhausen, Traunstein, Schwandorf, Kronach, Ansbach, Erlangen-Höchstadt, Rhön-Grabfeld, Miltenberg, Ostallgäu, Chemnitz(DFC), Magdeburg(DFC), Nordfriesland, Mönchengladbach(DFC), Mettmann, Minden-Lübbecke, Märkischer Kreis, Wetteraukreis, Frankenthal (Pfalz, DFC), Worms(DFC), Germersheim, Rems-Murr-Kreis, Heidenheim, Freudenstadt, Lörrach, Zollernalbkreis, Miesbach, Landshut, Weißenburg-Gunzenhausen, Aschaffenburg, Kitzingen, Lindau(Bodensee), Oberallgäu, Saarlouis, Greifswald(DFC), Rostock(DFC), Leipzig(DFC), Bitterfeld, Suhl(DFC)	
Club 7: Dithmarschen, Segeberg, Steinburg, Göttingen, Diepholz, Hameln-Pyrmont, Hildesheim, Soltau-Fallingbostel, Verden, Cloppenburg, Oberhausen(DFC), Solingen(DFC), Kleve, Rhein-Erft-Kreis, Euskirchen, Oberbergischer-Kreis, Gelsenkirchen(DFC), Borken, Warendorf, Lippe, Paderborn, Ennepe-Ruhr-Kreis, Hochsauerlandkreis, Soest, Odenwaldkreis, Limburg-Weilburg, Kassel, Schwalm-Eder-Kreis, Werra-Meißner-Kreis, Mayen-Koblenz, Neuwied, Rhein-Hunsrück-Kreis, Westerwaldkreis, Bernkastel-Wittlich, Eifelkreis-Bitburg-Prüm, Neustadt an der Weinstraße, Mainz-Bingen, Göppingen, Neckar-Odenwald-Kreis, Rhein-Neckar-Kreis, Enzkreis, Waldshut, Tübingen, Alb-Donau-Kreis, Sigmaringen, Berchtesgadener-Land, Ebersberg, Eichstätt, Erding, Garmisch-Partenkirchen, Landsberg am Lech, Kelheim, Passau, Regen, Rottal-Inn, Tirschenreuth, Hof, Kulmbach, Wunsiedel im Fichtelgebirge, Nürnberger-Land, Roth, Bad-Kissingen, Haßberge, Würzburg, Aichach-Friedberg, Dillingen an der Donau, Unterallgäu, St. Wendel, Berlin(DFC), Brandenburg an der Havel, Oder-Spree, Uckermark, Stralsund(DFC), Annaberg, Chemnitzer Land, Freiberg, Riesa-Großenhain, Döbeln, Bernburg, Halle (Saale, DFC), Aschersleben-Staßfurt, Jerichower Land, Wernigerode, Altmarkkreis Salzwedel, Gera(DFC), Wartburgkreis, Schmalkalden-Meiningen, Gotha, Sonneberg, Saale-Orla	

Club 8: Ostholstein, Rendsburg-Eckernförde, Goslar, Northeim, Holzminden, Nienburg Weser, Celle, Lüchow-Dannenberg, Lüneburg, Rotenburg Wümme, Uelzen, Delmenhorst(DFC,)Ammerland, Friesland, Grafschaft Bentheim, Leer, Osnabrück, Viersen, Wesel, Aachen, Düren, Rheinisch-Bergischer Kreis, Rhein-Sieg-Kreis, Coesfeld, Recklinghausen, Steinfurt, Höxter, Hamm(DFC), Unna, Darmstadt-Dieburg, Rheingau-Taunus-Kreis, Vogelsbergkreis, Altenkirchen (Westerwald), Bad Kreuznach, Birkenfeld, Cochem Zell, Rhein-Lahn-Kreis, Vulkaneifel, Alzey-Worms, Donnersbergkreis, Südliche Weinstraße, Calw, Breisgau Hochschwarzwald, Emmendingen, Fürstenfeldbruck, Freyung-Grafenau, Straubing-Bogen, Amberg-Sulzbach, Neustadt an der Waldnaab, Regensburg, Bamberg, Coburg, Forchheim, Fürth, Neustadt an der Aisch-Bad Windsheim, Merzig-Wadern, Neunkirchen, Dahme (Spreewald), Oberhavel, Oberspreewald-Lausitz, Ostprignitz-Ruppin, Prignitz, Demmin, Müritz, Rügen, Plauen(DFC), Mittweida, Stollberg, Bautzen, Meißen, Kamenz, Torgau-Oschatz, Wittenberg, Weißenfels, Bördekreis, Weimar, Eichsfeld, Hildburghausen, Ilm-Kreis, Saalfeld-Rudolstadt	
Club 9: Herzogtum Lauenburg, Plön, Schleswig-Flensburg, Helmstedt, Peine, Schaumburg, Cuxhaven, Harburg, Osterholz, Aurich, Oldenburg, Wittmund, Heinsberg, Bottrop(DFC), Herne(DFC), Ahrweiler, Trier-Saarburg, Bad Dürkheim, Kaiserslautern, Kusel, Bayreuth, Schweinfurt, Elbe-Elster, Potsdam-Mittelmark, Spree-Neiße, Bad Doberan, Güstrow, Ludwigslust, Parchim, Vogtlandkreis, Mittlerer Erzgebirgskreis, Aue-Schwarzenberg, Görlitz, Hoyerswerda, Niederschlesischer-Oberlausitzkreis, Löbau-Zittau, Sächsische Schweiz, Weißeritzkreis, Delitzsch, Muldentalkreis, Anhalt-Zerbst, Köthen, Burgenlandkreis, Mansfelder Land, Saalkreis, Sangerhausen, Halberstadt, Stendal, Quedlinburg, Schönebeck, Nordhausen, Unstrut-Hainich-Kreis, Weimarer Land, Saale-Holzland-Kreis, Greiz, Altenburger Land	
Club 10: Gifhorn, Wolfenbüttel, Barnim, Havelland, Märkisch Oderland, Mecklenburg-Strelitz, Nordvorpommern, Nordwestmecklenburg, Ostvorpommern, Uecker-Randow, Zwickauer Land, Leipziger Land, Kyffhäuserkreis	
Group 11: Groß Gerau, Rhein-Pfalz-Kreis, Südwestpfalz, München	

Table A.27: Resulting clubs for Japanese prefecture level data

Club 1: Hokkaido, Miyagi, Fukushima, Niigata, Ibaraki, Tochigi, Gumma, Saitama, Chiba, Tokyo, Kanagawa, Yamanashi, Nagano, Shizuoka, Gifu, Aichi, Mie, Shiga, Kyoto, Osaka, Hyogo, Nara, Hiroshima, Fukuoka, Kumamoto, Oita, Kagoshima, Okinawa
Club 2: Aomori, Iwate, Akita, Yamagata, Toyama, Ishikawa, Fukui, Tottori, Shimane, Okayama, Yamaguchi, Tokushima, Kagawa, Ehime, Saga, Nagasaki, Miyazaki, Wakayama, Kochi

Table A.28: German regional data. Summary statistics of the continuous variables.

Variable	Min.	1.Quart.	2.Quart.	3.Quart.	Max.	Mean	Std.-Dev.
grschool	-0.208	0.136	0.219	0.289	0.801	0.205	0.149
$\texttt{school}_0$	0.019	0.038	0.054	0.080	0.222	0.063	0.033

Table A.29: German regional data. Regression output for Equation (5.6).

	Estimate	Std. Error	t value	Pr(>|t|)
west	0.08816	0.04038	2.183	0.0296
1-west	-0.19617	0.08016	-2.447	0.0148
west:log(school0)	-0.05825	0.01318	-4.421	0.0000
(1-west)east:log(school0)	-0.09214	0.03233	-2.850	0.0046

$PR^2 = 0.504$, $AIC = -725.31$, $SIC = -704.89$

Table A.30: German regional data. Occupation frequency for the category combinations of the discrete covariates.

club	1	2	3	4	5	6	7	8	9	10	11	total
east	1	5	6	18	52	24	6	0	0	0	0	112
west	3	14	18	41	87	86	48	13	11	3	3	327
total	4	19	24	59	139	110	54	13	11	3	3	439

Table A.31: German regional data. Regression output for Equation (5.7).

	Estimate	Std. Error	t value	Pr(>|t|)
Club 1	-0.6634	0.2349	-2.82	0.0050
Club 2	-0.8543	0.0867	-9.86	0.0000
Club 3	-1.0384	0.0876	-11.86	0.0000
Club 4	-0.9782	0.0685	-14.28	0.0000
Club 5	-1.2090	0.0593	-20.39	0.0000
Club 6	-1.2928	0.0760	-17.01	0.0000
Club 7	-1.3670	0.1082	-12.63	0.0000
Club 8	-1.6421	0.3127	-5.25	0.0000
Club 9	-1.7918	0.4597	-3.90	0.0001
Club 10	-0.9817	1.5052	-0.65	0.5146
Group 11	-0.9182	1.3978	-0.66	0.5116
West	0.0525	0.0483	1.09	0.2772
Club 1:log(school0)	-0.4763	0.1379	-3.46	0.0006
Club 2:log(school0)	-0.5105	0.0417	-12.23	0.0000
Club 3:log(school0)	-0.5439	0.0377	-14.43	0.0000
Club 4:log(school0)	-0.4658	0.0279	-16.71	0.0000
Club 5:log(school0)	-0.4973	0.0231	-21.53	0.0000
Club 6:log(school0)	-0.4730	0.0273	-17.32	0.0000
Club 7:log(school0)	-0.4629	0.0354	-13.08	0.0000
Club 8:log(school0)	-0.5194	0.0890	-5.84	0.0000
Club 9:log(school0)	-0.5462	0.1274	-4.29	0.0000
Club 10:log(school0)	-0.2972	0.3975	-0.75	0.4551
Group11:log(school0)	-0.2923	0.3647	-0.80	0.4234
West:log(school0)	0.0071	0.0188	0.38	0.7063

$PR^2 = 0.896$, $AIC = -1372.89$, $SIC = -1270.78$

Table A.32: German regional data. AIC and SIC for baseline, spatial error, spatial lag, and spatial Durbin model with and without convergence club variable

	baseline	spatial error	spatial lag	spatial Durbin
Models without `club` variable				
	baseline	spatial error	spatial lag	spatial Durbin
AIC	-725.31	-734.54	-729.10	-735.36
SIC	-704.90	-710.04	-704.55	-698.60
Models including `club` variable				
	baseline	spatial error	spatial lag	spatial Durbin
AIC	-1373.89	-1372.01	-1374.70	-1370.73
SIC	-1270.78	-1265.81	-1268.51	-1170.59

Table A.33: German regional data. Results of LM-tests for spatial dependencies in the residuals of Equation (5.6) and (5.7)

	Statistic	df	p.value
Test results for Equation (5.6)			
	Statistic	df	p.value
LM-test for spatial error	10.87	1.00	0.00
LM-test for spatial lag	4.30	1.00	0.04
LM-test for spatial error and spatial lag	17.50	2.00	0.00
Test results for Equation (5.7)			
	Statistic	df	p.value
LM-test for spatial error	0.00	1.00	0.98
LM-test for spatial lag	3.20	1.00	0.07
LM-test for spatial error and spatial lag	4.26	2.00	0.12

Table A.34: German regional data. Estimated bandwidths using least-squares cross-validation for nonparametric mixed-kernel regression.

covariate	kernel function	$h_k \in$	$\hat{h}_k$
$\log($`school0`$)$	of Equation (5.13)	$]0,\infty[$	0.2820
`club`	of Equation (5.15)	$[0,1]$	0.0027
`west`	of Equation (5.14)	$[0,1]$	0.1711

Table A.35: German regional data. Virtual number of observations for the category combinations of the discrete covariates according to the estimated bandwidths of the discrete covariates for nonparametric mixed-kernel regression.

`club`	1	2	3	4	5	6	7	8	9	10	11	total
east	1.5	7.4	9.2	25.2	67.1	38.9	14.3	2.3	1.9	0.5	0.5	168.8
west	3.2	14.9	19.2	44.4	96.3	90.5	49.3	13.2	11.0	3.0	3.0	348.0
total	4.7	22.3	28.4	69.6	163.4	129.4	63.6	15.5	12.9	3.5	3.5	516.8

Table A.36: Penn World Tables data. Regression results for conditional convergence analysis without (second column) and including (third column) club variable (p-values in parentheses).

Dep.Var.	model without `club` variable		model with `club` variable	
`Intercept`	0.030	(0.359)	0.131	(0.000)
`lny60`	-0.007	(0.000)	-0.015	(0.000)
`lns`	0.021	(0.000)	0.009	(0.000)
`lnngd`	-0.032	(0.008)	-0.018	(0.011)
`factor(club)2`			-0.056	(0.020)
`factor(club)3`			-0.110	(0.001)
`factor(club)4`			-0.064	(0.009)
`factor(club)5`			-0.078	(0.023)
`factor(club)6`			-0.333	(0.000)
`factor(club)2:lny60`			0.005	(0.074)
`factor(club)3:lny60`			0.010	(0.009)
`factor(club)4:lny60`			0.004	(0.221)
`factor(club)5:lny60`			0.004	(0.351)
`factor(club)6:lny60`			0.034	(0.000)
$Moran's\ I$(W1)	0.230	(0.000)	0.058	(0.151)
$Moran's\ I$ (W2)	0.264	(0.000)	0.087	(0.094)
AIC	-522.15		-635.28	
SIC	-509.60		-597.62	

Table A.37: Penn World Tables data. Results of LM-tests for spatial dependencies in the residuals of conditional convergence analysis (Equation (21), p. 1041 in Ertur & Koch, 2007).

model without `club`-variable			
	Statistic	df	p.value
LM-test for spatial error	2.725	1.00	0.099
LM-test for spatial lag	0.20	1.00	0.652
LM-test for spatial error and spatial lag	11.441	2.00	0.003
model with `club`-variable			
	Statistic	df	p.value
LM-test for spatial error	0.011	1.00	0.915
LM-test for spatial lag	2.012	1.00	0.156
LM-test for spatial error and spatial lag	2.717	2.00	0.257

Table A.38: Penn World Tables data. AIC and SIC for different models estimating conditional convergence.

model without `club` variable				
	baseline	spatial error	spatial lag	spatial Durbin
AIC	-522.15	-532.62	-526.86	-537.56
SIC	-509.60	-517.56	-511.80	-514.96
model including `club` variable				
	baseline	spatial error	spatial lag	spatial Durbin
AIC	-635.28	-634.52	-635.83	-625.17
SIC	-597.62	-594.34	-595.66	-552.36

Table A.39: Penn World Tables data. Estimated bandwidths using least-squares cross-validation for nonparametric mixed-kernel regression.

covariate	kernel function	$h_k \in$	$\hat{h}_k$
`lny60`	of Equation (5.13)	$]0,\infty[$	1.0427
`club`	of Equation (5.15)	$[0,1]$	0.0231
`lns`	of Equation (5.13)	$]0,\infty[$	422,205.1
`lnngd`	of Equation (5.13)	$]0,\infty[$	0.0878

B Figures

Figure B.1: Cross-validation of linear parametric versus nonparametric approaches for conditional mean. Graph displays density estimate of relative MSEP based on $B = 10,000$ sub-sample replications.

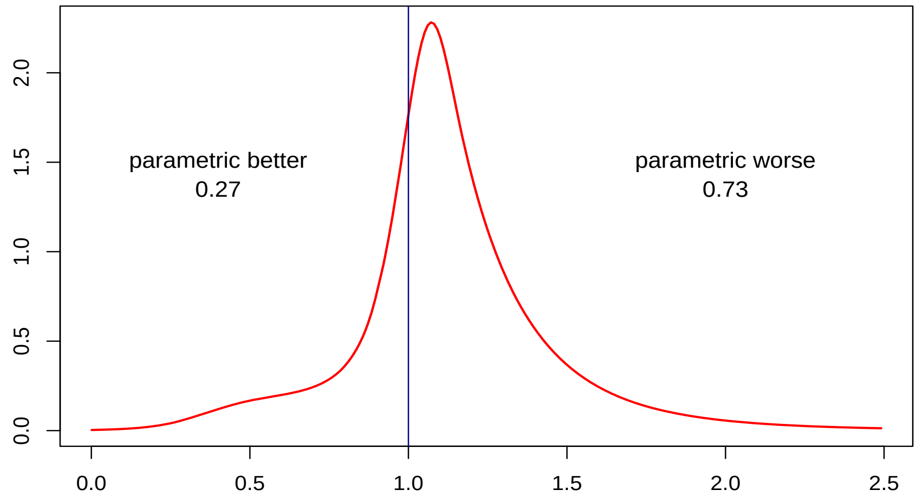


Figure B.2: Estimated manifold of nonparametric mean regression.

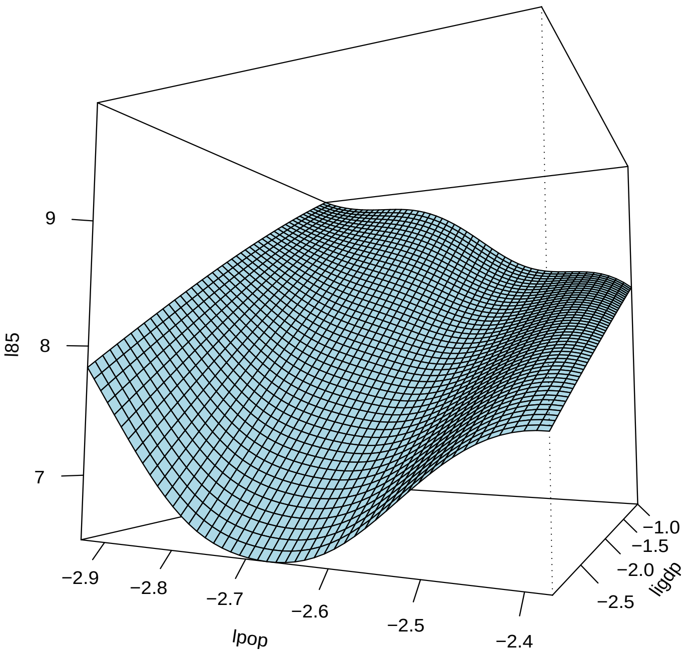
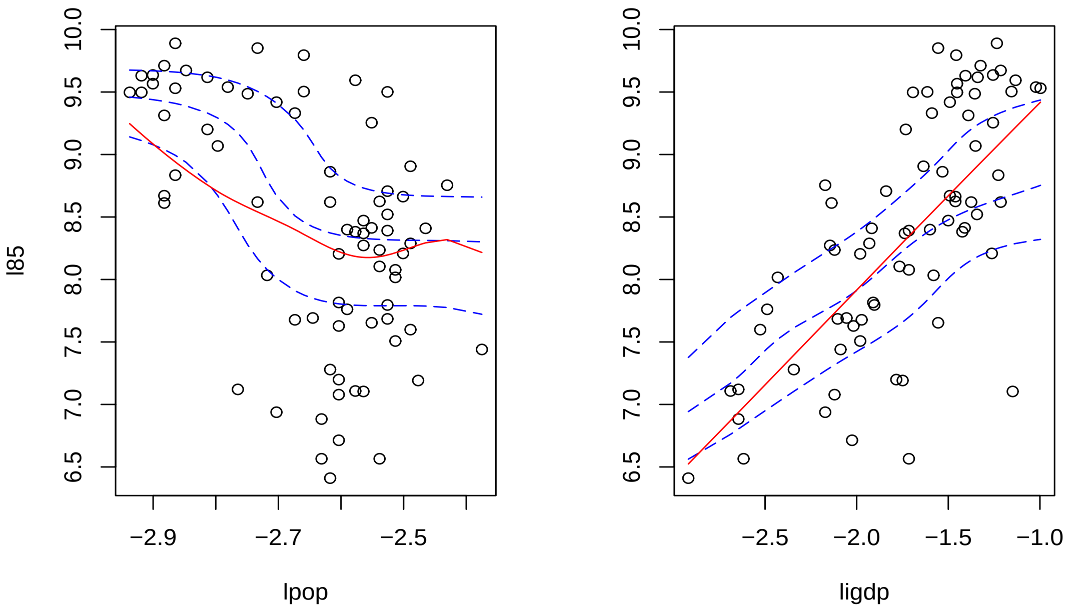


Figure B.3: Estimated conditional partial effects of `lpop` (left panel) and of `ligdp` (right panel). Solid red curve shows nonparametric mean regression, dashed blue curve shows nonparametric quartile regressions.



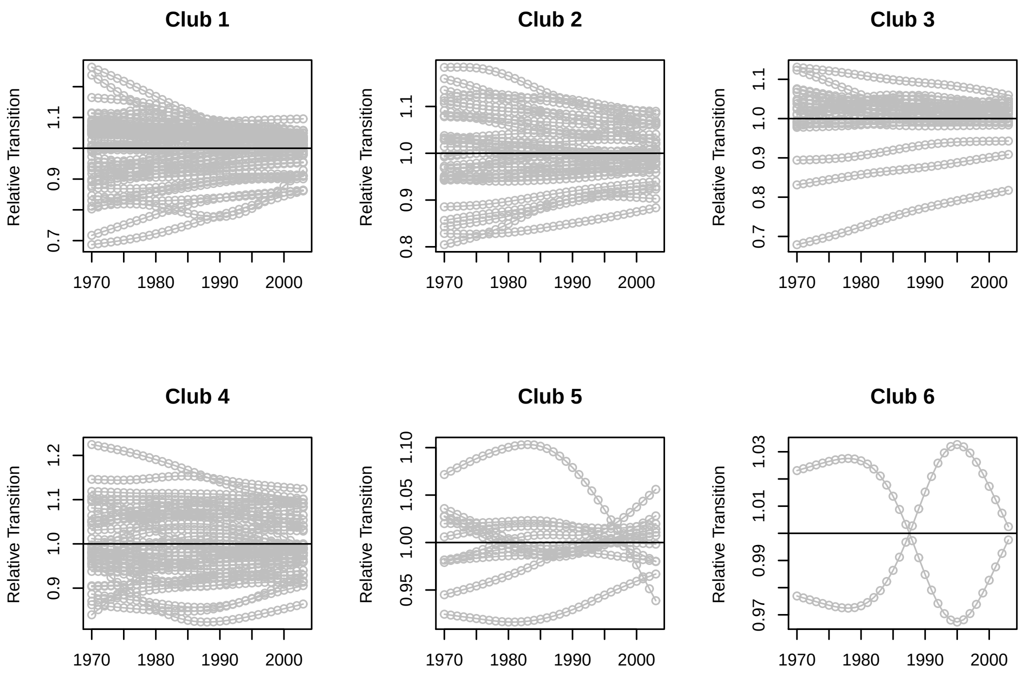


Figure B.4: Relative transition coefficients over time for convergence clubs resulting from final ordering after merging, corresponding to (I) b) in Table A.8.

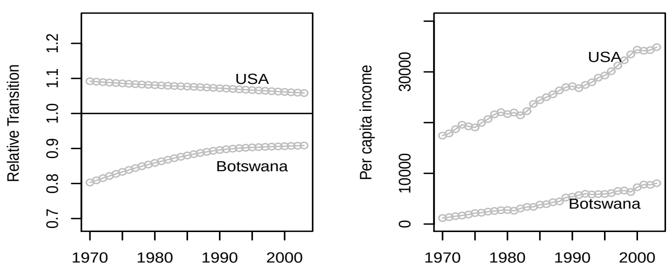


Figure B.5: Relative transition coefficients over time and absolute per capita income over time of USA and Botswana.

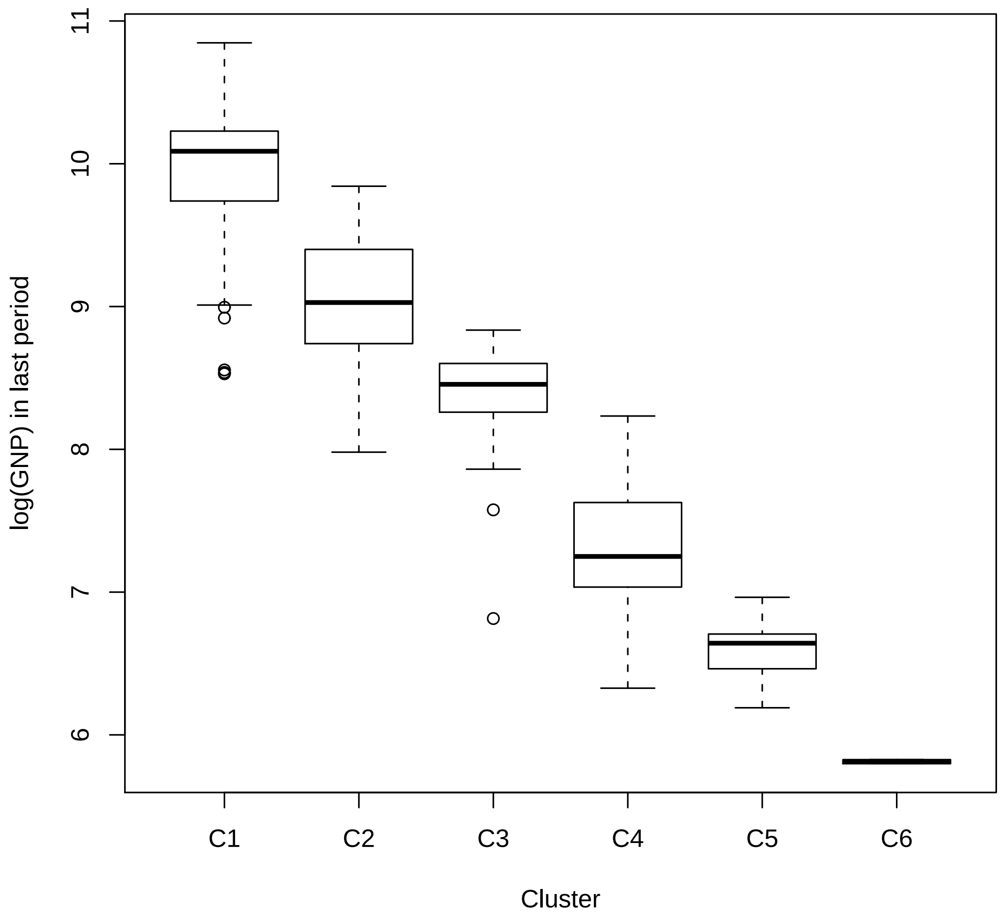


Figure B.6: Boxplots of income in final period divided by the convergence clubs resulting from final ordering after merging, corresponding to (I) b) in Table A.8.

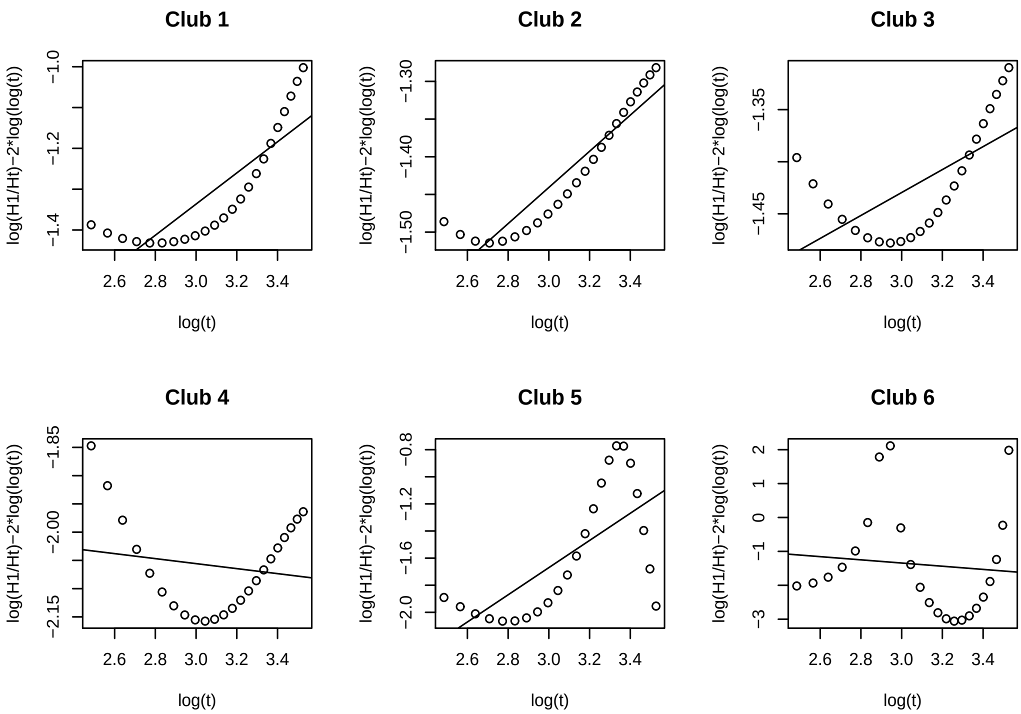


Figure B.7: Scatterplots of the log t regressions (3.6) for clubs found by final ordering after merging, corresponding to (I) b) in Table A.8. Solid line is ordinary least squares estimate.

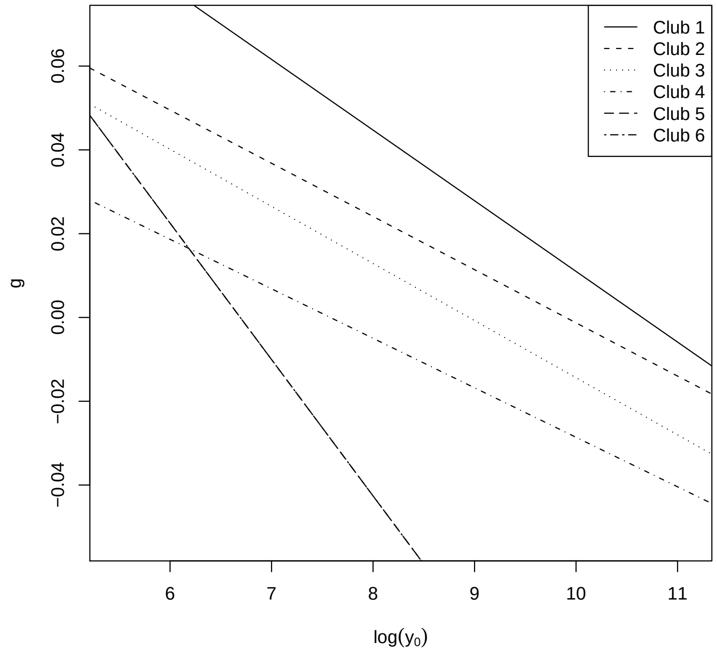


Figure B.8: Estimated regression lines from the estimation displayed in Table par-resultsps for the five convergence clubs for PWT data.

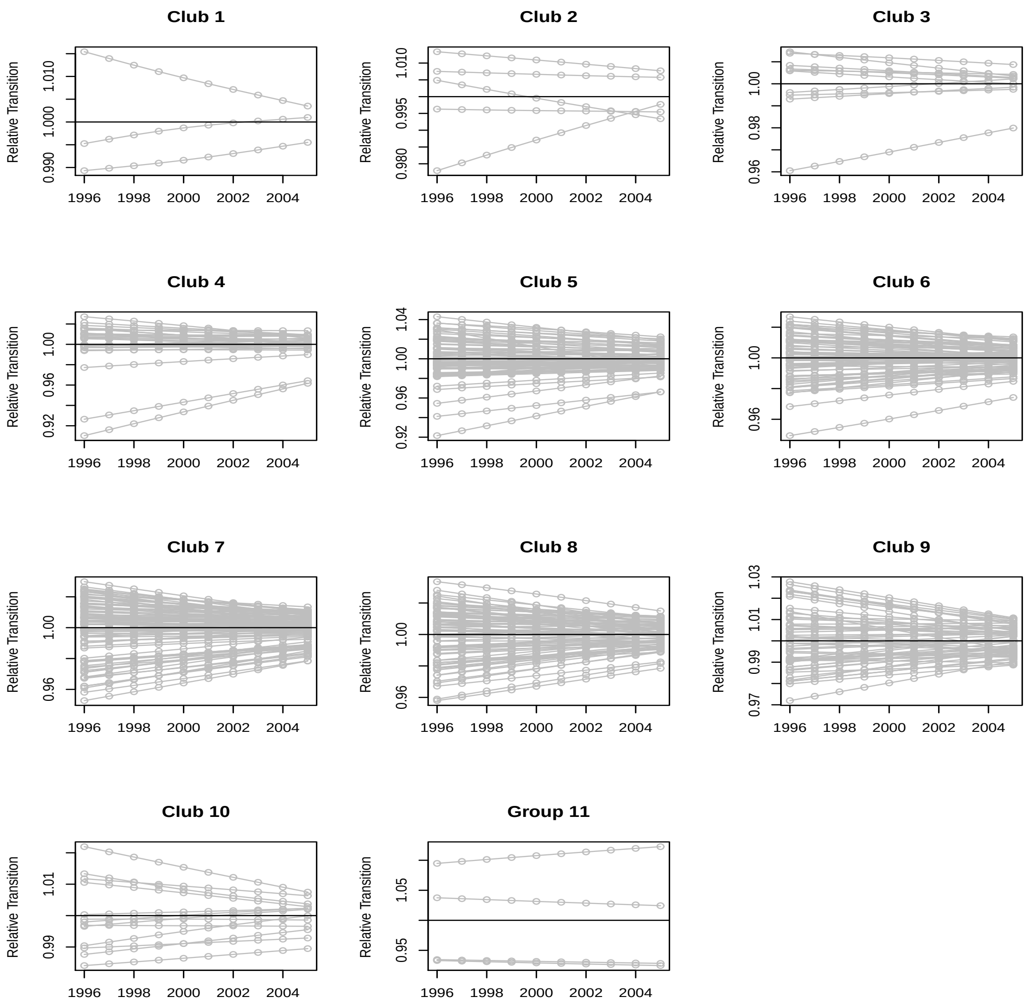


Figure B.9: Relative transition coefficients over time for convergence clubs resulting from final ordering after merging, corresponding to (I) b) in Table A.15.

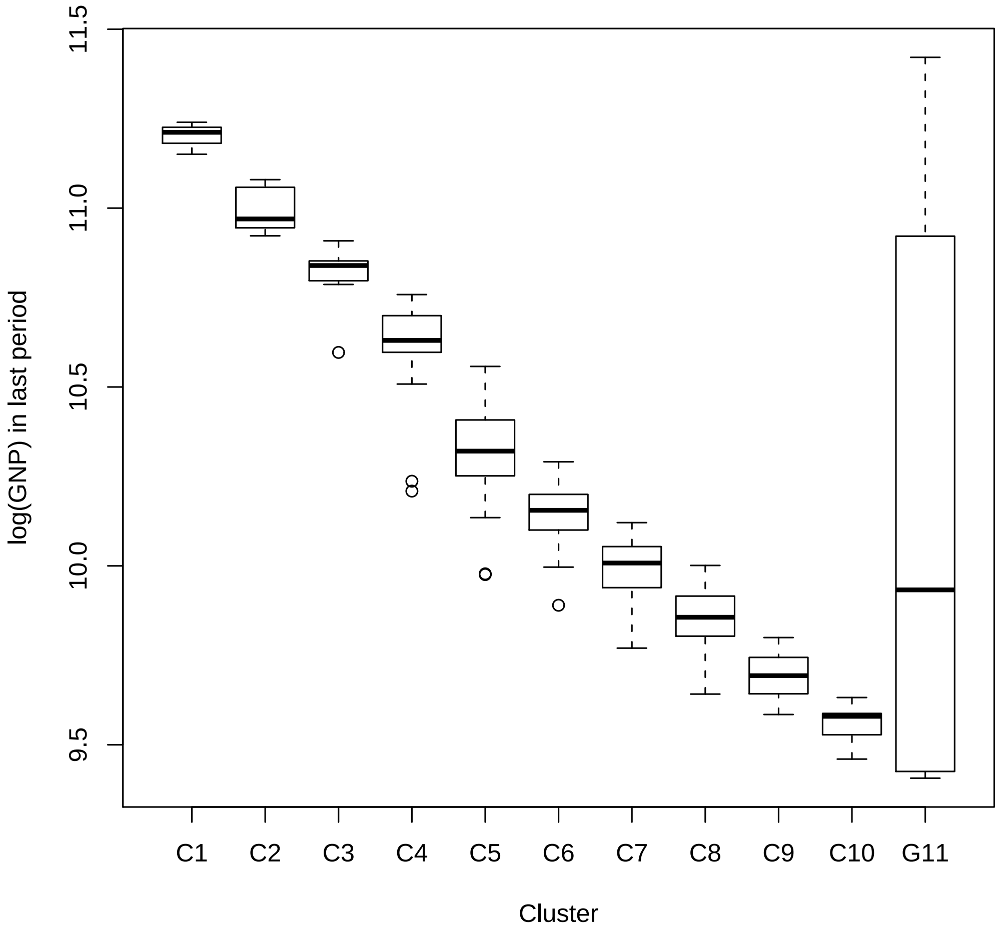


Figure B.10: Boxplots of income in final period divided by the convergence clubs resulting from final ordering after merging, corresponding to (I) b) in Table A.15.

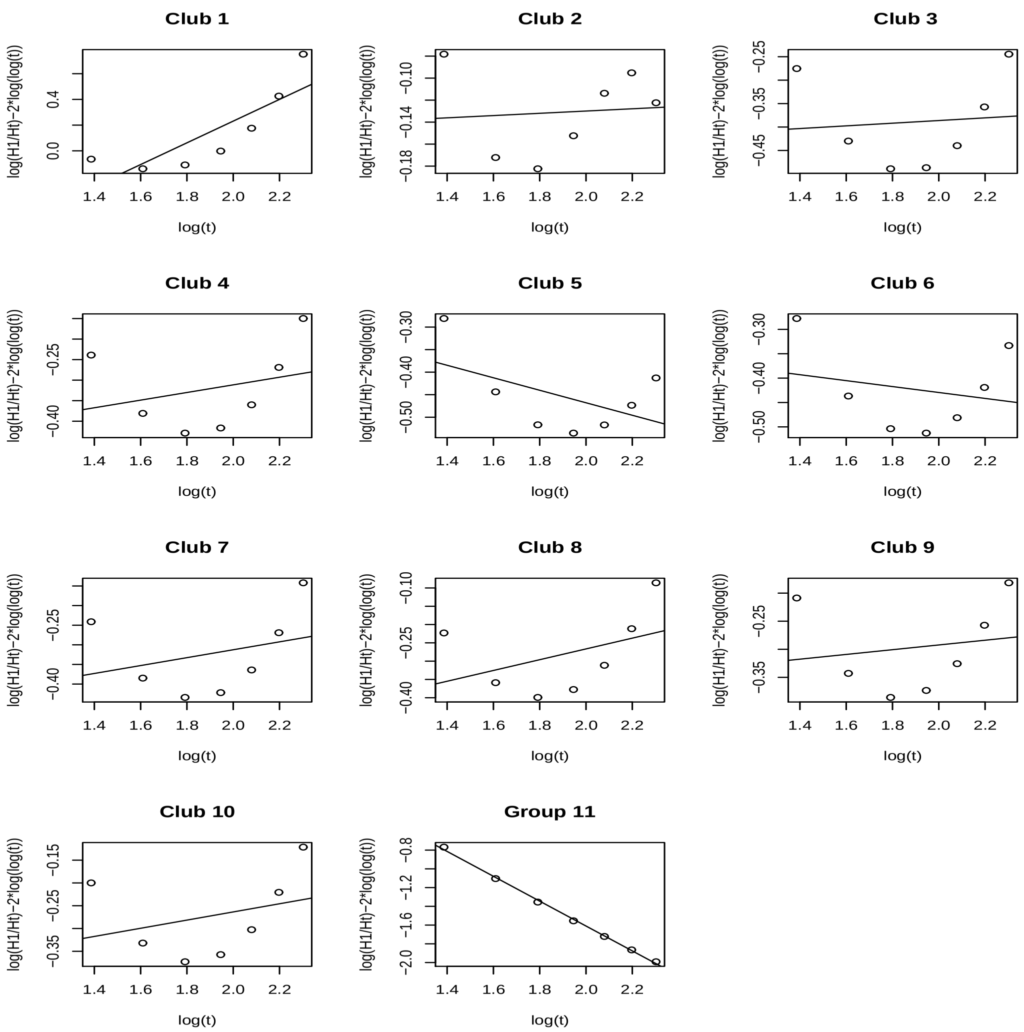


Figure B.11: Scatterplots of log t regression for clubs found by final ordering after merging, corresponding to (I) b) in Table A.15. Solid line is ordinary least squares estimate.

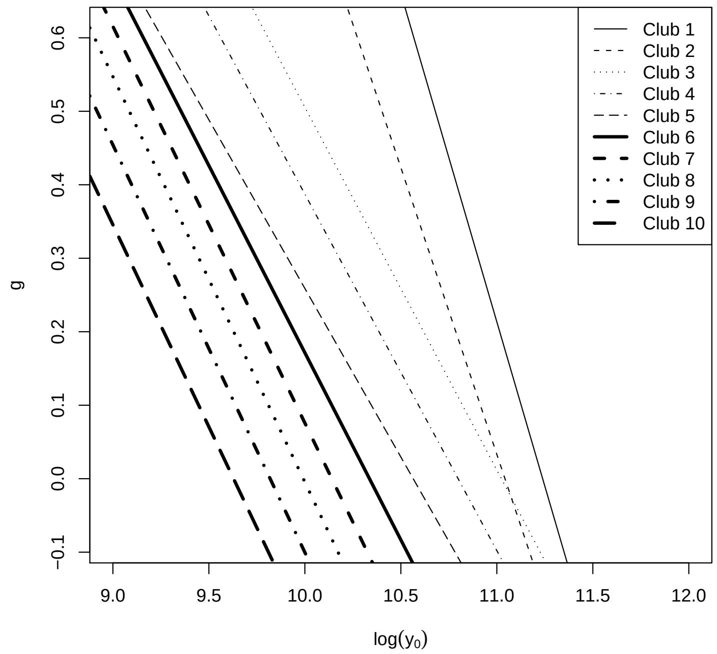


Figure B.12: Estimated regression lines from the estimation displayed in Table A.17 for the ten convergence clubs for German district data.

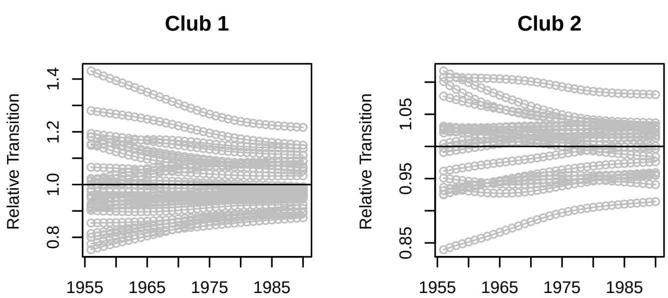


Figure B.13: Relative transition coefficients over time for convergence clubs resulting from final ordering after merging, corresponding to (I) b) in Table A.20.

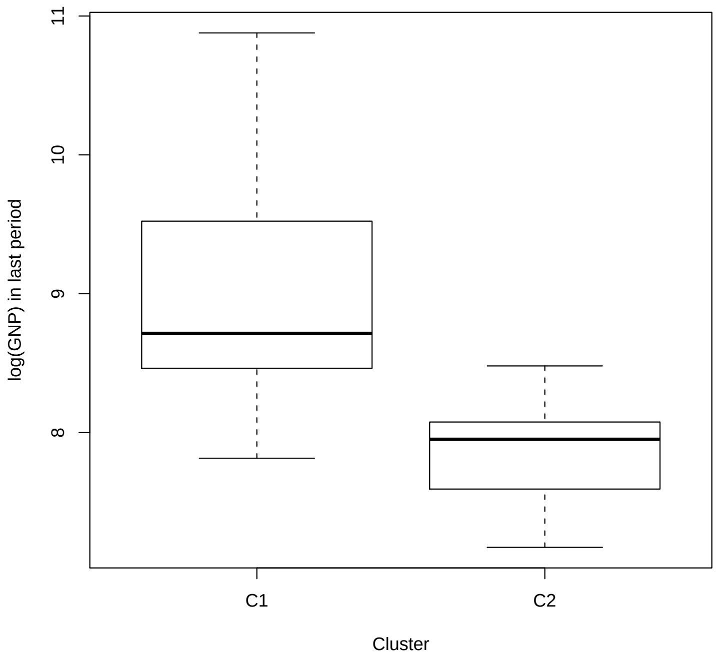


Figure B.14: Boxplots of income in final period divided by the convergence clubs resulting from final ordering after merging, corresponding to (I) b) in Table A.20.

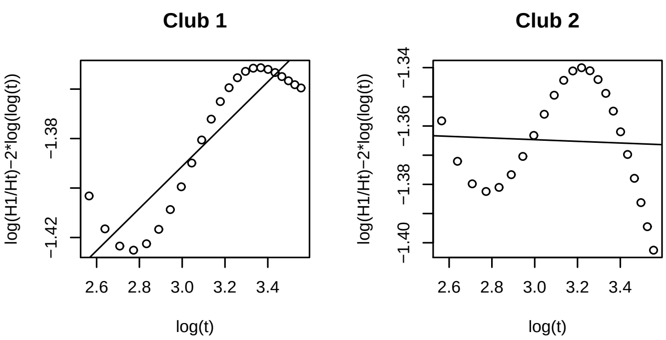


Figure B.15: Scatterplots of log t regressions for clubs found by final ordering after merging, corresponding to (I) b) in Table A.20. Solid line is ordinary least squares estimate.

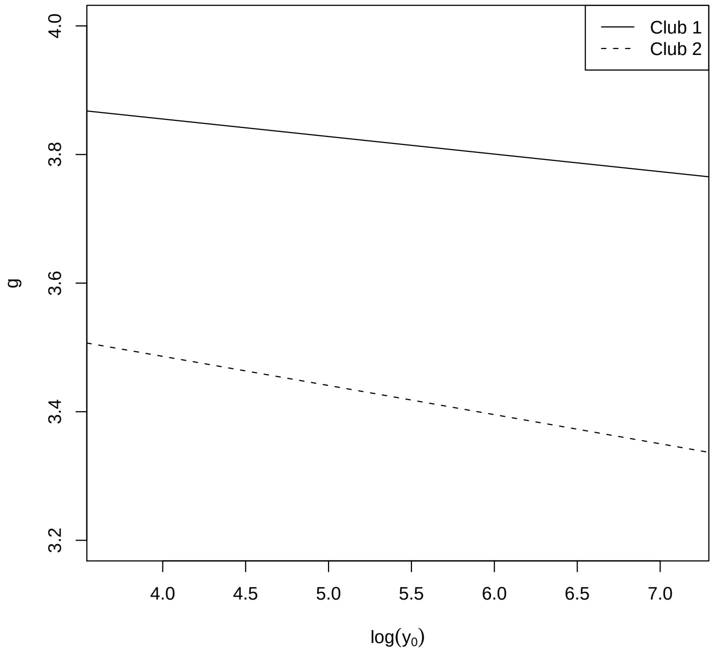
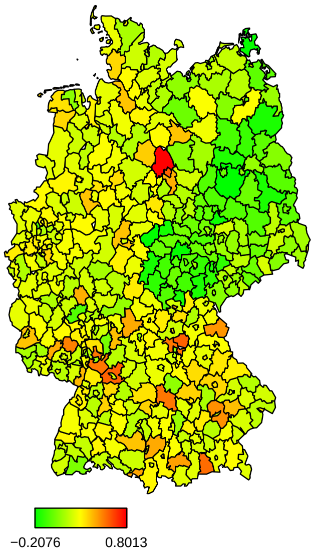
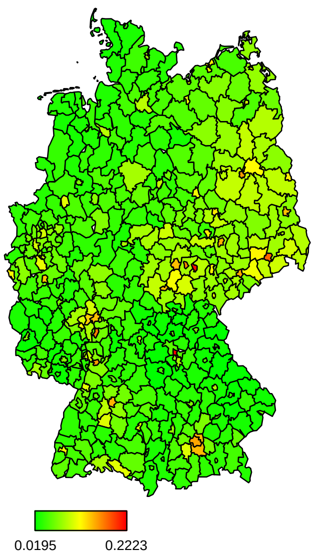


Figure B.16: Estimated regression lines from the estimation displayed in Table A.22 for the two convergence clubs for Japanese prefecture-level data.



(a) `grschool`



(b) $\texttt{school}_0$

Figure B.17: German regional data. Spatial maps for `grschool` and $\texttt{school}_0$.

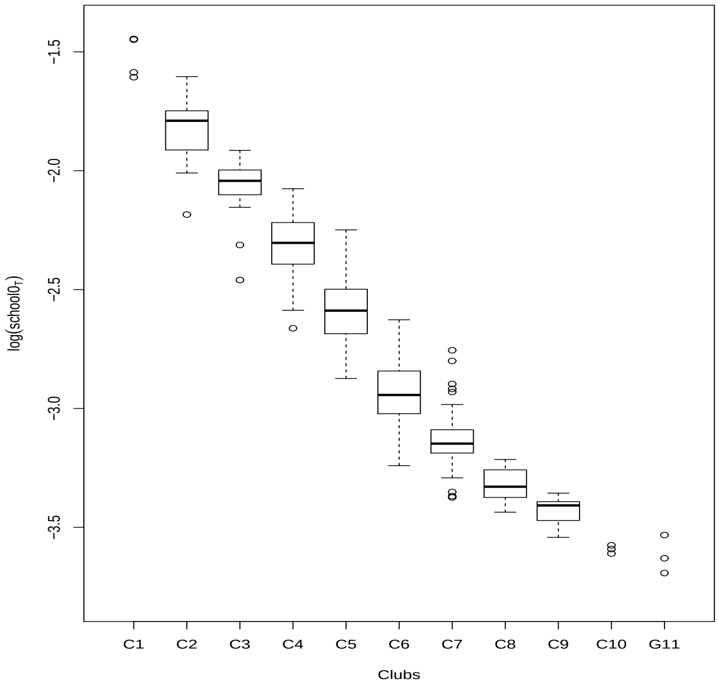


Figure B.18: German regional data. Boxplots of $\log(\mathtt{school}_T)$ for ten convergence clubs (C) and one divergence group (G).

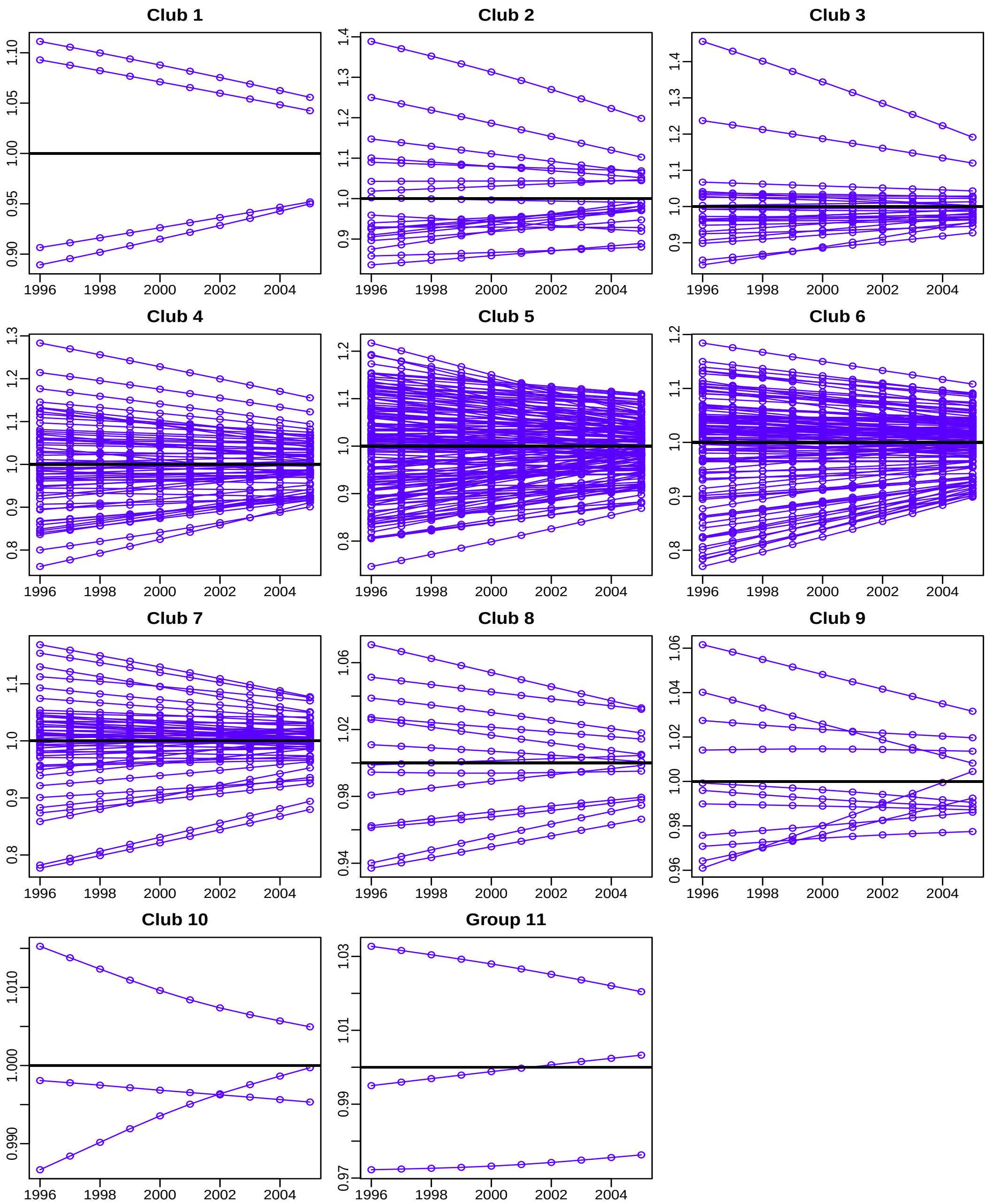


Figure B.19: German regional data. Relative transition paths from time period 0 to T for the convergence clubs and the divergence group.

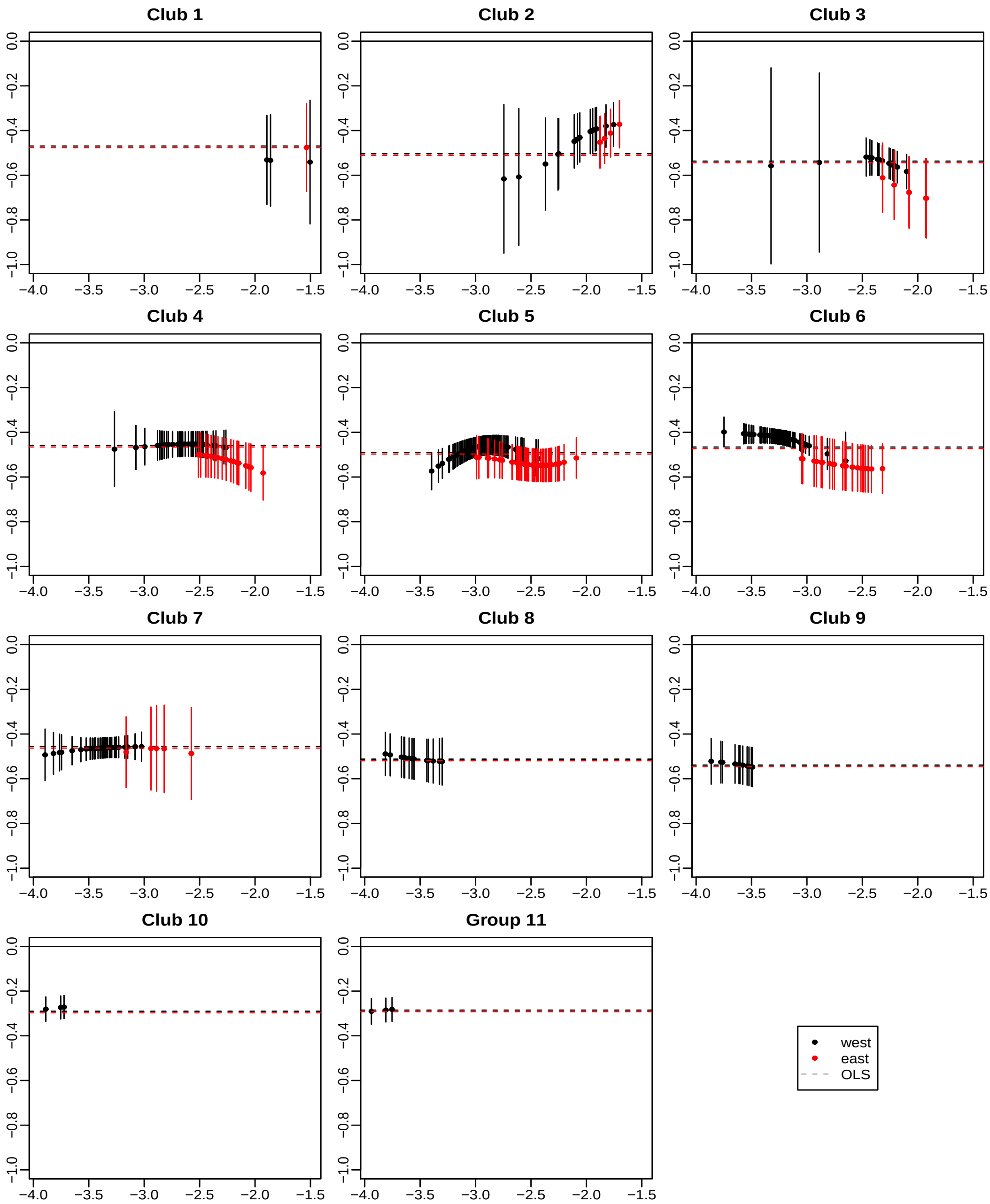


Figure B.20: German regional data. Plots of $\log(\mathtt{school}_0)$ (abscissa) and the estimated partial effects (w.r.t. $\log(\mathtt{school}_0)$, ordinate axis) for the nonparametric regression model (points) of Equation (5.11) and the parametric model (horizontal dashed lines) of Equation (5.7).

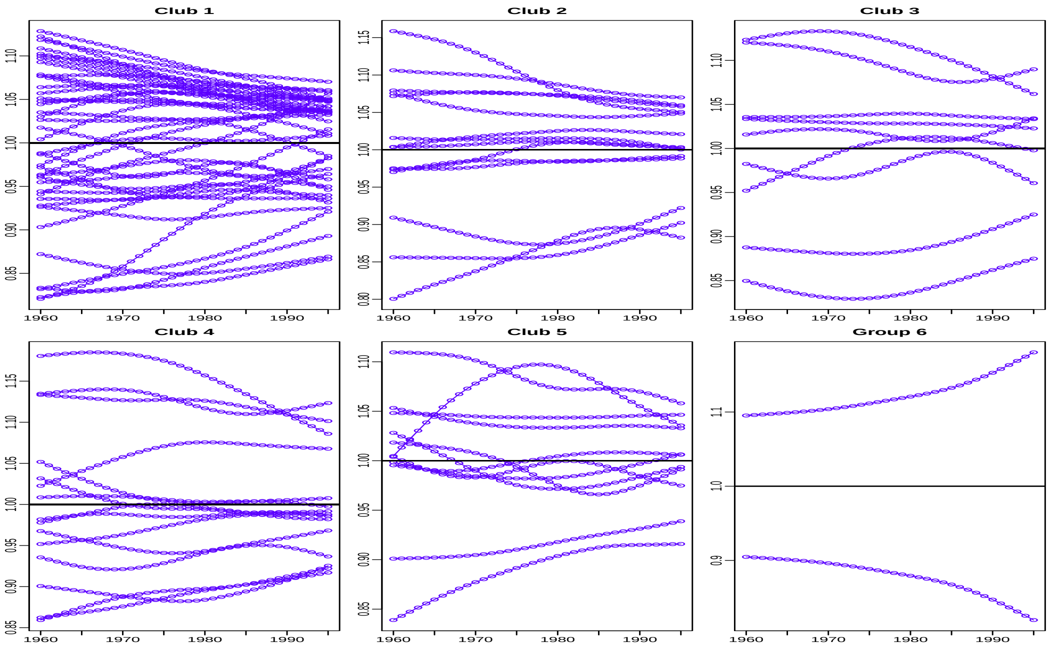


Figure B.21: Penn World Tables data: Relative transition paths from time period 0 to T for the convergence clubs and the divergence group.